Gulf Journal of Mathematics ISSN: 2309-4966
Vol 22, Issue 2 (2026) 1-26 https://doi.org/10.56947/gjom.v22i2.4130

ENHANCED 3D SHAPE ANALYSIS VIA INFORMATION GEOMETRY

AMIT VISHWAKARMA[1] and K. S. SUBRAHAMANIAN MOOSATH[1*]

Abstract. Three-dimensional point clouds are essential in computer graphics, computer vision, and robotics. Traditional metrics fail to capture global statistical structure, while existing KL divergence approximations for Gaussian mixture models (GMMs) are unbounded and numerically unstable. We establish an information-geometric framework where point clouds, represented as GMMs, form a statistical manifold. We introduce the modified symmetric KL (MSKL) divergence, defined via a generalized KL functional on square-root densities, and prove explicit upper and lower bounds guaranteeing finite, stable values. Experiments on MPI-FAUST and G-PCD datasets demonstrate MSKL outperforms traditional distances and existing KL approximations in shape discrimination tasks.

Keywords. Information Geometry, Point Cloud, Gaussian Mixture Model, Statistical Manifold, Divergence
2020 Mathematics Subject Classification. 53B12 , 62B11

1. Introduction and preliminaries

Three-dimensional point clouds provide highly accurate digital representations of objects, making them important for applications in computer graphics, photogrammetry, computer vision, construction, and remote sensing [36, 5, 1]. Recent work has addressed various geometric analysis tasks including robust symmetry detection in the presence of outliers [27]. Extracting meaningful information from point clouds is fundamental for shape analysis tasks such as pose estimation, deformation tracking, registration, retrieval and autoencoding [18, 39, 15]. However, comparing point clouds faces unique challenges due to their unstructured nature and the complex geometry of the surfaces they represent. Unlike structured grid data with natural Euclidean metrics, point clouds lack a common coordinate system, making direct comparisons difficult. When comparing shapes sampled from 3D surfaces, the objective is to capture global geometric similarity rather than point-wise correspondence.

Date: Received: Jan 19, 2026 ; Revised: Feb 18, 2026; Accepted: March 2, 2026.
* Corresponding author
Preprint: A preprint version of this article is available at [33].

Classical geometric distances such as Hausdorff [13] and Chamfer [38] measure spatial proximity between point sets. While effective for certain tasks, they often fail to capture broader statistical structure, exhibit sensitivity to outliers and sampling irregularities, and do not characterize the global distribution of points.

Earth Mover's Distance (EMD) [30] computes similarity as the minimum cost of transforming one distribution into another, providing a transport-based metric. However, EMD requires solving an expensive optimization problem that scales poorly with point cloud size and often struggles to capture intrinsic geometric relationships between the shapes.

Recent works have explored probabilistic representations using GMMs for point cloud registration and comparison [16]. While these methods make use of statistical modeling, they generally rely on approximations of the Kullback–Leibler (KL) divergence and these approximations have some limitations. KL Weighted-Average ($\mathrm{KL}_{\mathrm{WA}}$) [9, 14] approximation provides a lower bound by averaging pairwise KL divergences between Gaussian components weighted by mixing coefficients. However, this approximation can significantly underestimate the true divergence, particularly when mixture components have poor overlap. KL Matching-Based ($\mathrm{KL}_{\mathrm{MB}}$) [9, 14] approximation matches each component of one mixture to the closest component in another, yielding an upper bound. This matching strategy can overestimate divergence and is sensitive to the number and initialization of components. Furthermore, the standard KL divergence is asymmetric ($\mathrm{KL}(p||q) \neq \mathrm{KL}(q||p)$) and can be unbounded when comparing GMMs, leading to numerical instability and undefined comparisons in practical scenarios. In related statistical estimation contexts, MMD-based approaches have been shown to provide robustness to outliers and data contamination in distribution comparison tasks [4], highlighting the broader need for stable and well-behaved measures in probabilistic modeling.

