This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2017.2662686, IEEE Transactions on Pattern Analysis and Machine Intelligence

The Manhattan Frame Model – Manhattan World Inference in the Space of Surface Normals

Julian Straub, *Member, IEEE,* Oren Freifeld, *Member, IEEE,* Guy Rosman, *Member, IEEE,*
John J. Leonard, *Fellow, IEEE,* and John W. Fisher III, *Member, IEEE*

Abstract—Objects and structures within man-made environments typically exhibit a high degree of organization in the form of orthogonal and parallel planes. Traditional approaches utilize these regularities via the restrictive, and rather local, Manhattan World (MW) assumption which posits that every plane is perpendicular to one of the axes of a single coordinate system. The aforementioned regularities are especially evident in the surface normal distribution of a scene where they manifest as orthogonally-coupled clusters. This motivates the introduction of the Manhattan-Frame (MF) model which captures the notion of a MW in the surface normals space, the unit sphere, and two probabilistic MF models over this space. First, for a single MF we propose novel real-time MAP inference algorithms, evaluate their performance and their use in drift-free rotation estimation. Second, to capture the complexity of real-world scenes at a global scale, we extend the MF model to a probabilistic mixture of Manhattan Frames (MMF). For MMF inference we propose a simple MAP inference algorithm and an adaptive Markov-Chain Monte-Carlo sampling algorithm with Metropolis-Hastings split/merge moves that let us infer the unknown number of mixture components. We demonstrate the versatility of the MMF model and inference algorithm across several scales of man-made environments.

Index Terms—Manhattan World, Manhattan Frame, Mixture of Manhattan Frames, World Representation, Surface Normals

1 Introduction

Simplifying assumptions about the structure of the surroundings facilitate reasoning about complex environments. On a wide range of scales, from the layout of a city to structures such as buildings, furniture and many other objects, man-made environments lend themselves to a description in terms of parallel and orthogonal planes. This intuition is formalized as the Manhattan World (MW) assumption [1] which posits that most man-made structures may be approximated by planar surfaces that are parallel to one of the three principal planes of a common orthogonal coordinate system.

At a local level, this assumption holds for parts of city layouts, most buildings, hallways, offices and other man-made environments. However, the strict MW assumption cannot represent many real-world scenes on a global level: a rotated desk inside a room, more complex room and city layouts (as opposed to planned cities like Manhattan). While local parts of the scene can be modeled as an MW, the entire scene cannot. This suggests a more flexible description of a scene—one that is composed of multiple MWs of different orientations.

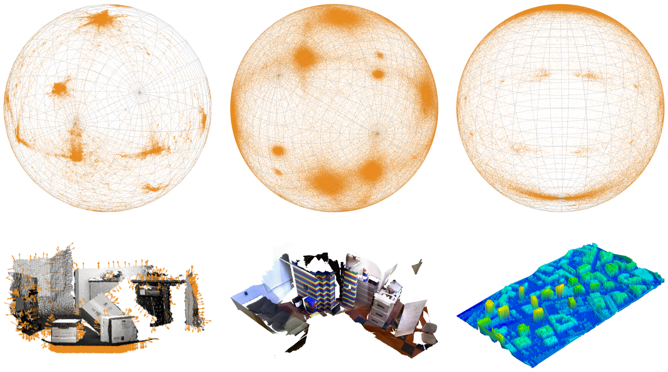


Fig. 1: Across scales, surface normals of man-made environments exhibit characteristic patterns. This work establishes the connection between 3D Manhattan-World structures and their surface-normal distributions via the Manhattan-Frame model.

When reasoning about the composition of a scene in terms of MWs, we directly utilize the surface-normal distribution of the scene, rather than working with a plane segmentation. This is motivated by the observation that across a wide range of scales, man-made environments exhibit structured, low-entropy surface-normal distributions as displayed in Fig. 1. Additionally, surface-normal distributions are invariant to translation and scale [2] which makes the proposed approach largely independent of the measurement and 3D reconstruction process. Finally, surface normals are straightforward to extract from most 3D scene representations such as depth images [3], [4], point clouds [5] and meshes [6], [7].

We introduce the notion of the Manhattan Frame (MF) which represents the MW structure in the space of surface normals, i.e., the unit sphere, as orthogonally-coupled clusters. Modeling surface-normal noise with two different distributions on the sphere, we formulate two probabilistic generative MF models. Depending on the model, real-time MAP inference is carried out in closed form or via optimization on the manifold of rotation matrices $\mathrm{SO}(3)$. Additionally, these models are used to construct a mixture of MFs (MMF) to represent com-

- *J. Straub, Oren Freifeld, Guy Rosman and John W. Fisher III are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 02139.*
- *John J. Leonard is with the Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, 02139.*

0162-8828 (c) 2016 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2017.2662686, IEEE Transactions on Pattern Analysis and Machine Intelligence

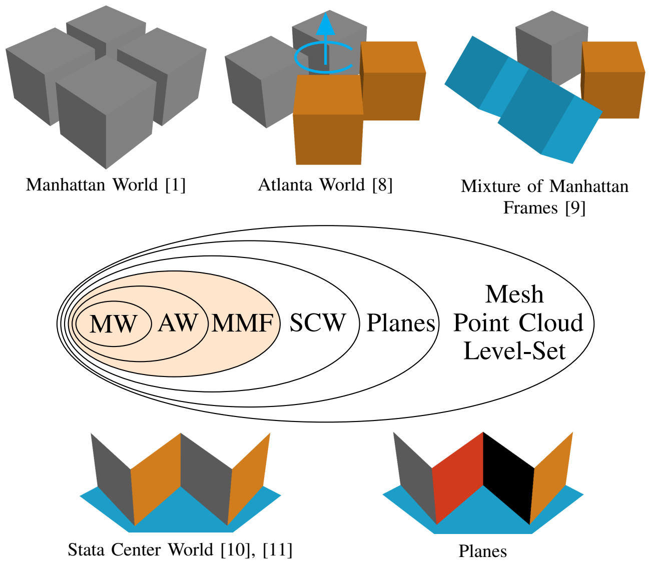


Fig. 2: Structure assumptions about scenes in terms of expressiveness. The proposed MMF subsumes both the Atlanta and the Manhattan World models.

plex scenes composed of multiple MFs. For the MMF model, we derive a simple MAP inference algorithm and a Gibbs-sampling-based algorithm that using Metropolis-Hastings [12] split/merge proposals [13], adapts the number of MFs to best capture the surface-normal distribution of a scene.

We examine the properties of the proposed models and inference algorithms in a variety of qualitative and quantitative experiments. These demonstrate the expressiveness and versatility of the MF and MMF model across scales: depth images of a single view of a scene, larger indoor reconstructions and large-scale aerial LiDAR data of an urban center.

2 Related Work

The different assumptions made in the literature about the environment can be categorized in terms of their expressiveness as depicted in Fig. 2. The assumptions range from mostly unrestricted representations such as point clouds, mesh and level-set, which can in the limit represent any surface exactly, to the rather strict MW assumption as indicated by the inclusion diagram. The proposed MMF assumption subsumes the Atlanta World (AW) which in turn subsumes the MW assumption. The MMF provides a directional segmentation under the orthogonality constraints imposed by the MW assumption. Relaxing the orthogonality constraints completely, we arrive at what we term the Stata Center World (SCW)[1]. It captures only the directional composition of a scene. Planar scene representations differ from the SCW in that different planes with the same orientation are separated in space.

These different assumptions about scenes can be observed directly in the 3D structure or indirectly in the projection of the 3D structure into a camera [14]. Models and inference algorithms based on the former utilize 3D representations such as meshes, point clouds and derived data such as surface normals. Intersections of planes in 3D are lines which can be observed as lines in the image space. A Vanishing Point (VP) is the intersection of multiple such lines where the 3D lines are all parallel to each other. Models built on VPs usually use image gradient orientations directly or indirectly via line segment extraction. Specifically, the MW is manifested as orthogonally-coupled VPs (OVPs) in the image space and an MF in the surface-normal space. Multiple MWs cause multiple OVPs and MFs. The SCW can be observed via independent VPs in the image or independent surface normal clusters.

2D image-space There is a vast literature on VP estimation from RGB images. The goals for VP estimation range from single-image scene parsing [15] and 3D reconstruction [16], [17], [18], [19], VP direction estimation for rotation estimation with respect to man-made environments [1], [20], [21], [22] to VP direction tracking over time to estimate camera rotation and scene structure [23], [24], [25], [26], [27].

While early VP extraction algorithms relied on image gradients [1], [8], most modern algorithms operate on line segments extracted from the image. This has been found to yield superior direction estimation results over dense image-gradient approaches [28]. Generally, VPs are extracted by intersecting lines in the image. These intersections are often found after mapping lines to the unit sphere [23], [29], [30], or into other accumulator spaces [31]. Introduced in [15], horizon estimation has emerged as a benchmark for VP estimation algorithms [32], [33].

Many VP extraction algorithms rely on the MW assumption [1], [20], [22], [25], [26], [32], [34], [35] which is manifested as three OVPs. Incorporating the MW assumption into the VP estimation algorithms not only increases estimation accuracy (if the MW assumption holds) [31] but also allows estimation of the focal length of the camera [20], [31], [32], [35], [36], [37], and rejection of spurious VP detections. Another avenue of research uses the MW assumption for single-image 3D reconstruction [16], [17], [18], [19]. The inferred MW and associations of lines to MW axes combined with geometric reasoning are used to reconstruct the 3D scene in [16], [17]. Hedau et al. [18] use an MW prior to iteratively infer the 3D room layout and segment out clutter in the room. Liu et al. [19] use a floor plan in conjunction with a set of monocular images to reconstruct whole apartments.

The AW model of Schindler et al. [8] assumes that the world is composed of multiple MWs sharing the same z-axis (which is assumed to be known). This facilitates inference from RGB images as only a single angle per MW has to be estimated as opposed to a full 3D rotation. The approach by Antunes et al. [38] infers the full MMF from RGB images. Relaxing the assumptions about the scene, VPs can be extracted independently [15], [21], [23], [24], [28], [30], [31], [33], [39] akin to the SCW assumption.

3D representations There are many approaches that rely purely on 3D representations of surfaces and scenes. Assumptions such as the MW or SCW, are used to align scenes into a common frame of reference for scene segmentation and understanding [40], [41], and to regularize 3D reconstruction [42], [43]. The AW and MMF model could be used similarly.

[1] see http://web.mit.edu/facilities/construction/completed/stata.html

0162-8828 (c) 2016 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2017.2662686, IEEE Transactions on Pattern Analysis and Machine Intelligence

Similar to the image space, the MW assumption has been used most commonly [40], [41]. This is probably due to the fact that man-made environments exhibit strong MW characteristics on a local scale, i.e. on the level of a single RGBD frame of a scene. In the application of Simultaneous Localization and Mapping (SLAM) [44], the MW assumption has been used to impose constraints on the inferred map [43]. Our original idea of the MF [9] has been adapted by Ghanem et al. [45] who propose a robust inference scheme for MF estimation (RMF) and by Kyungdon et al. [46] who use a branch-and-bound scheme to perform real-time globally optimal MF inference (MF BB).

To the best of our knowledge the assumption of multiple MWs in the 3D data setting (as opposed to RGB 2D-images) has not been explored prior to our own work in [9], which was a preliminary and partial version of this manuscript.

Similar to the MF and MMF model, the SCW can be inferred solely from surface-normal distributions [10], [11]. Monszpart et al. [42] couple a local plane-based approach with global directional regularity constraints to regularize 3D reconstructions of man-made environments from point clouds. Gupta et al. [41] assume the only relevant direction for semantic scene segmentation is the direction of gravity to enable alignment of the ground plane across scenes. They propose a simple algorithm to segment the scene into the gravity and all other directions based on surface-normal observations. Triebel et al. [47] extract the main directions of planes in a scene using a hierarchical Expectation-Maximization (EM) approach. Using the Bayesian Information Criterion (BIC) they infer the number of main directions as well. Note, that the MF and MMF model could be inferred from the SCW by grouping inferred directions into MFs.

An alternative to the MW, MMF or SCW model describes man-made structures by individual planes with no constraints on their relative normal directions. The orthogonality constraints in the MW or MMW models enable statistical pooling of measurements across different orientations. This means not only that fewer measurements (per plane) are needed to achieve the same amount of accuracy as without those constraints but also that reliable measurements from one or more directions help in handling cases where there are only few reliable measurements from other directions.

2D & 3D The connection between VPs in images and 3D MW structures has been used to infer dense 3D structure from sets of images by Furukawa et al. [48]. They employ a greedy algorithm for a single-MF extraction from normal estimates that works on a discretized sphere. Neverova et al. [49] integrate RGB images with associated depth data from a Kinect camera to obtain a 2.5D representation of indoor scenes under the MW assumption. Silberman et al. [40] infer the dominant MW using VPs extracted from the RGB image and surface normals computed from the depth image.