This work develops an information-geometric framework for stable shape similarity measurement by representing each point cloud as a GMM defined on a low-dimensional latent space. The framework includes preprocessing, extraction of local descriptors, manifold embedding via Isomap, and GMM fitting through the Expectation–Maximization (EM) algorithm. With this representation, the collection of all such GMMs forms a statistical manifold, enabling the use of divergence-based tools for shape analysis. We introduce the MSKL divergence, defined through a generalized KL functional on square-root densities. This addresses the instability issue of existing KL approximations. The theoretical properties, including bounds and stability for MSKL are established in the subsequent sections. In addition, we consider two well-known approximations for KL divergence on GMMs—the weighted-average (KL_{WA}) and matching-based (KL_{MB})—which provide complementary baselines frequently used in mixture model comparison.

The main contributions of this work are (i) establishing a rigorous mathematical framework demonstrating that point clouds represented as GMMs form a statistical manifold, enabling information-geometric methods; (ii) introducing the MSKL divergence with proven upper and lower bounds guaranteeing numerical stability; (iii) comparing MSKL with symmetric KL approximations ($\mathrm{SKL}_{\mathrm{WA}}$, $\mathrm{SKL}_{\mathrm{MB}}$) and classical geometric distances (Hausdorff, Chamfer).

To demonstrate the practical performance of the framework, we compare MSKL with Symmetric KL divergence (SKL), $\mathrm{SKL}_{\mathrm{WA}}$, $\mathrm{SKL}_{\mathrm{MB}}$ and the traditional distances Hausdorff, and Chamfer across two shape categories: (i) human shapes from the MPI FAUST dataset [25] and (ii) animal shapes from the G-PCD dataset [8]. These experiments highlight how statistical manifold representations capture both local geometry and global structure more effectively than metrics.

2. Related Works

In this section, we review existing methods for comparing point clouds, which can be broadly classified into two categories: metric-based distances and divergence-based statistical approaches.

2.1. **Metric-Based Approaches.**

1. Hausdorff Distance: For two finite point sets $P, Q \subset \mathbb{R}^m$ the Hausdorff distance [13] is defined as

$$D_H(P,Q) = \max\left\{\sup_{p\in P}\inf_{q\in Q}\|p-q\|, \sup_{q\in Q}\inf_{p\in P}\|p-q\|\right\}. \tag{2.1}$$

This measure emphasizes the largest difference between two sets of 3D points, which makes it particularly sensitive to unusual points that do not fit in the pattern.

2. Chamfer Distance: The Chamfer distance [38] provides an alternative that averages point-to-point distances, thereby reducing the impact of outliers. It is widely used in 3D reconstruction and mesh generation.

$$Ch(P,Q) = \frac{1}{|P|}\sum_{p\in P}\min_{q\in Q}\|p-q\| + \frac{1}{|Q|}\sum_{q\in Q}\min_{p\in P}\|p-q\|. \tag{2.2}$$

This method calculates the average of the squared distances between each point in one set to its nearest point in the other set, offering a balance between sensitivity to large discrepancies and outliers. This method does not account for differences in point cloud shape and distribution.

3. Earth Mover's Distance: The Earth Mover's Distance (EMD) formulates point cloud comparison as an optimal transport problem:

$$\mathrm{EMD}(A,B) = \min_{\phi: A\leftrightarrow B}\sum_{a\in A}\|a-\phi(a)\|, \tag{2.3}$$

where ϕ denotes a bijection between point clouds A and B. While EMD provides a principled transport-based metric, it requires solving an expensive optimization problem that scales poorly with point cloud size ($O(n^3 \log n)$), making it impractical for large-scale applications [30].