3 The Manhattan Frame (MF)

The Manhattan Frame (MF) is the image of a 3D MW structure under the Gauss Map as depicted in Fig. 3. In other words, the MF describes the notion of the MW in the space of surface normals. In a noise-free, perfect MW the surface normals would align with the six directions collected as columns in:

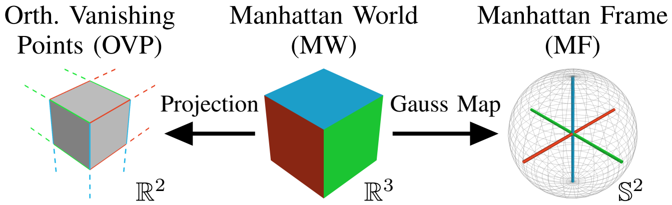


Fig. 3: A MW structure maps to a MF in the surface normal space and to three orthogonal VPs in the image plane.

$$E = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}, \; e_j \text{ denotes the } j\text{th col. of } E. \quad (1)$$

In the camera coordinate system these six directions will appear rotated by R, an element of $\mathrm{SO}(3)$ the space of orthonormal matrices in 3D:

$$M = RE \,, \; \mu_j \text{ denotes the } j\text{th col. of } M. \quad (2)$$

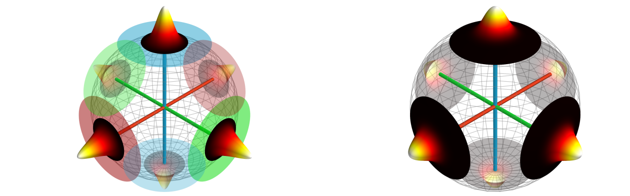
Note that the rotation of the camera, R, is unknown and hence a key parameter to be estimated by an inference algorithm. In other words, if a 3D scene consists of only planar surfaces such that the set of their surface normals is contained in the set $\{\mu_j\}_{j=1}^6$, then M captures all possible directions in the scene—the scene follows the MW assumption.

Specifically, let $q_i \in \mathbb{S}^2$ denote the ith observed surface normal. A latent label, $z_i \in \{1, \ldots, 6\}$, assigns q_i to a specific signed axis within the MF. Hence μ_{z_i} is the signed axis associated with q_i. In the following we will denote the set of all labels as $\mathbf{z} = \{z_i\}_{i=1}^N$ and the set of all surface normals as $\mathbf{q} = \{q_i\}_{i=1}^N$. The unit normals are elements of the unit sphere in $\mathbb{R}^3$, denoted by $\mathbb{S}^2$, a 2D manifold whose geometry is outlined in Sec. 3.2.1. Commonly, in 3D processing pipelines (e.g. in surface fairing or reconstruction [50], [51]), the unit normals are estimated from noisy measurements of the 3D scene structure such as depth images [3], [4], point clouds [5] and meshes [7]. Due to these noise sources and potentially imperfect 3D MW structure, the surface normals $\mathbf{q}$ may deviate from their associated MF axis.

In order to fit the parameters of an MF, one would seek to penalize those deviates. While, in principle, this can be formulated directly as a deterministic optimization, we adopt a probabilistic modeling approach. This allows us to derive a real-time algorithm for single MF inference as well as MCMC inference for the more complicated MMF model all from the same base MF model. Another key advantage over deterministic approaches is that the MCMC algorithm infers all model parameters to facilitate reasoning about uncertainty which is important for scene understanding. Furthermore using MCMC inference, the MMF model can be integrated into larger and more complex environment models, as we showed recently in [7]. Another approach would be to utilize a non-parametric directional segmentation algorithm such as [11] or [10] and to fit MFs to the inferred modes of the surface normal distribution. The advantage of directly inferring an MF model is that data from the different (orthogonal and opposing directions)

0162-8828 (c) 2016 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2017.2662686, IEEE Transactions on Pattern Analysis and Machine Intelligence



(a) Tangent Space Gaussian MF (b) von-Mises-Fisher MF

Fig. 4: Depictions of the two proposed MF noise models.

all jointly contributed to the estimation of the MF orientation. This is especially important in scenes like the urban scene in Fig. 1 where there is only few data points for some of the directions. To this end, we propose two different noise models to describe those random deviations: tangent space Gaussian (TG) noise as well as von-Mises-Fisher (vMF) noise.

3.1 The probabilistic MF model

Let $R \in \mathrm{SO}(3)$ denote the rotation of the MF. Making no a-priori assumptions about which orientation of the MF is more likely than others, R is distributed uniformly:

$$R \sim \mathrm{Unif}(\mathrm{SO}(3)) \,. \tag{3}$$

Since $\mathrm{SO}(3)$ is a manifold with finite support, we can compute its volume and obtain $8\pi^2$ [52] which implies that all rotations have equal likelihood of $^1/_{8\pi^2}$.

As is standard in Bayesian mixture modeling, the MF axis assignments z_i of a surface normal q_i to an MF axis are assumed to be distributed according to a categorical distribution $\mathrm{Cat}(w)$ with a Dirichlet distribution prior parameterized by γ:

$$z_i \sim \mathrm{Cat}(w); \; w \sim \mathrm{Dir}(\gamma) \tag{4}$$

The deviations of the observed normals from their signed axis are modeled by a directional distribution parameterized by Θ. We only require this directional distribution to have the assigned MF axis μ_{z_i} as its mode. Following a Bayesian approach we assume a prior $p(\Theta; \lambda)$ for the parameters:

$$q_i \sim p(q_i \mid z_i, R, \Theta); \; \Theta \sim p(\Theta; \lambda) \tag{5}$$

$$\text{s.t. } \mu_{z_i} = \arg\max_{q \in \mathbb{S}^2} \; p(q \mid z_i, R, \Theta) \,, \tag{6}$$

where λ are the so-called hyperparameters of the prior. Many directional distributions exist (e.g., [53], [54], [55]) and most are valid choices for the distribution of surface normals. We focus on the tangent-space Gaussian (Sec. 3.2) and the von-Mises-Fisher distribution (Sec. 3.3) as depicted in Fig. 4.

Finally, the joint distribution for the MF model is given as:

$$\begin{aligned} p(\mathbf{z}, \mathbf{q}, w, R, \Theta; \gamma, \lambda) = & \tfrac{1}{8\pi^2} p(w; \gamma) p(\Theta; \lambda) \\ & \textstyle\prod_{i=1}^{N} w_{z_i} p(q_i \mid z_i, R, \Theta) \,. \end{aligned} \tag{7}$$

3.2 Tangent Space Gaussian MF Model

The tangent-space Gaussian MF (TG-MF) model describes the deviates not on $\mathbb{S}^2$ directly but in a tangent plane. To explain this concept, we touch upon some differential-geometric notions before describing the TG-MF model.

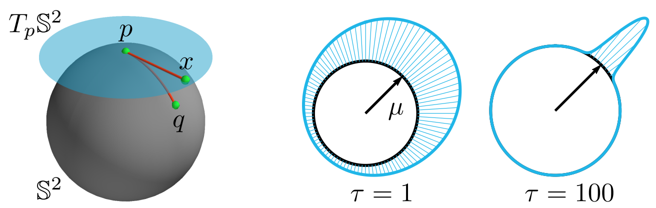


Fig. 5: Left: The unit sphere $\mathbb{S}^2$ in 3D. The blue plane on the sphere illustrates $T_p\mathbb{S}^2$, the tangent space to $\mathbb{S}^2$ at point p. Middle and right: 2D vMF distributions.

3.2.1 The Manifold of the Unit Sphere $\mathbb{S}^2$

As alluded to earlier, surface normals lie on the unit sphere in 3D, $\mathbb{S}^2$. This space is a 2-dimensional Riemannian manifold whose properties we review in the following.

Let p and q be two points on the unit sphere in 3D, $\mathbb{S}^2$, and let $T_p\mathbb{S}^2$ denote the tangent space to $\mathbb{S}^2$ at point p. Then

$$p^T p = q^T q = 1 \text{ and } T_p\mathbb{S}^2 = \{x : x \in \mathbb{R}^3 \; ; \; x^T p = 0\} \,. \tag{8}$$

Note that while $\mathbb{S}^2$ is a nonlinear manifold, $T_p\mathbb{S}^2 \subset \mathbb{R}^3$ is a 2-dimensional linear space as depicted in Fig. 5. It can be shown ([56], [57]) that the length of the shortest path along the manifold, also called the geodesic, between p and q is given by the angle between p and q:

$$d_G(p, q) = \arccos(p^T q) \,. \tag{9}$$

Furthermore, the Riemannian exponential map $\mathrm{Exp}_p : T_p\mathbb{S}^2 \to \mathbb{S}^2$ maps a point x in the tangent space $T_p\mathbb{S}^2$ around p onto the sphere $\mathbb{S}^2$:

$$x \mapsto p \cos(||x||) + \tfrac{x}{||x||} \sin(||x||) \,. \tag{10}$$

The inverse of Exp_p, the Riemannian logarithm map $\mathrm{Log}_p : \mathbb{S}^2/\{-p\} \to T_p\mathbb{S}^2$ can be computed as:

$$q \mapsto (q - p \cos\theta) \tfrac{\theta}{\sin\theta} \,, \tag{11}$$

where $\theta = d_G(p, q)$. Since θ is computed as the angle between p and q it is upper bounded by π for $q = -p$. In that case the logarithm map becomes singular and is thus defined over the entire sphere except the antipodal point $-p$. The Riemannian logarithm map can be thought of as a linearization of $\mathbb{S}^2 \backslash \{-p\}$ to a disk with radius π excluding the boundary of the disk.

In other words, the geodesic distance between two unit normals is the angle between them, Exp_p maps $T_p\mathbb{S}^2$ onto $\mathbb{S}^2$ and Log_p performs the inverse mapping. Note that Exp_p and Log_p depend on p. For further details and an introduction to Riemannian geometry see [57].

Note that in the formulas for the Riemannian exp and log map points $x \in T_p\mathbb{S}^2$ are vectors in the ambient Euclidean space $\mathbb{R}^3$ fulfilling Eq. (8). In the following we express x in local tangent space coordinates as $\tilde{x}$ by parallel transporting all data into the tangent space around the north pole $n = (0, 0, 1)$ using a rotation R. Vectors $x \in T_n\mathbb{S}^2$ all have $x_z = 1$ and hence $\tilde{x}$ is obtained from x by simply discarding the z coordinate. In the following all this process is implicitly assumed whenever a point on the sphere is mapped into a tangent space via the log map. Likewise whenever the exp

0162-8828 (c) 2016 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2017.2662686, IEEE Transactions on Pattern Analysis and Machine Intelligence

map is used the process is performed in the inverse direction: first augment the z direction with 1, then rotate the resulting point down to the tangent space via R^T.

The rotation R can be computed via the axis-angle formulation with axis $w = \frac{n \times p}{\|n \times p\|_2}$ and angle $\theta = \arccos(n^T p)$ using Rodrigues' formula [58]:

$$R(w, \theta) = \mathrm{I} + (\sin\theta)[w]_\times + (1 - \cos\theta)[w]_\times^2 \tag{12}$$

where $[w]_\times$ denotes the construction of a skew-symmetric matrix from a vector w as:

$$[w]_\times = \begin{bmatrix} 0 & -w_3 & w_2 \\ w_3 & 0 & -w_1 \\ -w_2 & w_1 & 0 \end{bmatrix} \tag{13}$$

Note that since $\theta[w]_\times$ is a skew-symmetric matrix, it is an element of the Lie Algebra $\mathrm{so}(3)$ associated with $\mathrm{SO}(3)$. The Exponential map from $\mathrm{so}(3)$ to $\mathrm{SO}(3)$ is equivalent to Rodrigues' formula in Eq. (12) [59].

3.2.2 The TG-MF Model

Under the TG-MF model the deviations of the observed normals from their assigned axis are modeled by a 2D zero-mean Gaussian distribution with covariance $\Theta = \mathbf{\Sigma} \in \mathbb{R}^{2\times 2}$ in the tangent space around the axis. The conjugate prior distribution for covariance matrices $\mathbf{\Sigma}$ is the inverse Wishart (IW) distribution [60] parameterized by $\lambda = \{\Delta \in \mathbb{R}^{2\times 2}, \nu \in \mathbb{R}\}$:

$$p(q_i \mid z_i, R, \Theta) = \mathcal{N}(\mathrm{Log}_{\mu_{z_i}}(q_i); 0, \Sigma_{z_i}) \,, \tag{14}$$

$$p(\Theta; \lambda) = \mathrm{IW}(\Sigma_{z_i}; \Delta, \nu) \tag{15}$$

where $\mathrm{Log}_{\mu_{z_i}}(q_i) \in T_{\mu_{z_i}} S^2$. In other words, we evaluate the probability density function (pdf) of $q_i \in \mathbb{S}^2$ by first mapping it into $T_{\mu_{z_i}}\mathbb{S}^2$ and then evaluating it under the Gaussian distribution with covariance $\Sigma_{z_i} \in \mathbb{R}^{2\times 2}$. The pdf of the surface normals over the nonlinear $\mathbb{S}^2$ is then induced by the Riemannian exponential map:

$$q_i \sim \mathrm{Exp}_{\mu_{z_i}}(\mathcal{N}(0, \Sigma_{z_i})) \tag{16}$$

The range of Log_p is contained within a disk of finite radius (π) while the Gaussian distribution has infinite support. Consequently, for probabilistic inference, we use the inverse Wishart prior to favor small covariances resulting in a probability distribution that, except a negligible fraction, is within the range of Log_p and concentrated about the respective axis.

3.3 von-Mises-Fisher MF Model

In this section we introduce modeling the surface normals as von-Mises-Fisher (vMF) [55], [61] distributed. This distribution is natively defined over the manifold of the sphere and commonly used to model directional data [11], [62], [63], [64]. It defines an isotropic distribution for D-dimensional directional data $q \in \mathbb{S}^{D-1}$ around a mode $\mu \in \mathbb{S}^{D-1}$ with a concentration $\tau \in \mathbb{R}$ and has the following form:

$$\mathrm{vMF}(q; \mu, \tau) = Z(\tau) \exp(\tau q_i^T \mu) \,, \tag{17}$$

where $Z(\tau)$ is the normalizing constant. See Fig. 5 for a 2D illustrative example. Under the vMF model with concentration $\Theta = \tau$, a surface normal $q_i \in \mathbb{S}^2$ is distributed as:

$$p(q_i \mid z_i, R, \Theta) = \mathrm{vMF}(q_i; \mu_{z_i}, \tau) \,, \tag{18}$$

$$p(\Theta; \lambda) \propto Z(\tau)^a \exp(b\tau) \,, \quad \lambda = \{a, b\} \,, \tag{19}$$

where the prior $p(\Theta; \lambda)$ on the concentration parameter of the vMF is only known up to proportionality [65].

4 Real-time MF MAP Inference

Based on the probabilistic generative models for the MF setup in the previous sections, we now develop real-time MF (RTMF) maximum-a-posteriori (MAP) inference methods. These algorithms are used to infer the local MF structure of an environment efficiently. Starting from the TG-MF model we first derive the MAP inference algorithm directly before employing an approximation that yields more efficient inference. Lastly, the vMF-MF MAP inference is derived. Those three MF algorithms are instantiations of the hard-assignment expectation maximization algorithm (EM): We iterate assigning surface normals to the most likely MF axis and updating the MF rotation estimate until convergence.

In this section, for efficiency reasons and in the absence of further knowledge about the scene, the surface normals are assumed to be generated with equal probability from any of the axes, i.e. all $w_j = \frac{1}{6}$. For the same reason we assume identical isotropic covariances $\Sigma_j = \sigma^2 \mathrm{I}$ for all TG-MF axes and identical concentration parameters τ_j for all vMF-MF axes. In Section 5, the TG-MF assumptions will be relaxed.

4.1 Direct MAP MF Estimation for the TG-MF

Starting from the tangent-space Gaussian MF model as set up in Sec. 3.2, we derive the direct MAP MF rotation estimation algorithm. The posterior over assignments z_i of surface normals q_i to axis of the MF is given by

$$p(z_i = j | R, q_i; \pi, \Sigma) \propto w_j \, \mathcal{N}(\mathrm{Log}_{\mu_j}(q_i); 0, \Sigma) \,. \tag{20}$$

Therefore the MAP estimate for the label z_i becomes:

$$\begin{aligned} z_i &= \arg\min\nolimits_{j \in \{1\ldots 6\}} \mathrm{Log}_{\mu_j}(q_i)^T \Sigma^{-1} \mathrm{Log}_{\mu_j}(q_i) \\ &= \arg\min\nolimits_{j \in \{1\ldots 6\}} \arccos^2(q_i^T \mu_j) \,, \end{aligned} \tag{21}$$

where we have used $\arccos(q_i^T \mu_j) = ||\mathrm{Log}_{\mu_j}(q_i)||_2$ and assumed that the covariance Σ is isotropic. With $p(R) = \mathrm{Unif}(\mathrm{SO}(3))$, the posterior over the MF rotation R is

$$\begin{aligned} p(R | \mathbf{q}, \mathbf{z}; \Sigma) &\propto p(\mathbf{q} | \mathbf{z}, R; \Sigma) p(R) \propto p(\mathbf{q} | \mathbf{z}, R; \Sigma) \\ &= \textstyle\prod_{i=1}^N \mathcal{N}(\mathrm{Log}_{\mu_{z_i}}(q_i); 0, \Sigma) \,. \end{aligned} \tag{22}$$

Working in the log-domain, the MAP estimate for R is

$$R^\star = \arg\min\nolimits_R - \log p(R | \mathbf{q}, \mathbf{z}; \Sigma) := \arg\min\nolimits_R f(R) \,. \tag{23}$$

With the posterior in Eq. (22), the cost function $f(R)$ is

$$\begin{aligned} f(R) &= -\log \left[\textstyle\prod_{i=1}^N \mathcal{N}(\mathrm{Log}_{\mu_{z_i}}(q_i); 0, \Sigma) \right] \\ &\propto \textstyle\sum_{i=1}^N \arccos^2(q_i^T \mu_{z_i}) \,, \end{aligned} \tag{24}$$

where we have used a derivation similar to Eq. (21). We call this method direct since the cost function directly penalizes a normal's deviation from its associated MF axis.

0162-8828 (c) 2016 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2017.2662686, IEEE Transactions on Pattern Analysis and Machine Intelligence

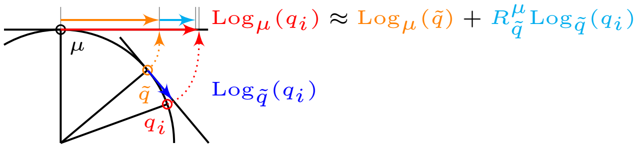

$\mathrm{Log}_\mu(q_i) \approx \mathrm{Log}_\mu(\tilde{q}) + R^\mu_{\tilde{q}}\mathrm{Log}_{\tilde{q}}(q_i)$
μ
$\mathrm{Log}_{\tilde{q}}(q_i)$
$\tilde{q}$
q_i

Fig. 6: The geometry of the approximation of $\mathrm{Log}_\mu(q_i)$.

We enforce the constraints on R ($R^TR = \mathrm{I}$ & $\det(R) = 1$) by explicitly optimizing the cost function on the $\mathrm{SO}(3)$ manifold using gradient descent with backtracking linesearch. More details can be found in [56], [66], [67]. To derive the Jacobian needed for the optimization we use that perturbations of R from R_0 can be written as $R(\omega) = R_0\mathrm{Exp}(\omega)$ where $\omega = \theta[w]_\times \in \mathrm{so}(3)$ and Exp is the exponential map from $\mathrm{so}(3)$ to $\mathrm{SO}(3)$ as introduced in Eq. (12). Then the Jacobian is $J = \frac{\partial f(R(\omega))}{\partial \omega}|_{\omega=0}$, the derivative of f with respect to the perturbation $\omega \in \mathbb{R}^3$ at $\omega = 0$:

$$\begin{aligned} J &= \sum_i \frac{\partial \arccos^2(x)}{\partial x} \frac{\partial}{\partial \omega} q_i^T R\mathrm{Exp}(\omega) e_{z_i}\Big|_{\omega=0} \\ &= \sum_i \frac{2\arccos(q_i^T Re_{z_i})}{\sqrt{1-(q_i^T Re_{z_i})^2}} q_i^T [Re_{z_i}]_\times \end{aligned} \tag{25}$$

Backtracking line search in the direction of the negative Jacobian until the Armijo conditions are met provides an appropriate step size δ which allows us to obtain a new rotation estimate using the exponential map:

$$R_t = R_{t-1}\mathrm{Exp}(-\delta J_t) \tag{26}$$

4.2 Approximate MAP MF Rotation Estimation

The direct approach derived in the previous section is inefficient since the cost function in Eq. (24) and the respective Jacobian involve a sum over all data-points. The Jacobian needs to be re-computed after each update to R and the cost function multiple times during the backtracking linesearch. To address this inefficiency, we derive an approximate estimation algorithm by exploiting the geometry of $\mathbb{S}^2$.

The approximation necessitates the computation of the Karcher means $\{\tilde{q}_j\}_{j=1}^6$ for each of the sets of normals, $\{q_i\}_{\mathcal{I}_j}$, associated with the respective MF axis. The Karcher mean is the generalization of the standard Euclidean sample mean to arbitrary manifolds. It is computed efficiently as described in Appendix A. After this preprocessing step, we approximate $\mathrm{Log}_{\mu_{z_i}}(q_i)$ using the Karcher mean $\tilde{q}_{z_i}$ as proposed in [10]:

$$\mathrm{Log}_\mu(q_i) \approx \mathrm{Log}_\mu(\tilde{q}) + R^\mu_{\tilde{q}}\mathrm{Log}_{\tilde{q}}(q_i)\,. \tag{27}$$

where the subscript z_i was omitted for the sake of clarity and $R^\mu_{\tilde{q}}$ rotates vectors in $T_{\tilde{q}}\mathbb{S}^2$ to $T_\mu\mathbb{S}^2$. Intuitively this approximates the mapping of q_i into μ_{z_i} with the mapping of the Karcher mean into μ_{z_i} plus a correction term that accounts for the deviation of q_i from the $\tilde{q}_{z_i}$. See Fig. 6 for an illustration of underlying geometry. With this the cost function $f(R)$ from Eq. (24) can be approximated by $\tilde{f}(R)$ as

$$\begin{aligned} f(R) \approx \tilde{f}(R) &\propto \sum_{i=1}^N \mathrm{Log}_{\mu_{z_i}}(q_i)^T\Sigma^{-1}\mathrm{Log}_{\mu_{z_i}}(q_i) \\ &\propto \sum_{j=1}^6 |\mathcal{I}_j| \arccos^2(\tilde{q}_j^T\mu_j)\,, \end{aligned} \tag{28}$$

where we have used that the sample mean in the tangent space of their Karcher mean $\sum_{i\in\mathcal{I}_j}\mathrm{Log}_{\tilde{q}_j}(q_i) = 0$ by definition. With $\mu_k = Re_k$ the Jacobian for $\tilde{f}(R)$ is

$$J = \frac{\partial \tilde{f}(R(\omega))}{\partial \omega} = \sum_k \frac{2|\mathcal{I}_k|\arccos(\tilde{q}_j^T Rm_k)}{\sqrt{1-(\tilde{q}_k^T Re_k)^2}} \tilde{q}_k^T [Re_k]_\times\,. \tag{29}$$

Thus the gradient descent optimization over R only utilizes the Karcher means $\{\tilde{q}_j\}_{j=1}^6$, which can be pre-computed since the labels are fixed for the rotation estimation. This eliminates the costly iteration through all data-points at each gradient descent iteration.

4.3 MAP Inference in the MF-vMF Model

In this section, we derive the MAP inference for the vMF-MF model and show that the structure of the vMF distribution allows the MF rotation to be computed in closed form.

With the uniform distribution over labels, i.e. $\pi_j = \frac{1}{6}$, the posterior distribution over label z_i follows the proportionality

$$p(z_i = j|q_i, R; \tau) \propto \mathrm{vMF}(q_i; \mu_j, \tau) \propto \exp(\tau q_i^T\mu_j)\,. \tag{30}$$

Since we assume equal concentration parameter τ for the six vMF distributions, the MAP assignment for z_i is

$$z_i = \arg\max_{j\in\{1,\ldots,6\}} q_i^T\mu_j\,. \tag{31}$$

The posterior distribution over the MF's rotation is:

$$\begin{aligned} p(R|\mathbf{q}, \mathbf{z}; \tau) &\propto p(\mathbf{q}|\mathbf{z}, R; \tau)p(R) \propto p(\mathbf{q}|\mathbf{z}, R; \tau) = \\ &= \prod_{i=1}^N \mathrm{vMF}(q_i|\mu_{z_i}; \tau)\,. \end{aligned} \tag{32}$$

With $N = \sum_{j=1}^6 e_j \sum_{i\in\mathcal{I}_j} q_i^T$ we can find the optimum of the log posterior in closed form by noticing

$$\begin{aligned} \max_{R\in\mathrm{SO}(3)} \sum_{i=1}^N \tau q_i^T\mu_{z_i} &= \max_{R\in\mathrm{SO}(3)} \sum_{j=1}^6 \sum_{i\in\mathcal{I}_j} q_i^T Re_j \\ &= \max_{R\in\mathrm{SO}(3)} \mathrm{tr}\,\{NR\}\,. \end{aligned} \tag{33}$$

This has the same form as the orthogonal Procrustes problem [68]. Hence, the optimal rotation can be computed in closed form using the SVD $N = USV^T$ as

$$R^\star = V\,\mathrm{diag}(1, 1, \mathrm{sign}(\det(UV^T)))\,U^T\,. \tag{34}$$

This has been used before to align point patterns by Umeyama [69] and applied to the MF rotation estimation in a slightly different way by [45].