4. Gromov-Hausdorff Distance: In [13] and [38] comparing two objects reduced to the comparison between the corresponding interpoint distance matrices in the multidimensional scaling. These methods primarily capture rigid similarities and do not account for isometric objects. To address this limitation, Facundo Mémoli and Guillermo Sapiro [23] introduced a method based on the Gromov-Hausdorff

distance, which provides a theoretical framework for isometric-invariant recognition. The Gromov-Hausdorff distance, d_{GH} measures the similarity between two metric spaces X and Y by embedding them into a common metric space Z and minimizing the Hausdorff distance $d_H^Z(X,Y)$. The Gromov-Hausdorff distance is defined as

$$d_{GH}(X,Y) = \inf_{Z,f,g} d_H^Z(f(X), g(Y)), \tag{2.4}$$

where, $d_H^Z(f(X), g(Y)) = \max\left(\sup_{x\in f(X)} d_Z(x, g(Y)), \sup_{y\in g(Y)} d_Z(y, f(X))\right)$. Here, f and g are isometric embeddings of X and Y into Z, and d_Z is the intrinsic geodesic distance on Z. Since there is no efficient way to directly compute the Gromov-Hausdorff distance, they introduced a metric d_J,

$$d_J(X,Y) = \min_{\pi\in\mathcal{P}_n} \max_{1\le i,j\le n} \frac{1}{2} \left| d_X(x_i,x_j) - d_Y(y_{\pi i}, y_{\pi j})\right|, \tag{2.5}$$

where $\mathcal{P}_n$ is the set of all permutations of $\{1,2,...,n\}$, n is the number of points in the point cloud. This metric satisfies $d_{GH} \le d_J$. The metric d_J is computable and can be used to replace d_{GH}.

The framework for comparing point clouds developed in [23] may face challenges with highly complex or higher-dimensional geometric structures due to computational constraints.

Beyond distance-based comparisons, recent approaches address robust point cloud processing using statistical methods. For instance, Gaussian process regression has been applied to point cloud denoising and outlier detection [17], demonstrating the broader landscape of statistical techniques for point cloud analysis.

2.2. **Divergence and Distance Based Measures on GMMs.** Recent works have explored representing point clouds as probability distributions, most commonly using GMMs. Several registration and comparison methods have been developed. GMM-based registration by Jian and Vemuri [16] used Gaussian mixtures with L^2 distance for point set registration. Myronenko and Song [26] developed the Coherent Point Drift algorithm, treating one point cloud as a GMM and aligning it to another through probabilistic correspondence. Some Divergence-based methods by Wang et al. [35] introduced cumulative distribution functions with Jensen-Shannon divergence for groupwise registration. Hasanbelliu et al. [12] employed correntropy and Cauchy-Schwarz divergences. Li et al. [22] proposed novel bandwidth estimation strategies for GMM-based comparison.

While the above methods use GMMs for registration, they do not exploit the underlying geometric structure of the space of GMMs. Lee et al. [21] introduced a statistical manifold framework using kernel density estimation, defining parametric probability density functions as statistical representations of point clouds. In contrast, we interpret point clouds as samples from GMMs and endow the space of GMMs with a statistical manifold structure. This enables the use of information-geometric tools such as divergences, connections, and curvature—for rigorous shape analysis. Our recent work [34] further embeds this statistical manifold into the manifold of symmetric positive definite matrices, enriching the geometric structure.

3. Information Geometry

Information geometry studies non-Euclidean geometric structures on spaces of probability distributions, with applications to statistical inference and estimation as well as modern areas such as machine learning and deep learning. Metrics or divergences between probability distributions are important in practical applications to look at similarity or dissimilarity among the given set of samples. A brief account of the geometry of statistical manifolds is given [2], [11]. For basic knowledge in differential geometry see [20].

3.1. **Statistical manifold.** The manifold structure and its geometry for a statistical model are discussed here. Let $(\mathcal{X}, \Sigma, p)$ be a probability space, where $\mathcal{X} \subseteq \mathbb{R}^n$. Consider a family $\mathcal{S}$ of probability distributions on $\mathcal{X}$. Suppose each element of $\mathcal{S}$ can be parametrized using n real-valued variables $(\theta^1, ..., \theta^n)$ so that

$$\mathcal{S} = \{p_\theta = p(x;\theta) \ \mid \ \theta = (\theta^1, ..., \theta^n) \in \Theta\} \tag{3.1}$$

where Θ is an open subset of $\mathbb{R}^n$ and the mapping $\theta \mapsto p_\theta$ is injective. The family $\mathcal{S}$ is called an n-dimensional statistical model or a parametric model.

For a model $\mathcal{S} = \{p_\theta \ \mid \ \theta \in \Theta\}$, the mapping $\varphi : \mathcal{S} \longrightarrow \mathbb{R}^n$ defined by $\varphi(p_\theta) = \theta$ allows us to consider $\varphi = (\theta^i)$ as a coordinate system for $\mathcal{S}$. Suppose there is a c^∞ diffeomorphism $\psi : \Theta \longrightarrow \psi(\Theta)$, where $\psi(\Theta)$ is an open subset of $\mathbb{R}^n$. Then, if we use $\rho = \psi(\theta)$ instead of θ as our parameter, we obtain $\mathcal{S} = \{p_{\psi^{-1}(\rho)} \mid \rho \in \psi(\Theta)\}$. This expresses the same family of probability distributions $\mathcal{S} = \{p_\theta\}$. If parametrizations which are c^∞ diffeomorphic to each other is considered to be equivalent then $\mathcal{S}$ is a c^∞ differentiable manifold, called the statistical manifold (see Figure 1)[2, 3, 28].

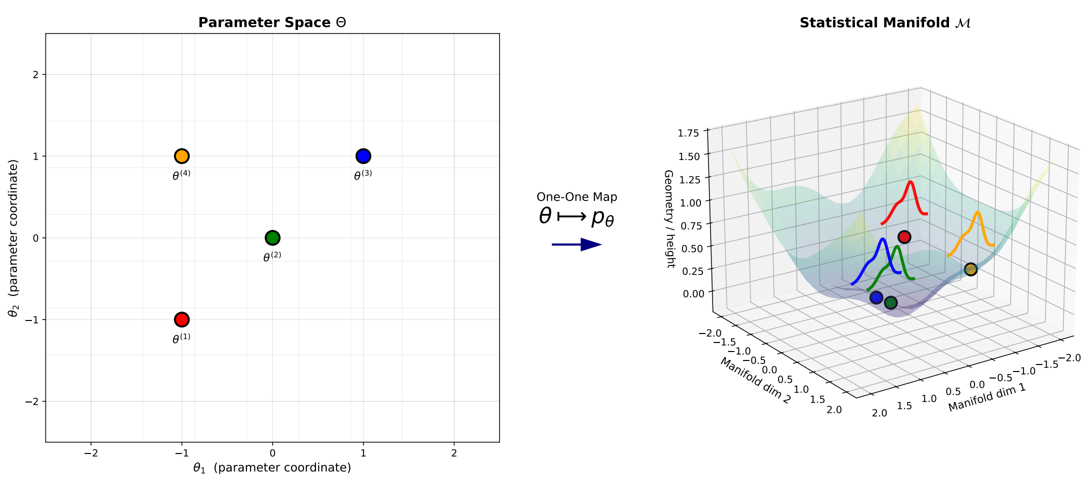


Figure 1. Statistical manifold representation shows the map between the parameter space and the space of probability distributions.

3.2. **Divergence Measures on Statistical Manifolds.** Divergence is a distance-like measure between two points (probability density functions) on a statistical manifold. The divergence D on $\mathcal{S}$ is defined as $D = D(\cdot\|\cdot) : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$, a smooth function satisfying, for any $p, q \in \mathcal{S}$

$$D(p\|q) \geq 0 \quad \text{and} \quad D(p\|q) = 0 \iff p = q.$$

The Kullback-Leibler (KL) divergence for probability density functions is defined as [19]

$$D_{KL}(p\|q) = \int p(x;\theta_1) \log \frac{p(x;\theta_1)}{q(x;\theta_2)}\, dx \tag{3.2}$$

where $p(x;\theta_1)$ and $q(x;\theta_2)$ are probability density functions satisfying $\int p\, dx = \int q\, dx = 1$.