4.4 Real-time MF inference on streaming data

In the case of a stream of batches of surface normals obtained, for example, from an RGB-D camera, we impose a matrix vMF [70] diffusion model with concentration τ_R. The conditional distribution of the current rotation R given the previous rotation R_- is:

$$p(R|R_-, \tau_R) \propto \exp(\tau_R\,\mathrm{tr}\{R_-^TR\})\,. \tag{35}$$

This distribution is uniform over the rotation space for $\tau_R = 0$ and concentrates on R_- as τ_R increases. See [71] Ex. 2 for a modern treatment of the matrix vMF distribution. In practice we chose $\tau = 1$. This adds the negative log likelihood term

$$f_R = -\tau_R\,\mathrm{tr}\{R_-^TR\} \tag{36}$$

0162-8828 (c) 2016 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2017.2662686, IEEE Transactions on Pattern Analysis and Machine Intelligence

to the MAP cost functions derived in the previous sections. For the direct and the approximate RTMF algorithm this means an additional term in the Jacobian:

$$\begin{aligned}\frac{\partial f_R(R(\omega))}{\partial \omega} &= -\tau_R \frac{\partial}{\partial \omega}\,\mathrm{tr}\{R_-^T R(\omega)\} \\ &= -\tau_R \left[\mathrm{tr}\{R_-^T G_1 R\}\;\; \mathrm{tr}\{R_-^T G_2 R\}\;\; \mathrm{tr}\{R_-^T G_3 R\}\right] .\end{aligned} \tag{37}$$

For the vMF-based algorithm we can still derive a closed from rotation MAP estimate:

$$\begin{aligned}R^\star &= \arg\max_{R\in \mathrm{SO}(3)} \log p(R|\mathbf{q},\mathbf{z};\tau) + \tau_R\,\mathrm{tr}\{R_-^T R\} \\ &= \arg\max_{R\in \mathrm{SO}(3)} \mathrm{tr}\{\tilde{N} R\}\,,\;\; \tilde{N} = N + \tau_R R_-^T\,.\end{aligned} \tag{38}$$

Note that the additional term stemming from the matrix vMF distribution acts as a regularizer if only one MF axis has associated observations.

5 The Mixture of Manhattan Frames

As alluded to in the introduction, the description of man-made environments on a global scale necessitates a more flexible model that can capture Manhattan Worlds with some relative rotation between them. This motivates the extension of the MF framework described in Sec. 3 to the MMF. In practice, scene representations may be composed of multiple intermediate representations, which may include MMFs, to facilitate higher-level reasoning (e.g. [7]). As such, adopting a probabilistic model allows one to describe and propagate uncertainty in the representation. Prior knowledge and model inherent measurement noise can be incorporated in a principled way. Conditional independence allows drawing samples in parallel and hence leads to tractable inference.

In the proposed MMF representation scenes consist of K MFs, $\{M_1,\dots,M_K\}$ which jointly define $6K$ signed axes. For $K=1$, the MMF coincides with the MF. Specifically, let $q_i \in \mathbb{S}^2$ denote the ith observed normal. In the MMF, each q_i has two levels of association. The first, $c_i \in \{1,\dots,K\}$, assigns q_i to the c_ith MF. The second, $z_i \in \{1,\dots,6\}$, assigns q_i to a specific signed axis within the MF M_{c_i} as described in Sec. 3. In the following sections it will be convenient to collect all variables of the kth MF into $\Psi_k = \{\mathbf{c}_k, \mathbf{z}_k, w_k, R_k, \mathbf{\Sigma}_k\}$ where $\mathbf{\Sigma}_k = \{\Sigma_{kj}\}_{j=1}^6$, $\mathbf{c}_k = \{c_i\}_{i:c_i=k}$ and $\mathbf{z}_k = \{z_i\}_{i:c_i=k}$ denote all labels c_i or z_i which are associated to the kth MF via $\mathbf{c}$. The MF axes of the kth MF are a function of the rotation R_k according to Eq. (2) and will be denoted $\{\mu_{kj}\}_{j=1}^6$.

First we define the MMF's probabilistic model before we outline a sampling-based-inference scheme. We restrict the analysis and inference method to the TG-MF model because the vMF distributions in the vMF-MF model necessitate more involved inference methods since the prior on the concentration does not have a closed form as mentioned in Sec. 3.3. Hence an internal slice sampler would be required to sample posterior concentration parameters for the vMF-MF.

5.1 Probabilistic Model

Figure 7 depicts a graphical representation of the probabilistic MMF model. It is a Bayesian finite mixture model that takes into account the geometries of both $\mathbb{S}^2$ and SO(3). In this probabilistic model, the MMF parameters are regarded as random variables and we avoid assumptions from Sec. 4 about weights and covariances of the individual MFs.

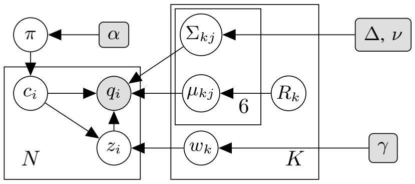


Fig. 7: Graphical model for a mixture of K MFs.

A surface normal q_i is associated with an MF via the assignment variable c_i. These MF-level assignments are assumed to be distributed according to a categorical distribution with a Dirichlet distribution prior with parameters α:

$$c_i \sim \mathrm{Cat}(\pi);\; \pi \sim \mathrm{Dir}(\alpha)\,. \tag{39}$$

Each MF follows the mixture distribution outlined in Sec. 3.2 and hence a surface normal is distributed as

$$q_i \sim p\left(\Psi_{c_i}; \gamma, \Delta, \nu\right)\,. \tag{40}$$

We set $\alpha < 1$ to favor models with few MFs, which is typical for man-made scenes. Contemporary buildings, for example, customarily exhibit a small number of MFs. To encourage the association of equal numbers of normals to all MF axes, we place a strong prior $\gamma \gg 1$ on the distribution of axis assignments z_i. Intuitively, this encourages an MF to explain several normal directions and not just a single one.

5.2 Metropolis-Hastings MCMC Inference

We perform inference over the probabilistic MMF model described in Sec. 5.1 using Gibbs sampling with Metropolis-Hastings [12] split/merge proposals [13]. Specifically, the sampler iterates over the latent assignment variables $\mathbf{c}$ and $\mathbf{z}$, their categorical distribution parameters π and $\mathbf{w} = \{w_k\}_{k=1}^K$, as well as the covariances in the tangent spaces around the MF axes $\mathbf{\Sigma} = \{\mathbf{\Sigma}_k\}_{k=1}^K$ and the MF rotations $\mathbf{R} = \{R_k\}_{k=1}^K$. We first explain all posterior distributions needed for Gibbs sampling before we outline the algorithm.

5.2.1 *Posterior Distributions for MCMC Sampling*

The posterior distributions of both mixture weights are:

$$p(\pi|\mathbf{c};\alpha) = \mathrm{Dir}(\alpha_1 + N_1, \dots, \alpha_K + N_K) \tag{41}$$

$$p(w_k|\mathbf{c},\mathbf{z};\gamma) = \mathrm{Dir}(\gamma_1 + N_{k1}, \dots, \gamma_{k6} + N_{k6})\,, \tag{42}$$

where $N_k = \sum_{i=1}^N \mathbb{1}_{[c_i=k]}$ is the number of normals assigned to the kth MF and $N_{kj} = \sum_{i=1}^N \mathbb{1}_{[c_i=k]}\mathbb{1}_{[z_i=j]}$ is the number of normals assigned to the jth axis of the kth MF. The indicator function $\mathbb{1}_{[a=b]}$ is 1 if $a=b$ and 0 otherwise

Using the likelihood of q_i from Eq. (14), the posterior distributions for labels c_i and z_i are given as:

$$p(c_i = k|\pi, q_i, \Theta) \propto \pi_k \textstyle\sum_{j=1}^6 w_{kj}\, p(q_i; \mu_{kj}, \Sigma_{kj}) \tag{43}$$

$$p(z_i = j|c_i, q_i, \Theta) \propto w_{c_i j}\, p(q_i; \mu_{c_i j}, \Sigma_{c_i j})\,, \tag{44}$$

where $\Theta = \{\mathbf{w}, \mathbf{\Sigma}, \mathbf{R}\}$. We compute $x_i = \mathrm{Log}_{\mu_{c_i z_i}}(q_i)$, the mapping of q_i into $T_{\mu_{c_i z_i}}\mathbb{S}^2$, to obtain the scatter matrix $S_{kj} =$

0162-8828 (c) 2016 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2017.2662686, IEEE Transactions on Pattern Analysis and Machine Intelligence

$\sum_i^N \mathbb{1}_{[c_i=k]}\mathbb{1}_{[z_i=j]} x_i x_i^T$ in $T_{\mu_{kj}}\mathbb{S}^2$. Using S_{kj} the posterior distribution over covariances Σ_{kj} is:

$$p(\Sigma_{kj}|\mathbf{c},\mathbf{z},\mathbf{q},\mathbf{R};\Delta,\nu) = \mathrm{IW}\left(\Delta + S_{kj}, \nu + N_{kj}\right) . \quad (45)$$

Since there is no closed-form posterior distribution for an MF rotation given axis-associated normals, we approximate it as a narrow Gaussian distribution on $\mathrm{SO}(3)$ around the optimal rotation $R_k^\star$ under normal assignments $\mathbf{z}$ and $\mathbf{c}$:

$$p(R_k|\mathbf{z},\mathbf{c},\mathbf{q}) \approx \mathcal{N}(R_k; R_k^\star(R_k^0,\mathbf{z},\mathbf{c},\mathbf{q}), \Sigma_{\mathrm{so}(3)}) , \quad (46)$$

where $\Sigma_{\mathrm{so}(3)} \in \mathbb{R}^{3\times 3}$ and R_k^0 is set to R_k from the previous Gibbs iteration. Refer to Appendix B for details on how to evaluate and sample from this distribution.

The (locally-) optimal rotation $R_k^\star \in \mathrm{SO}(3)$ of MF M_k given the assigned normals $\mathbf{q} = \{q_i\}_{i:c_i=k}$ and their associations z_i to one of the six axes μ_{kz_i} can be found using any of the MAP MF inference algorithms (i.e. Sec. 4.1 or 4.3).

5.2.2 Metropolis-Hastings MCMC Sampling

The Gibbs sampler with Metropolis-Hastings split/merge proposals is outlined in Algorithm 1. For K MFs and N normals the computational complexity per iteration is $O(K^2N)$. To let the order of the model adapt to the complexity of the distribution of normals on the sphere, we implement Metropolis-Hastings-based split/merge proposals. The details of the algorithm are described in the following sections.

Algorithm 1 One Iteration of the MMF Inference

```
Draw $\pi \mid \mathbf{c}; \alpha$ using Eq. (41)
Draw $\mathbf{c} \mid \pi, \mathbf{q}, \mathbf{R}, \Sigma$ in parallel using Eq. (43)
for $k \in \{1, \dots, K\}$ do
    Draw $w_k \mid \mathbf{c}, \mathbf{z}; \gamma$ using Eq. (42)
    Draw $\mathbf{z} \mid \mathbf{c}, \mathbf{w}, \mathbf{q}, \mathbf{R}, \mathbf{\Sigma}$ in parallel using Eq. (44)
    Draw $R_k | \mathbf{z}, \mathbf{c}, \mathbf{q}; \Sigma_{\mathrm{so}(3)}$ using Eq. (46)
    Draw $\{\Sigma_{kj}\}_{j=1}^6 \mid \mathbf{c}, \mathbf{z}, \mathbf{q}, \mathbf{R}; \Delta, \nu$ using Eq. (45)
end for
Propose splits for all MFs
Propose merges for all MF combinations
```

5.3 Split/Merge Proposals

Here we derive split and merge proposals for the MMF model as well as their acceptance probability in an approach similar to Richardson and Green [13]. Note that a merge involves moving all points from MF l and m into a new MF n and then removing MFs l and m. Similarly, a split creates two new MFs l and m from a single MF n. Hence, both a split and a merge change the number of parameters in the model. Specifically the parameters that change their dimension are the set of MF rotations, $\mathbf{R}$, and the set of covariances on the MF axes, $\mathbf{\Sigma}$. The labels $\mathbf{z}$ and $\mathbf{c}$ remain of the same dimensions; only the range for $\mathbf{c}$ changes from $[1, K]$ to $[1, K-1]$ (merge) or from $[1, K]$ to $[1, K+1]$ (split). Therefore, we employ the theory of Reversible Jump Markov Chain Monte Carlo (RJMCMC) [72] to derive a proper acceptance probability. RJMCMC is a generalization of Metropolis-Hastings MCMC [12] and provides a way of computing an acceptance probability when the number of parameters changes between moves. We will see that the split/merge proposals as well as the acceptance probabilities are similar to what one would expect from the Metropolis-Hastings algorithm. For this reason and because the MH algorithm is more well-known, we chose to refer to the inference algorithm as to Metropolis-Hastings MCMC.

5.3.1 RJMCMC Split/Merge Moves in an MMF

RJMCMC utilizes auxiliary variables to propose deterministic moves to change between model orders. In the following, we will give the RJMCMC algorithm for a merge proposal between two MFs. The inverse proposal of a split of an MF into two MFs follows the same but inverted process.

Let a MF A be parameterized by the random variables Ψ_A. An RJMCMC merge proposal between MFs m and l is executed in three steps. First, an auxiliary MF v is sampled from $q(\text{merge})$ to propose a merge of the current MFs l and m as will be described in Section 5.3.2. Second, the deterministic function $f([\Psi_l, \Psi_m, \Psi_v]) = [\mathbf{u}_1, \mathbf{u}_2, \hat{\Psi}_n]$ is used to obtain the merged MF n parameterized by $\hat{\Psi}_n$. The auxiliary MFs $\mathbf{u}_1$ and $\mathbf{u}_2$ absorb the MFs l and m from before the merge. The function $f([\Psi_l, \Psi_m, \Psi_v])$ is hence defined as

$$\mathbf{u}_1 = \Psi_l\,, \quad \mathbf{u}_2 = \Psi_m\,, \quad \hat{\Psi}_n = \Psi_v\,. \quad (47)$$

Therefore, the Jacobian J_f of the function $f([\Psi_l, \Psi_m, \Psi_v])$ is

$$J_f = \frac{\partial f([\Psi_l, \Psi_m, \Psi_v])}{\partial [\Psi_l, \Psi_m, \Psi_v]} = \mathrm{I}\,, \quad (48)$$

where I is the identity matrix with determinant 1.

Third, the proposed merge is accepted with probability

$$\min\left\{1, \frac{\prod_{k=1}^{K-1} p(\hat{\Psi}_k; \alpha, \gamma, \Delta, \nu)}{\prod_{k=1}^{K} p(\Psi_k; \alpha, \gamma, \Delta, \nu)} \frac{q(\text{split})}{q(\text{merge})} \det(J_f)\right\} , \quad (49)$$

where parameters after the merge are designated with a hat. The proposal distributions for a split of MFs l and m into MF n is denoted $q(\text{split})$.

The RJMCMC split proposal of an MF n into MFs l and m follows the same process except that a split is proposed according to Sec. 5.3.3 instead of a merge. The deterministic transformation is the inverse of $f(\cdot)$. This means that the determinant of the Jacobian is 1 and the acceptance probability for the split is Eq. 49 where the ratio has been inverted.

Note that the RJMCMC acceptance probability for split/merge moves in an MMF looks like the Metropolis-Hastings acceptance probability, because $|\det(J_f)| = 1$. However, since the model orders in the nominator and denominator of the fractions are different, it technically is not a Metropolis-Hastings acceptance probability.

5.3.2 Merge Proposal in an MMF

Let the two MFs l and m be parameterized by the random variables Ψ_l and Ψ_m. A merged MF n can be sampled from the current MFs l and m as follows. We first assign all normals of MF l and m to MF n: $\mathbf{c}_{\mathbf{c}\in\{l,m\}} = n$, which corresponds to the proposal distribution:

$$q(\mathbf{c}_l, \mathbf{c}_m|\mathbf{c}) = \delta(\{\mathbf{c}_l, \mathbf{c}_m\} - n)\,. \quad (50)$$

0162-8828 (c) 2016 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2017.2662686, IEEE Transactions on Pattern Analysis and Machine Intelligence

Second, we sample the axes assignments $\mathbf{z}_n$ according to

$$q(z_i = j | w_l, R_l, \mathbf{\Sigma}_l, \mathbf{q}) \quad \propto \quad w_{lj}\, p(q_i; \mu_{lj}, \Sigma_{lj})\,. \tag{51}$$

Next, given associations $\mathbf{c}_n$ and $\mathbf{z}_n$, we find the optimal rotation using the closed form solution of the vMF-based model derived in Sec. 4.3. This is justified because the direct and the vMF-based algorithms generally found the same optimum in our experiments. Then we sample R_n from a narrow Gaussian distribution over rotations with mean $R_n^\star$:

$$\begin{aligned} q(R_n|\mathbf{z},\mathbf{c},\mathbf{q},R_l) &= \mathcal{N}(R_n; R_n^\star(\mathbf{z}_n,\mathbf{c}_n,\mathbf{q}), \Sigma_{\mathrm{so}(3)}) \\ &= \mathcal{N}((R_n^{\star T} \mathrm{Log}_{R_n^\star(\cdot)}(R_n))^\vee; 0, \Sigma_{\mathrm{so}(3)})\,, \end{aligned} \tag{52}$$

where $\mathrm{Log}_{R_n^\star}(R) : \mathrm{SO}(3) \to T_{R_n^\star}\mathrm{SO}(3)$ denotes the logarithm map of R into the tangent space $T_{R_n^\star}\mathrm{SO}(3)$ around $R_n^\star$. The vee operator $^\vee$ [52] extracts the unique elements of a skew-symmetric matrix $W \in \mathbb{R}^{3\times 3}$ into a vector w: $W^\vee = w = [-W_{23}; W_{13}; -W_{12}] \in \mathbb{R}^3$. $\Sigma_{\mathrm{so}(3)} \in \mathbb{R}^{3\times 3}$ is the covariance of the Normal distribution in $T_{R_n^\star}\mathrm{SO}(3)$. Refer to Appendix B for an in depth discussion.

Finally, we obtain samples for the axis covariances $\mathbf{\Sigma}_n$ according to the proposal distribution

$$q(\mathbf{\Sigma}_n|\mathbf{c},\mathbf{z},R_n,\mathbf{q}) = \textstyle\prod_{j=1}^6 p(\Sigma_{nj}|\mathbf{z}_n,\mathbf{c}_n,\mathbf{q},R_n)\,, \tag{53}$$

where $p(\Sigma_{nj}|\mathbf{z}_n,\mathbf{c}_n,\mathbf{q},R_n;\Delta,\nu)$ is the posterior distribution over covariance Σ_{nj} under the IW prior given the assigned normals in the tangent space $T_{\mu_{nj}}\mathbb{S}^2$.

The proposal of merging MF l and m into MF n factors as

$$\begin{aligned} q(\mathrm{merge}) = q(\Psi_n|\Psi_l,\Psi_m,\mathbf{q};\alpha,\gamma,\Delta,\nu) = q(\mathbf{c}_l,\mathbf{c}_m|\mathbf{c}) \\ q(R_n|R_l,\mathbf{z},\mathbf{c},\mathbf{q}) q(\mathbf{\Sigma}_n|\mathbf{c},\mathbf{z},\mathbf{q},R_n;\Delta,\nu) \\ \textstyle\prod_{i:c_i=n} q(z_i|w_l,R_l,\mathbf{\Sigma}_l,\mathbf{q})\,. \end{aligned} \tag{54}$$

5.3.3 Split Proposal in an MMF

First, we randomly assign normals in MF n to MF l or m by drawing MF labels according to the Dirichlet Multinomial (DirMult) distribution:

$$\begin{aligned} q(\mathbf{c}_n|\mathbf{c};\alpha) &= \mathrm{DirMult}\left(\mathbf{c}_n;\alpha_l,\alpha_m\right) \\ &= \textstyle\int_\pi \mathrm{Cat}(\mathbf{c}|\pi)\,\mathrm{Dir}(\pi;\alpha_l,\alpha_m)\,\mathrm{d}\pi\,; \end{aligned} \tag{55}$$

$$\mathrm{DirMult}(\mathbf{c};\alpha) = \tfrac{\Gamma(\sum_{k=1}^K \alpha_k)}{\Gamma(\sum_{k=1}^K \alpha_k + N_k)} \textstyle\prod_{k=1}^K \tfrac{\Gamma(\alpha_k+N_k)}{\Gamma(\alpha_k)}\,. \tag{56}$$

and the counts N_k of labels $c_i = k$ are $N_k = \sum_{i=1}^N \mathbb{1}_{[c_i=k]}$.

Within each of the MFs l and m we assign normals $\mathbf{q}$ to an axis by drawing the assignments $\mathbf{z}_n$ as

$$q(z_i = j|w_n, R_n, \mathbf{\Sigma}_n, \mathbf{q}) \propto w_{nj}\, p(q_i;\mu_{nj},\Sigma_{nj})\,. \tag{57}$$

Using these assignments, we find optimal rotations $R_l^\star$ and $R_m^\star$ and draw R_l and R_m:

$$\begin{aligned} q(R_l,R_m|\mathbf{z},\mathbf{c},\mathbf{q},R_n) = \mathcal{N}(R_l;R_l^\star(R_n,\mathbf{z},\mathbf{c},\mathbf{q}),\Sigma_{\mathrm{so}(3)}) \\ \mathcal{N}(R_m;R_m^\star(R_n,\mathbf{z},\mathbf{c},\mathbf{q}),\Sigma_{\mathrm{so}(3)})\,. \end{aligned} \tag{58}$$

Given rotations as well as labels, we can draw axis covariances $\mathbf{\Sigma}_{\{l,m\}}$ from the respective posterior:

$$\begin{aligned} & q(\mathbf{\Sigma}_{\{l,m\}}|\mathbf{c},\mathbf{z},\mathbf{q},R_{\{l,m\}};\Delta,\nu) \\ & = \textstyle\prod_{j=1}^6 p(\Sigma_{lj}|\mathbf{z},\mathbf{c},\mathbf{q},R_l) p(\Sigma_{mj}|\mathbf{z},\mathbf{c},\mathbf{q},R_m) \end{aligned} \tag{59}$$

The split proposal distribution factors as

$$\begin{aligned} q(\mathrm{split}) = q(\mathbf{x}_l,\mathbf{x}_m|\mathbf{x}_n,\mathbf{q};\alpha,\gamma,\Delta,\nu) = q(\mathbf{c}_n|\mathbf{c};\alpha) \\ q(R_{\{l,m\}}|\mathbf{z},\mathbf{c},\mathbf{q},R_n) q(\mathbf{\Sigma}_{\{l,m\}}|\mathbf{c},\mathbf{z},\mathbf{q},R_{\{l,m\}}) \\ \textstyle\prod_{i:c_i=n} q(z_i|w_n,R_n,\mathbf{\Sigma}_n,\mathbf{q})\,. \end{aligned} \tag{60}$$

5.3.4 RJMCMC Acceptance Probability

After introducing the RJMCMC merge and the split proposals in the previous sections, we will now derive the acceptance probabilities for those two moves by detailing the distributions involved in the computation of Eq. (49).

The joint distribution for the MMF model is defined by the graphical model depicted in Fig. 7. For the evaluation of the acceptance probability, we marginalize over the categorical variables π and $\mathbf{w}$ as in the split proposal in Eq. (55):

$$p(\mathbf{c};\alpha) = \textstyle\int_\pi p(\mathbf{c}|\pi)p(\pi;\alpha)\mathrm{d}\pi = \mathrm{DirMult}(\mathbf{c};\alpha) \tag{61}$$

$$\begin{aligned} p(\mathbf{z}_k|\mathbf{c};\gamma) &= \textstyle\int_{w_k} p(\mathbf{z}_k|\mathbf{c},w_k)p(w_k;\gamma)\mathrm{d}w_k \\ &= \mathrm{DirMult}(\mathbf{z}_k;\gamma)\,, \end{aligned} \tag{62}$$

After marginalization of π and $\mathbf{w}$, the joint distribution is:

$$\begin{aligned} p(\mathbf{q},\mathbf{c},\mathbf{z},\mathbf{\Sigma},\mathbf{R};\alpha,\gamma,\Delta,\nu) = p(\mathbf{c};\alpha) \textstyle\prod_j^6 p(\Sigma_{kj};\Delta,\nu) \\ \textstyle\prod_{i=1}^N p(q_i|c_i,z_i,R_{c_i},\Sigma_{c_i z_i}) \prod_{k=1}^K p(R_k)p(\mathbf{z}_k|\mathbf{c};\gamma)\,, \end{aligned} \tag{63}$$

where we have assumed that the prior over rotations factors according to $p(\mathbf{R}) = \prod_{k=1}^K p(R_k)$. Therefore, the ratio of joint probabilities in the merge move acceptance probability in Eq. (49) becomes

$$\begin{aligned} & \tfrac{p(\mathbf{q},\hat{\mathbf{c}},\hat{\mathbf{z}},\hat{\mathbf{\Sigma}},\hat{\mathbf{R}};\alpha,\gamma,\Delta,\nu)}{p(\mathbf{q},\mathbf{c},\mathbf{z},\mathbf{\Sigma},\mathbf{R};\alpha,\gamma,\Delta,\nu)} = \tfrac{p(\hat{\mathbf{c}};\alpha)p(\mathbf{q}|\hat{\mathbf{c}},\hat{\mathbf{z}},\hat{\mathbf{R}},\hat{\mathbf{\Sigma}})p(\hat{\mathbf{z}}|\hat{\mathbf{c}};\gamma)p(\hat{\mathbf{\Sigma}};\Delta,\nu)p(\hat{\mathbf{R}})}{p(\mathbf{c};\alpha)p(\mathbf{q}|\mathbf{c},\mathbf{z},\mathbf{R},\mathbf{\Sigma})p(\mathbf{z}|\mathbf{c};\gamma)p(\mathbf{\Sigma};\Delta,\nu)p(\mathbf{R})} = \\ & \tfrac{8\pi^2 p(\hat{\mathbf{c}};\alpha)\left(\prod_i^N p(q_i|\hat{c}_i,\hat{z}_i,\hat{\mathbf{R}},\hat{\mathbf{\Sigma}})\right)\prod_{k=1}^{\hat{K}} p(\hat{\mathbf{z}}_k|\hat{\mathbf{c}};\gamma)\prod_{j=1}^6 p(\hat{\Sigma}_{kj};\Delta,\nu)}{p(\mathbf{c};\alpha)\left(\prod_i^N p(q_i|c_i,z_i,\mathbf{R},\mathbf{\Sigma})\right)\prod_{k=1}^K p(\mathbf{z}_k|\mathbf{c};\gamma)\prod_{j=1}^6 p(\Sigma_{kj};\Delta,\nu)}\,, \end{aligned} \tag{64}$$

where $\hat{K} = K - 1$. For a split proposal this ratio is inverted.