3.2.1. *Generalized KL Divergence for Positive Measures.* To construct a stable, bounded divergence measure suitable for GMM comparison, we consider applying KL-type functionals to transformed densities. However, if p is a probability density, then $\sqrt{p}$ is a positive function that does not integrate to unity. To handle this rigorously, we employ the *Generalized KL divergence* for positive normalizable measures [40].

Definition 3.1. For positive measurable functions $f, g : \mathbb{R}^m \to \mathbb{R}_+$ with $0 < \int f\, dx < \infty$ and $0 < \int g\, dx < \infty$, the Generalized KL divergence is defined as

$$D_{Gen}(f\|g) = \int \left(f(x) \log \frac{f(x)}{g(x)} - f(x) + g(x) \right) dx \tag{3.3}$$

This extends the standard KL divergence to unnormalized positive measures. When $\int f = \int g = 1$, the mass terms cancel and this reduces to the standard KL divergence.

3.2.2. *MSKL Divergence.* We now define the central divergence measure of this work.

Definition 3.2. For probability density functions p and q on $\mathbb{R}^m$, the *MSKL divergence* is defined as the symmetrized Generalized KL divergence between their square-root representations $\sqrt{p}$ and $\sqrt{q}$

$$D_{\text{MSKL}}(p\|q) = \frac{1}{2}\left[D_{Gen}(\sqrt{p}\|\sqrt{q}) + D_{Gen}(\sqrt{q}\|\sqrt{p}) \right]. \tag{3.4}$$

Remark 3.3. Substituting Equation (3.3) into Equation (3.4), the linear mass terms $(-\int \sqrt{p} + \int \sqrt{q})$ from the forward divergence and $(-\int \sqrt{q} + \int \sqrt{p})$ from the reverse divergence sum to zero. Consequently, the MSKL simplifies to a sum of logarithmic integrals

$$D_{MSKL}(p\|q) = \frac{1}{2} \int \left(\sqrt{p(x)} \log \frac{\sqrt{p(x)}}{\sqrt{q(x)}} + \sqrt{q(x)} \log \frac{\sqrt{q(x)}}{\sqrt{p(x)}} \right) dx$$

This simplification justifies the analysis in Section 5.

Remark 3.4 (Relationship to Hellinger Distance and Bhattacharyya Coefficient). The MSKL divergence operates on square-root transformed densities, placing it in the same mathematical family as several classical divergences. The squared Hellinger distance is defined as

$$H^2(p,q) = \frac{1}{2}\int \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 dx = 1 - \int \sqrt{p(x)q(x)}\, dx = 1 - BC(p,q)$$

where $BC(p,q) = \int \sqrt{p(x)q(x)}\, dx$ is the Bhattacharyya coefficient. The MSKL divergence differs from these measures by applying the logarithmic structure of KL divergence to the square-root transformed densities.

3.2.3. *Baseline KL Divergence Approximations.* Since no closed-form expression exists for KL divergence between GMMs, we also consider two established approximations as baselines for comparison [9, 14]. For GMMs $p(x) = \sum_{i=1}^{K} \alpha_i p_i(x)$ and $q(x) = \sum_{j=1}^{L} \beta_j q_j(x)$, where p_i and q_j denote Gaussian components with mixing weights α_i and β_j, the Weighted Average approximation leverages the convexity of KL divergence:

$$KL_{WA}(p\|q) = \sum_{i=1}^{K}\sum_{j=1}^{L} \alpha_i\beta_j\, D_{KL}(p_i\|q_j). \tag{3.5}$$

While computationally efficient, this provides a lower bound that can significantly underestimate the true divergence when mixture components are well-separated. The Matching-Based approximation assumes dominant contributions arise from the closest components:

$$KL_{MB}(p\|q) = \sum_{i=1}^{K} \alpha_i \min_{j=1,\ldots,L}\left[D_{KL}(p_i\|q_j) + \log\left(\frac{\alpha_i}{\beta_j}\right)\right]. \tag{3.6}$$

This provides an upper bound but becomes unreliable when components overlap.