The acceptance probability of splits and merges of MFs can be computed, by plugging Eq. (64) into Eq. (49).

5.4 MAP inference in the vMF-MMF model

To design a simple inference algorithm for MAP estimation in the vMF-MMF model we add a hierarchy level to the vMF-MF model from Sec. 4.3 as described in the previous MMF section. For hard-assignment EM-based inference we only additionally require the assignment of data points to an MF. According to the MMF model the labels c_i are distributed as:

$$p(c_i = k|q_i, R;\tau) \propto w_k \textstyle\sum_j \pi_{kj} \exp(\tau q_i^T \mu_{kj})\,. \tag{65}$$

Assuming $\pi_{kj} = \frac{1}{6}$ and τ all equal, like for the derivation of the vMF MAP algorithm, leads to the hard assignment rule:

$$\begin{aligned} c_i &= \arg\max_k w_k \textstyle\sum_j \pi_{kj} \exp(\tau q_i^T \mu_{kj}) \\ &\approx \arg\max_k \arg\max_j q_i^T \mu_{kj} + \log(w_k)\,. \end{aligned} \tag{66}$$

Given the MF-level assignment, data is assigned to MF internal axes as dictated by Eq. 31 and the kth MF rotation updated according to Eq. 34. With the number of normals assigned to the kth MF, $N_k = \sum_{i=1}^N \mathbb{1}_{[c_i=k]}$, we estimate the MF proportions w_k under the Dirichlet distribution prior as:

$$w_k = \tfrac{N_k+\alpha}{N+K\alpha}\,. \tag{67}$$

0162-8828 (c) 2016 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2017.2662686, IEEE Transactions on Pattern Analysis and Machine Intelligence

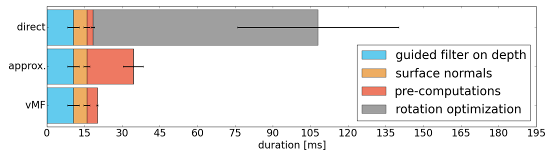


Fig. 8: Timing breakdown for the three different RTMF algorithms. The error bars show the one-σ range.

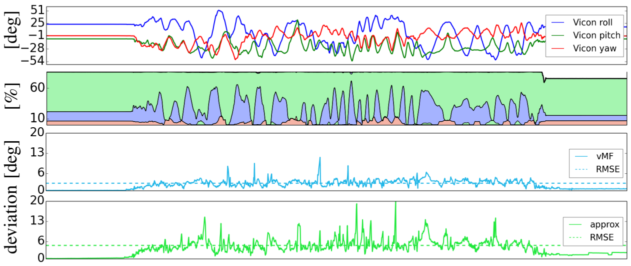


Fig. 9: Rotation estimation accuracy. Percentages of points assigned to each MF axis is color-coded in the second row.

6 Evaluation

We evaluate the properties and performance of the real-time MF (RTMF) before the MMF inference algorithms. All evaluations were run on an Intel Core i7-3940XM CPU at 3.00GHz with an NVIDIA Quadro K2000M GPU.

6.1 Evaluation of real-time MAP Inference

We show run-times and rotation estimation accuracy of all three derived real-time MF inference (RTMF) algorithms on a dataset with groundtruth (GT) camera rotations from a Vicon motion-capture system. The dataset was obtained by waving the camera randomly in full 3D motion up-down as well as left-right in front of a simple MW scene for 90 s as can be seen in the GT yaw-pitch-roll angles in the first row of Fig. 9.

The approximate RTMF algorithm was run for 25 iterations at most while the direct RTMF algorithm was run for at most ten to keep computation time low. Any fewer iterations rendered the direct MF rotation estimation unusable.

Timings We split the computation times into the following stages: (1) applying a guided filter to the raw depth image, (2) computing surface normals from the smoothed depth image, (3) pre-computing of data statistics and (4) optimization for the MF rotation. The timings shown in Fig. 8 were computed over all frames of the dataset. At 111 ms per frame the direct method cannot be run in real-time. While the approximate method improves the runtime, it is 15 ms slower than the vMF-based approach which runs in 18 ms. The approximate algorithm is slower since the Karcher mean pre-computation is iterative whereas the vMF pre-computation is single pass. As intended by shifting all computations over the full data into a pre-processing step, the latter two RTMF algorithms can be run at a camera frame-rate of 30 Hz. In the following we omit the direct method from the evaluation due to its slow runtime.

Accuracy Over the whole dataset the algorithms obtain an angular RMSE from the Vicon groundtruth rotation of $2.5°$ for the vMF-based approach and $4.56°$ for the approximate method showing the higher precision of the vMF-based closed-form rotation estimates. The percentages of surface normals associated with the MF axes displayed in the second row of Fig. 9 support the intuition that a less uniform distribution of normals across the MF axes results in a worse rotation estimate: large angular deviations occur when there are surface normals on only one or two MF axes for several frames. Note that the rotation estimates of the RTMF algorithms are drift free—an important property for visual odometry systems.

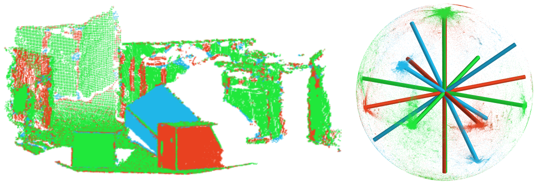


Fig. 10: Inferred MMF in scene and surface normal space.

6.2 Evaluation of MMF Inference

We now evaluate MMF inference on various datasets across scales and compare against MF and VP estimation algorithms.

With the RJMCMC-based approach, we infer an MMF in a coarse-to-fine approach. First, we down-sample to 120k normals and run the algorithm for $T = 150$ iterations, proposing splits and merges throughout as described in Sec. 5.2. We use the following parameters: $\Sigma_{\mathrm{so}(3)} = (2.5°)^2 \, \mathrm{I}_{3\times3}$, $\alpha = 0.01$, $\gamma = 120$, $\nu = 12k$, and $\Delta = (11°)^2\nu \, \mathrm{I}_{2\times2}$. For the purpose of displaying results, we obtain MAP estimates from samples from the posterior distribution of the MMF. First, we find the most likely number of MFs $K^\star$ from all samples after a burn-in of 100 RJMCMC iterations. We then run MCMC starting form the latest sample that has $K^\star$ MFs using all data without proposing splits and merges. All MMF results displayed herein show the last MMF sample of that chain.

The vMF-MMF MAP inference algorithm is sensitive to the initial MF rotations. Hence, we run it 11 times each time starting from 6 randomly rotated MFs and choose one of the models with the moste likely number of MFs after discarding MFs with less than 10% of surface normals.

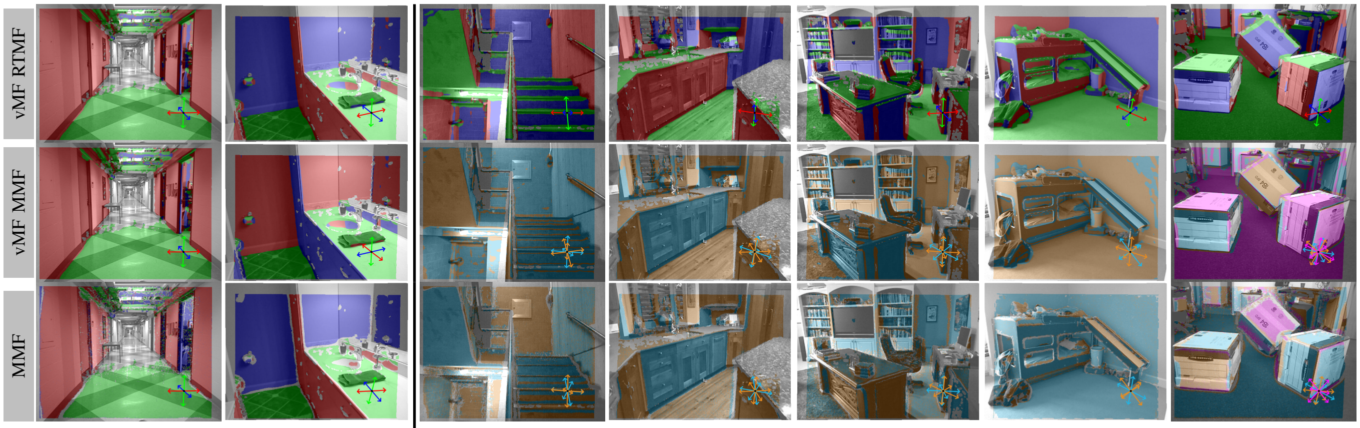
6.2.1 MMF Inference from Depth Images

We first highlight different aspects and properties of the inference using the 3-box scene depicted in Fig. 10. For this scene, we initialized the number of MFs to $K = 6$. The algorithm correctly infers $K = 3$ MFs corresponding to the three differently rotated boxes as displayed in Fig. 10 on the sphere and in the point cloud. While the blue MF consists only of the single box standing on one corner, the green and red MFs contain planes of the surrounding room in addition to their respective boxes. This highlights the ability of our model to pool normal measurements from the whole scene.

In Fig. 11 we show several typical indoor scenes of varying complexity and the inferred MF using the vMF-based RTMF algorithm, the MMF inferred by the MAP-MMF algorithm and the MMF inferred by the RJMCMC algorithm. The MMF inference algorithms were started with six MFs in all cases.

0162-8828 (c) 2016 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2017.2662686, IEEE Transactions on Pattern Analysis and Machine Intelligence



(a) 1 MF (b) 1 MF (c) 2 MFs (d) 2 MFs (e) 2 MFs (f) 2 MFs (g) 3 MFs

Fig. 11: Segmentation and inferred (M)MF of various indoor scenes partly taken from the NYU V2 depth dataset [40]. For single-MF scenes we color-code the assignment to MF axes and for MMF scenes the assignments to MFs.

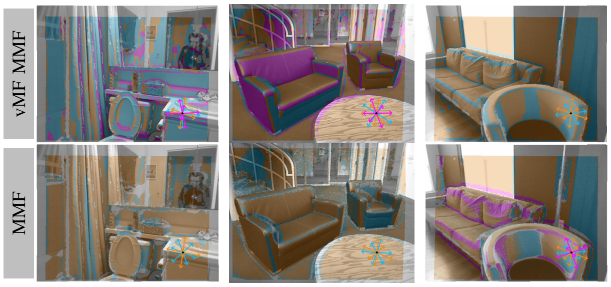


Fig. 12: Common failure cases of the MMF inference.

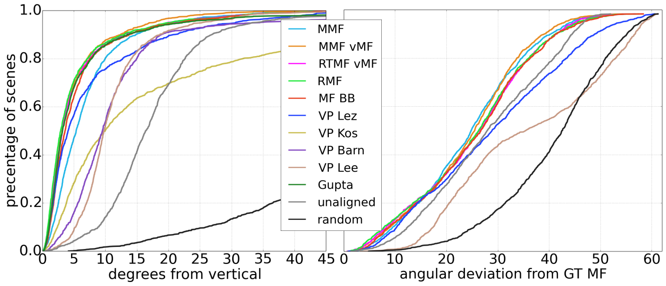


Fig. 13: CDFs of gravity and MW orientation estimation.

For the single MW scenes, all these algorithms infer the same MF, for the multiple-MW scenes the MMF and the MAP-MMF algorithm infer the same reasonable MFs while the vMF RTMF algorithm seems to pick the most prominent MF.

Besides poor depth measurements due to reflections, strong ambient light, black surfaces, or range limitations of the sensor, the inference converged to the wrong number of MFs mainly because of violations of the MW assumption such as round objects or significant clutter in the scene. We observe that the algorithm fails gracefully, approximating round objects with several MFs or adding a “noise MF” to capture clutter as can be seen in Fig. 12. Hence, to eliminate “noise MFs”, we consider only MFs with more than 10% of all normals for the following quantitative evaluation.

To evaluate the performance of the MMF inference algorithms, we ran them on the NYU V2 dataset [40] which contains 1449 RGB-D images of various indoor scenes.

MF count inference For each scene, we compare the number of MFs the algorithm infers to the number of MFs a human annotator perceives. The confusion matrices for the two MMF algorithms are:

$$C_{\text{MMF}} = \left[\begin{smallmatrix} 557 & 467 & 108 & 3 & 0 \\ 130 & 152 & 28 & 1 & 1 \end{smallmatrix}\right] C_{\text{vMF MMF}} = \left[\begin{smallmatrix} 528 & 283 & 186 & 138 \\ 37 & 118 & 83 & 74 \end{smallmatrix}\right] \quad (68)$$

The MMF algorithm infers the human perceived MMFs in 49.0% of the scenes while vMF-MMF is slightly worse with 44.6% and a tendency to overestimate the number of MFs.

MW orientation accuracy We use the groundtruth MW orientation of the most prominent MW provided by [45] to directly evaluate MW orientation estimation accuracy. We take into account that the same MF axes defined according to Eq. (2) can be described by 24 rotations $\{R_{\text{MF},i}\}_{i=1}^{24}$. These are constructed as: for all six permutations of choosing two columns from $R_{\text{MF},i}$, r and r', construct four rotation matrices:

$$[r, r', r_\times],\ [-r, r', -r_\times],\ [r, -r', -r_\times],\ [-r, -r', r_\times] \quad (69)$$

where $r_\times = r \times r'$. To compute the angular deviation θ of an estimated MF to the ground truth MF rotation we construct the set $\{R_{\text{MF},i}\}_{i=1}^{24}$ from the inferred MF rotation, compute all angular rotation deviations to the groundtruth R_{GT} and choose the smallest deviation:

$$\theta = \min_i \arccos(\tfrac{1}{2}\,\text{trace}(R_{\text{GT}}^T R_{\text{MF},i}) - \tfrac{1}{2}) \quad (70)$$

In case of MMFs we choose the smallest deviation across MFs.

Figure 13 (right) depicts cumulative distribution functions (CDF) for the angular deviation of the different MF and MMF algorithms, and two VP extraction algorithms, [31] (VP Lez) and [17] (VP Lee), which extract three OVPs. Evidently, MF algorithms estimate the MW rotation more accurately than the OVP algorithms. The MMF algorithms show higher accuracy on the whole dataset than single MF algorithms.

Gravity direction estimation Like the algorithms by Silberman et al. [40] and Gupta et al. [41] the proposed MF and MMF inference algorithms can be used to estimate the gravity

0162-8828 (c) 2016 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2017.2662686, IEEE Transactions on Pattern Analysis and Machine Intelligence

TABLE 1: Algorithm timings on NYU V2 dataset

Method	RTMF	MMF vMF	MMF	RMF	MF BB	Lez	Lee	Kos	Barn
Time [s]	**0.037**	0.18	3312	23.6	0.061	3.99	5.76	0.21	0.015

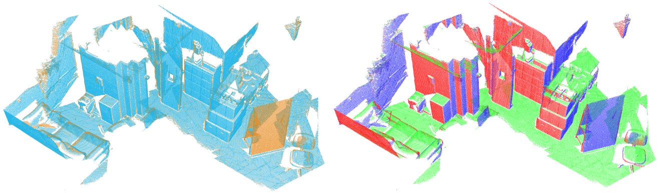


Fig. 14: MMF extracted from a mesh of an indoor couch area (see middle of Fig. 1) obtained using Kintinuous [75].

direction to facilitate rotating scenes into a canonical frame for scene understanding. VPs are also indicative of the gravity direction and we show the performance of two additional VP algorithms [20] (VP Kos) and [29] (VP Barn). The mean direction of surface normals in the scene parts labeled as "floor" serves as a proxy for the true gravity direction in the evaluation.

The CDFs of angular deviation from the gravity direction in Fig. 13 (left) demonstrates that all MF inference algorithms match the performance of Gupta et al. and clearly outperform all VP-based estimates. The inferred MMF models outperform all other methods because of the higher flexibility of the model. The MF algorithms all show similar performance.

Timing Table 1 gives an overview of run-times for the different algorithms averaged over the 1449 scenes from the NYU V2 dataset. RTMF-vMF is the fastest algorithm while the sampling-based algorithm is, unsurprisingly, the slowest. It could, however, be sped up, e.g., by employing a sub-cluster approach for split-merge proposals [10], [74].

6.2.2 *Additional Qualitative MMF Inference Results*

Triangulated meshes Algorithms such as Kintinuous [75] and Elastic Fusion [76] allow dense larger scale indoor reconstructions from a stream of RGBD frames as depicted in the middle of Fig. 1 for a couch area with some boxes, shelves and a Lego house. Using the triangles' surface normals, the MMF can be inferred. The associations to one of the inferred MFs is shown in the middle of Fig. 14 while the associations to the MF axes within each of the MFs is shown to the right.

LiDAR data The large-scale LiDAR scan of Cambridge (Fig. 1 right) has few measurements on the sides of buildings due to reflections off the glass facades and inhomogeneous point density because of overlapping scan-paths. To handle these properties, we implement a variant of robust moving-least-squares normal estimation [77]. Figure 15 shows the point cloud colored according to MF assignment of the normals overlaid on a gray street-map. The inferred MFs share the upward direction without imposing any constraints. The inferred MFs capture large scale organizational structure in this man-made environment: blue and green are the directions of Boston and Harvard respectively, and red is aligned with the Charles river. The locations belonging to the MFs are spatially separated supporting that the MW assumption is best treated as a local property as argued in the introduction.

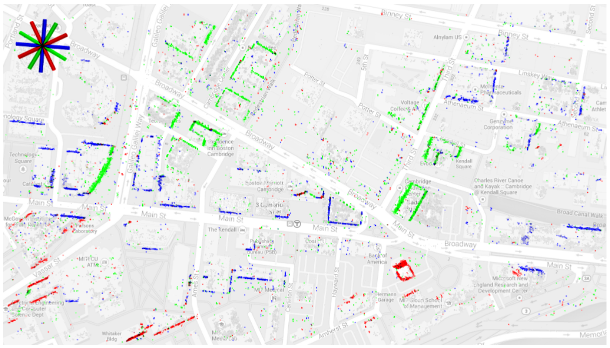


Fig. 15: Inferred MMFs (top left) and MMF labels overlayed on top of a street map of Kendal Square in Cambridge, MA. Normals associated with upward axes are hidden for clarity.

7 Conclusion

Guided by the observation that regularities of the orientations of planes composing man-made environments manifest in structured surface normal distributions, we have proposed the MF model which captures the MW assumption in the space of surface normals. We have formalized the notion of the MF and explored two different probabilistic models and resulting MAP inference algorithms. These real-time capable inference algorithms are useful for extracting the local MW orientation and segmentation of a scene from a stream of data. Motivated by the observation that on a larger scale the commonly-made MW assumption is easily broken, we have extended the MF model to a mixture of MFs. Besides a simple MAP inference algorithm, a manifold-aware Gibbs sampling algorithm with Metropolis-Hastings split/merge proposals, allows adaptive and robust inference of MMFs. This enables the proposed model to describe both complex small-scale-indoor and large-scale-urban scenes. We have demonstrated the versatility of our model by extracting MMFs from 1.5k indoor scenes, larger indoor reconstructions and from aerial LiDAR data of Cambridge, MA. Since VP estimation algorithms showed lower accuracy in both gravity and MF orientation estimation we conclude that if surface normal data is available MF inference algorithms should be used. Beyond the herein demonstrated applications of gravity direction estimation, drift free rotation estimation and scene segmentation, we envision that the MF and MMF models can enable robust and flexible modeling of real-world scenes for a variety of tasks, such as scene understanding, 3D reconstruction and 3D perception.

Appendix A
The Karcher Mean

The Karcher mean $\tilde{q}$ of a set of points on a manifold $\{q_i\}_{i=1}^N$ is a generalization of the sample mean in Euclidean space [78]. It is a local minimizer of the following weighted cost function:

$$\tilde{q} = \arg\min_{p \in M} \sum_{i=1}^{N} w_i d^2(p, q_i) \,. \tag{71}$$

Here, $w_i = 1$, $M = \mathbb{S}^2$, and $d(\cdot,\cdot) = d_G(\cdot,\cdot)$ In this case, excepting degenerate sets, it has a single minimum. It may be computed by the following iterative algorithm:

0162-8828 (c) 2016 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2017.2662686, IEEE Transactions on Pattern Analysis and Machine Intelligence

1) project all $\{q_i\}_{i=1}^N$ into $T_{\tilde{q}_t}\mathbb{S}^2$ and compute their sample mean $\bar{x} = \frac{1}{N}\sum_{i=1}^N \mathrm{Log}_{\tilde{q}_t}(q_i)$.
2) project $\bar{x}$ from $T_{\tilde{q}_t}\mathbb{S}^2$ back onto the sphere to obtain $\tilde{q}_{t+1} = \mathrm{Exp}_{\tilde{q}_t}(\bar{x})$.
3) iterate until $||\bar{x}||_2$ is sufficiently close to 0.

APPENDIX B NORMAL DISTRIBUTION OVER $\mathrm{SO}(3)$

A matrix $R \in \mathbb{R}^{3\times3}$ is called a rotation matrix if it is an element of $\mathrm{SO}(3)$, the Special Orthogonal group; namely, $R^T R = \mathrm{I}$ and $\det(R) = 1$. Probability distributions over rotation matrices can be defined by exploiting the manifold structure of $\mathrm{SO}(3)$ (e.g. [52], [78]). In particular, a way to construct the analog of a Gaussian distribution utilizes the linearity of the tangent spaces. Let $\mathrm{Log}_{R_\mu}(R) : \mathrm{SO}(3) \to \mathrm{so}(3)$ denote the logarithm map of R into the associated Lie Algebra $\mathrm{so}(3)$. With $\theta = \arccos\left(\frac{1}{2}(\mathrm{trace}(R_\mu^T R) - 1)\right)$:

$$\mathrm{Log}_{R_\mu}(R) = \left(\frac{\theta}{2\sin(\theta)}\left(R_\mu^T R - R^T R_\mu\right)\right) . \tag{72}$$

As a member of $\mathrm{so}(3)$, the matrix $W = \mathrm{Log}_{R_\mu}(R)$ is skew-symmetric. The *vee operator* $^\vee$ [52] inverts $[w]_\times$ from Eq. 13 by extracting the unique elements of W into a vector w: $W^\vee = w = [-W_{23}; W_{13}; -W_{12}] \in \mathbb{R}^3$. With this we can define a normal distribution with mean rotation R_μ and covariance $\Sigma_{\mathrm{so}(3)} \in \mathbb{R}^{3\times3}$ in $\mathrm{so}(3)$:

$$p(R; R_\mu, \Sigma_{\mathrm{so}(3)}) = \mathcal{N}(\mathrm{Log}_{R_\mu}(R)^\vee; 0, \Sigma_{\mathrm{so}(3)}) . \tag{73}$$

In order to sample from the distribution in Eq. (73), we sample $w = W^\vee \sim \mathcal{N}(0, \Sigma_{\mathrm{so}(3)})$ and map from $\mathrm{so}(3)$ to $\mathrm{SO}(3)$ using the exponential map $\mathrm{Exp}_{R_\mu} : \mathrm{so}(3) \to \mathrm{SO}(3)$ from Eq. 12 and rotating by R_μ: $\mathrm{Exp}_{R_\mu} = R_\mu R(w, \|w\|_2)$. For further details on $\mathrm{SO}(3)$, the log and exp maps, and the relation to the Lie Algebra $\mathrm{so}(3)$, refer to [52], [57].

ACKNOWLEDGMENTS

We thank J. Chang for his help with the split/merge proposals, R. Cabezas for his help with the Cambridge dataset, T. Whelan for giving access to Kintinuous and R. Finman for helping to scan indoor environments. J.S., O.F., J.L. and J.F. were partially supported by the ONR MURI program, award N00014-11-1-0688. J.S., O.F. and J.F. were also partially supported by the DARPA, award FA8650-11-1-7154. G.R. was partially funded by a MIT-Technion Postdoctoral Fellowship.

REFERENCES

[1] J. M. Coughlan and A. L. Yuille, “Manhattan world: Compass direction from a single image by Bayesian inference,” in *ICCV*, 1999.

[2] B. K. P. Horn, “Extended Gaussian images,” *Proc. of the IEEE*, vol. 72, no. 12, pp. 1671–1686, 1984.

[3] D. Holz, S. Holzer, and R. B. Rusu, “Real-Time Plane Segmentation using RGB-D Cameras,” in *Proc. of the RoboCup Symposium*, 2011.

[4] J. Straub, N. Bhandari, J. J. Leonard, and J. W. Fisher III, “Real-time Manhattan world rotation estimation in 3D,” in *IROS*, 2015.

[5] N. J. Mitra, A. Nguyen, and L. Guibas, “Estimating surface normals in noisy point cloud data,” *International Journal of Computational Geometry & Applications*, vol. 14, no. 04n05, pp. 261–276, 2004.

[6] M. Botsch, L. Kobbelt, M. Pauly, P. Alliez, and B. Lévy, *Polygon mesh processing*. CRC press, 2010.

[7] R. Cabezas, J. Straub, and J. W. Fisher III, “Semantically-Aware Aerial Reconstruction from Multi-Modal Data,” in *ICCV*, 2015.

[8] G. Schindler and F. Dellaert, “Atlanta world: An expectation maximization framework for simultaneous low-level edge grouping and camera calibration in complex man-made environments,” in *CVPR*, 2004.