The MSKL divergence offers several advantages over these existing approximations. First as shown in Section 5, MSKL admits explicit upper and lower bounds for GMMs, whereas standard KL divergence can be unbounded. Second the square-root transformation reduces sensitivity to regions where one density approaches zero, providing better numerical stability particularly when comparing noisy point cloud data.

4. Statistical Manifold Representation for Point Cloud Data

In this section, we establish a statistical manifold framework for point clouds dataset. Consider the dataset X with n points in $\mathbb{R}^m$, i.e $X = \{x_1, ..., x_n \mid x_i \in \mathbb{R}^m\}$, assume that all the points are distinct. Denote the set of all point clouds by Z.

4.1. **Gaussian Mixture Model.** We model the point cloud $X \subset \mathbb{R}^m$ as i.i.d. samples from a K-component Gaussian mixture with parameters $\Theta = \{(\mu_k, \Sigma_k, \pi_k)_{k=1}^{K}\}$, where $\mu_k \in \mathbb{R}^m$, $\Sigma_k \succ 0$, $\pi_k \geq 0$, and $\sum_{k=1}^{K} \pi_k = 1$. Each component has density

$$\mathcal{N}(x; \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^m |\Sigma|}} \exp\left(-\tfrac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right),$$

and the mixture is

$$p(x; \theta) = \sum_{k=1}^{K} \pi_k \, \mathcal{N}(x; \mu_k, \Sigma_k).$$

The density $p(\cdot; \theta)$ is the statistical representation of X [29].

4.2. **Manifold Structure for Point Clouds.** Consider the space Z of point clouds, where each point cloud is represented as $X = \{x_1, \ldots, x_n \mid x_i \in \mathbb{R}^m\}$. Point cloud X has a statistical representation $p(x; \theta)$ where the parameter θ varies over the parameter space $\Theta = \{\theta = (\mu_1, ..., \mu_K, \Sigma_1, ..., \Sigma_K, \pi_1, ..., \pi_K) \mid \mu_i \in \mathbb{R}^m, \Sigma_i$ is an $m \times m$ symmetric matrix, $\sum_{i=1}^{K} \pi_i = 1\}$. The set $S = \{p(x; \theta) \mid \theta \in \Theta\}$ representing the space Z of point clouds is the statistical model representing Z. Our aim is to provide a geometric structure, called the statistical manifold, to the space Z.

A critical requirement for this construction is the identifiability of the mixture model, which ensures that the mapping from parameters to probability densities is injective. The identifiability of finite mixtures of Gaussians is a classical result in statistics, established by Teicher [31] and Yakowitz and Spragins [37]. We restate this property below as a foundational block for our manifold structure.

Theorem 4.1. *Let $p(x; \theta_1) = \sum_{i=1}^{K} \alpha_i \mathcal{N}(x; \mu_i, \sigma_i^2)$ and $p(x; \theta_2) = \sum_{j=1}^{L} \beta_j \mathcal{N}(x; \nu_j, \tau_j^2)$ be two univariate Gaussian mixture models with K and L components respectively. Suppose that $p(x; \theta_1) = p(x; \theta_2)$ for almost all $x \in \mathbb{R}$. If the parameters in each component of a GMM are distinct and they are ordered lexicographically by variance and mean, then $K = L$ and for each i, $\alpha_i = \beta_i$, $\mu_i = \nu_i$, and $\sigma_i^2 = \tau_i^2$.*

Proof. The proof is classical and can be found in [31, 37].

□

In this paper, we work with the diagonal covariance matrices to reduce the computational complexity.