[9] J. Straub, G. Rosman, O. Freifeld, J. J. Leonard, and J. W. Fisher III, “A mixture of Manhattan frames: Beyond the Manhattan world,” in *CVPR*, 2014.

[10] J. Straub, J. Chang, O. Freifeld, and J. W. Fisher III, “A Dirichlet process mixture model for spherical data,” in *AISTATS*, 2015.

[11] J. Straub, T. Campbell, J. P. How, and J. W. Fisher III, “Small-variance nonparametric clustering on the hypersphere,” in *CVPR*, 2015.

[12] W. K. Hastings, “Monte Carlo sampling methods using Markov chains and their applications,” *Biometrika*, 1970.

[13] S. Richardson and P. J. Green, “On Bayesian analysis of mixtures with an unknown number of components (with discussion),” *Journal of the Royal Statistical Society*, vol. 59, no. 4, pp. 731–792, 1997.

[14] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, Apr. 2004.

[15] O. Barinova, V. Lempitsky, E. Tretiak, and P. Kohli, “Geometric image parsing in man-made environments,” in *ECCV*, 2010.

[16] E. Delage, H. Lee, and A. Y. Ng, “Automatic single-image 3D reconstructions of indoor Manhattan world scenes,” in *Robotics Research*. Springer, 2007, pp. 305–321.

[17] D. C. Lee, M. Hebert, and T. Kanade, “Geometric reasoning for single image structure recovery,” in *CVPR*, 2009.

[18] V. Hedau, D. Hoiem, and D. Forsyth, “Recovering the spatial layout of cluttered rooms,” in *ICCV*, 2009.

[19] C. Liu, A. G. Schwing, K. Kundu, R. Urtasun, and S. Fidler, “Rent3d: Floor-plan priors for monocular layout estimation,” in *CVPR*, 2015.

[20] J. Košecká and W. Zhang, “Video compass,” in *ECCV*, 2002.

[21] M. E. Antone and S. Teller, “Automatic recovery of relative camera rotations for urban scenes,” in *CVPR*, 2000.

[22] J.-C. Bazin and M. Pollefeys, “3-line ransac for orthogonal vanishing point detection,” in *IROS*, 2012.

[23] T. Kroeger, D. Dai, and L. Van Gool, “Joint vanishing point extraction and tracking,” in *CVPR*, 2015.

[24] P. Moghadam and J. F. Dong, “Road direction detection based on vanishing-point tracking,” in *IROS*, 2012.

[25] M. Bosse, R. Rikoski, J. Leonard, and S. Teller, “Vanishing points and three-dimensional lines from omni-directional video,” *The Visual Computer*, vol. 19, no. 6, pp. 417–430, 2003.

[26] A. Flint, D. Murray, and I. Reid, “Manhattan scene understanding using monocular, stereo, and 3d features,” in *ICCV*, 2011.

[27] O. Saurer, F. Fraundorfer, and M. Pollefeys, “Homography based visual odometry with known vertical direction and weak Manhattan world assumption,” *ViCoMoR*, 2012.

[28] P. Denis, J. H. Elder, and F. J. Estrada, “Efficient edge-based methods for estimating Manhattan frames in urban imagery,” in *ECCV*, 2008.

[29] S. T. Barnard, “Interpreting perspective images,” *Artif. Intell.*, 1983.

[30] R. T. Collins and R. S. Weiss, “Vanishing point calculation as a statistical inference on the unit sphere,” in *ICCV*, 1990.

[31] J. Lezama, R. Gioi, G. Randall, and J.-M. Morel, “Finding vanishing points via point alignments in image primal and dual domains,” in *CVPR*, 2014.

[32] H. Wildenauer and A. Hanbury, “Robust camera self-calibration from monocular images of Manhattan worlds,” in *CVPR*, 2012.

[33] Y. Xu, S. Oh, and A. Hoogs, “A minimum error vanishing point detection approach for uncalibrated monocular images of man-made environments,” in *CVPR*, 2013.

[34] E. Lutton, H. Maitre, and J. Lopez-Krahe, “Contribution to the determination of vanishing points using hough transform,” *TPAMI*, 1994.

[35] C. Rother, “A new approach to vanishing point detection in architectural environments,” *Image and Vision Computing*, 2002.

[36] B. Caprile and V. Torre, “Using vanishing points for camera calibration,” *IJCV*, 1990.

[37] R. Cipolla, T. Drummond, and D. P. Robertson, “Camera calibration from vanishing points in image of architectural scenes.” in *BMVC*, 1999.

[38] M. Antunes and J. P. Barreto, “A global approach for the detection of vanishing points and mutually orthogonal vanishing directions,” in *CVPR*, 2013.

[39] J.-P. Tardif, “Non-iterative approach for fast and accurate vanishing point detection,” in *ICCV*, 2009.

[40] P. K. Nathan Silberman, Derek Hoiem and R. Fergus, “Indoor segmentation and support inference from RGBD images,” in *ECCV*, 2012.

[41] S. Gupta, P. Arbelaez, and J. Malik, “Perceptual organization and recognition of indoor scenes from rgb-d images,” in *CVPR*, 2013.

0162-8828 (c) 2016 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2017.2662686, IEEE Transactions on Pattern Analysis and Machine Intelligence

[42] A. Monszpart, N. Mellado, G. J. Brostow, and N. J. Mitra, "Rapter: Rebuilding man-made scenes with regular arrangements of planes," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 103:1–103:12, 2015.

[43] B. Peasley, S. Birchfield, A. Cunningham, and F. Dellaert, "Accurate on-line 3D occupancy grids using Manhattan world constraints," in *IROS*, 2012.

[44] J. J. Leonard and H. F. Durrant-Whyte, "Simultaneous map building and localization for an autonomous mobile robot," in *IROS*, 1991.

[45] B. Ghanem, A. Thabet, J. C. Niebles, and F. Caba Heilbron, "Robust Manhattan frame estimation from a single rgb-d image," in *CVPR*, 2015.

[46] K. Joo, T.-H. Oh, J. Kim, and I. S. KWean, "Globally optimal Manhattan frame estimation in real-time," in *CVPR*, 2016.

[47] R. Triebel, W. Burgard, and F. Dellaert, "Using hierarchical em to extract planes from 3d range scans," in *ICRA*, 2005.

[48] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski, "Reconstructing building interiors from images," in *ICCV*, 2009.

[49] N. Neverova, D. Muselet, and A. Trémeau, "2 1/2D scene reconstruction of indoor scenes from single RGB-D images," in *Computational Color Imaging*, 2013.

[50] A. E. Johnson and M. Hebert, "Surface registration by matching oriented points," in *3DIM*, 1997.

[51] M. Kazhdan, "Reconstruction of solid models from oriented point sets," in *SGP*, 2005.

[52] G. S. Chirikjian, *Stochastic Models, Information Theory, and Lie Groups, Volume 2: Analytic Methods and Modern Applications*. Springer, 2011, vol. 2.

[53] C. Bingham, "An antipodally symmetric distribution on the sphere," *The Annals of Statistics*, vol. 2, no. 6, pp. 1201–1225, 1974.

[54] J. T. Kent, "The Fisher-Bingham distribution on the sphere," *Journal of the Royal Statistical Society*, pp. 71–80, 1982.

[55] K. V. Mardia and P. E. Jupp, *Directional statistics*. John Wiley & Sons, 2009, vol. 494.

[56] P. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[57] M. P. do Carmo, *Riemannian Geometry*. Birkhäuser Verlag, 1992.

[58] O. Rodrigues, *Des lois géométriques qui régissent les déplacements d'un système solide dans l'espace: et de la variation des cordonnées provenant de ces déplacements considérés indépendamment des causes qui peuvent les produire*. 1840.

[59] S. L. Altmann, *Rotations, quaternions, and double groups*. Courier Corporation, 2005.

[60] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis*. CRC press, 2013.

[61] N. I. Fisher, *Statistical Analysis of Circular Data*. Cambridge University Press, 1995.

[62] A. Banerjee, I. S. Dhillon, J. Ghosh, S. Sra, and G. Ridgeway, "Clustering on the unit hypersphere using von Mises-Fisher distributions." *JMLR*, vol. 6, no. 9, 2005.

[63] M. Bangert, P. Hennig, and U. Oelfke, "Using an infinite von Mises-Fisher mixture model to cluster treatment beam directions in external radiation therapy," in *ICMLA*, 2010.

[64] S. Gopal and Y. Yang, "von Mises-Fisher clustering models," in *ICML*, 2014.

[65] G. Nunez-Antonio and E. Gutiérrez-Pena, "A Bayesian analysis of directional data using the von Mises-Fisher distribution," *Communications in StatisticsSimulation and Computation*, 2005.

[66] J. Gallier, "Basics of classical lie groups: The exponential map, lie groups, and lie algebras," in *Geometric Methods and Applications*. Springer, 2001, pp. 367–414.

[67] E. Eade, "Monocular simultaneous localisation and mapping," Ph.D. dissertation.

[68] P. H. Schönemann, "A generalized solution of the orthogonal procrustes problem," *Psychometrika*, vol. 31, no. 1, pp. 1–10, 1966.

[69] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *TPAMI*, no. 4, pp. 376–380, 1991.

[70] C. Khatri and K. Mardia, "The von mises-fisher matrix distribution in orientation statistics," *Journal of the Royal Statistical Society*, pp. 95–106, 1977.

[71] N. Boumal, A. Singer, P.-A. Absil, and V. D. Blondel, "Cramér–rao bounds for synchronization of rotations," *Information and Inference*, vol. 3, no. 1, pp. 1–39, 2014.

[72] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, 1995.

[73] B. K. Horn, "Closed-form solution of absolute orientation using unit quaternions," *JOSA A*, vol. 4, no. 4, pp. 629–642, 1987.

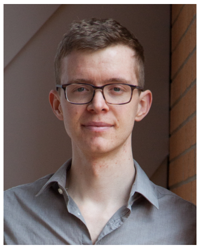
[74] J. Chang and J. W. Fisher III, "Parallel sampling of DP mixture models using sub-clusters splits," in *NIPS*, 2013.

[75] T. Whelan, M. Kaess, M. Fallon, H. Johannsson, J. Leonard, and J. McDonald, "Kintinuous: Spatially extended KinectFusion," in *RSS Workshop on RGB-D: Advanced Reasoning with Depth Cameras*, 2012.

[76] T. Whelan, S. Leutenegger, R. Salas-Moreno, B. Glocker, and A. Davison, "Elasticfusion: dense slam without a pose graph," in *RSS*, 2015.

[77] S. Fleishman, D. Cohen-Or, and C. T. Silva, "Robust moving least-squares fitting with sharp features," in *SIGGRAPH*, 2005.

[78] X. Pennec, "Probabilities and statistics on Riemannian manifolds: Basic tools for geometric measurements," in *NSIP*, 1999.



Julian Straub graduated from the Technische Universität München with a Diplom (2012) and the Georgia Institute of Technology with a M.Sc. (2012), both in Electrical Engineering, within the ATLAS Double Degree Program of the European Union. Currently, he is pursuing a Ph.D. at MIT in Computer Science specializing in artificial intelligence. His research is centered on computer vision, Bayesian inference and 3D environment perception for robots.

Oren Freifeld earned his his MSc (2007) and BSc (2005) in Biomedical Engineering from Tel-Aviv University and his ScM (2009) and PhD (2013) in Applied Mathematics from Brown University. Later, he was a postdoc at MIT Computer Science & Artificial Intelligence Laboratory. He is currently an Assistant Professor at Ben-Gurion University, Department of Computer Science. His main fields of research are computer vision, statistical inference, and machine learning.

Guy Rosman Guy Rosman is a post-doctoral fellow at MIT / CSAIL, where he received the Technion-MIT post-doctoral Fellowship and is working with the Distributed Robotics Lab and the Sensing, Learning and Inference group. He obtained in 2004 his BSc, in 2008 MSc, and in 2013 PhD at the Technion (with the Jacobs-Qualcomm fellowship), in the Computer Science Department. He has worked at several companies/labs, including IBM/HRL, RAFAEL, Medicvision, and Invision Biometrics.

John J. Leonard is Samuel C. Collins Professor of Mechanical and Ocean Engineering and Associate Department Head for Research in the MIT Department of Mechanical Engineering. He is also a member of the MIT Computer Science and Artificial Intelligence Laboratory. His research addresses the problems of navigation and mapping for autonomous mobile robots. He holds the degrees of B.S.E.E. in Electrical Engineering and Science from the University of Pennsylvania (1987) and D.Phil. in Engineering Science from the University of Oxford (1994). He is an IEEE Fellow (2014).

John W. Fisher III (M98) received the Ph.D. degree in Electrical and Computer Engineering from the University of Florida (UF), Gainesville, in 1997. He is currently a Principal Research Scientist in the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology (MIT), Cambridge, and affiliated with the Laboratory for Information and Decision Systems, MIT. Prior to joining MIT, he has been affiliated with UF as both a faculty member and graduate student since 1987. His current area of research focus includes information theoretic approaches to signal processing, multimodal data fusion, machine learning, and computer vision.

0162-8828 (c) 2016 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.