Theorem 4.2. *The parameter space $\Theta = \{\theta = (\mu_1, \ldots, \mu_K, \Sigma_1, \ldots, \Sigma_K, \alpha_1, \ldots, \alpha_K) \mid \mu_i \in \mathbb{R}^m$, Σ_i is an $m \times m$ diagonal covariance matrix, $\sum_{i=1}^{K} \alpha_i = 1\}$ for GMMs with K components of m-dimensional Gaussian is a topological manifold.*

Proof. For the probability density function

$$p(x; \theta) = \sum_{i=1}^{K} \alpha_i \mathcal{N}(x; \mu_i, \Sigma_i),$$

each μ_i is an m-dimensional vector. The $m \times m$ diagonal covariance matrix Σ_i is symmetric, thus having m distinct elements. The total number of mixing coefficients is K, and since $\sum_{i=1}^{K} \alpha_i = 1$, only $K - 1$ are independent. So the total number of independent parameters in Θ is $\dim(\Theta) = K(2m + 1) - 1$. Note that Θ is an open subset of $\mathbb{R}^{dim(\Theta)}$, so Θ is a topological space with standard

Euclidean topology. Also, around any point in Θ there exists a neighborhood that is homeomorphic to an open subset of $\mathbb{R}^{\dim(\Theta)}$. This satisfies the local Euclidean condition for a manifold. Therefore, the parameter space Θ of m-dimensional GMMs is a topological manifold of dimension $K(2m+1)-1$. □

Theorem 4.3. *The statistical model S representing the space Z of point clouds is a statistical manifold of dimension $K(2m+1)-1$.*

Proof. Consider the map $h : \Theta \to S$, defined by $h(\theta) = p(x; \theta)$ for $\theta \in \Theta$. Now to show that this map is injective. Consider two GMMs with parameter sets $\theta_1, \theta_2 \in \Theta$

$$p(x;\theta_1) = \sum_{i=1}^{K} \alpha_{i1}\mathcal{N}(x;\mu_{i1},\Sigma_{i1})$$

and

$$p(x;\theta_2) = \sum_{i=1}^{K} \alpha_{i2}\mathcal{N}(x;\mu_{i2},\Sigma_{i2})$$

where Σ_{i1} and Σ_{i2} are diagonal matrices.

Assume $p(x;\theta_1) = p(x;\theta_2)$ for all x. By the Cramér-Wold theorem [7], this equality holds if and only if their projections onto any direction $\ell \in \mathbb{R}^m$ are equal.

For any direction ℓ, the property of multivariate normal distributions ensures that if $X \sim \mathcal{N}(\mu, \Sigma)$, then

$$\ell^T X \sim \mathcal{N}(\ell^T \mu, \ell^T \Sigma \ell).$$

Therefore, projecting GMMs gives

$$\sum_{i=1}^{K} \alpha_{i1}\mathcal{N}(\ell^T x;\ell^T\mu_{i1},\ell^T\Sigma_{i1}\ell) = \sum_{i=1}^{K} \alpha_{i2}\mathcal{N}(\ell^T x;\ell^T\mu_{i2},\ell^T\Sigma_{i2}\ell).$$

These projected distributions are univariate GMMs, as each $\ell^T x$ is scalar-valued. By Theorem 1, the equality of these univariate GMMs for all directions ℓ implies

$$\alpha_{i1} = \alpha_{i2}, \quad \ell^T\mu_{i1} = \ell^T\mu_{i2}, \quad \ell^T\Sigma_{i1}\ell = \ell^T\Sigma_{i2}\ell$$

for all i and all ℓ. Since this holds for all projection directions ℓ, we conclude that $\theta_1 = \theta_2$, proving that h is injective. The parameter space Θ is a topological manifold and the mapping $h : \Theta \to S$ is injective hence the statistical model representing the space Z of point clouds can be viewed as a statistical manifold of dimension $dim(\Theta) = K(2m+1)-1$. □

5. MSKL Bounds

In this section, we establish upper and lower bounds for the MSKL divergence between GMMs. These bounds are the theoretical foundation that ensures MSKL is well-defined and numerically stable for the GMM comparison, addressing the unboundedness issues of existing KL approximations discussed in Section 3.2.3.

Theorem 5.1. *Let p, q be probability densities on $\mathbb{R}^m$. Then, the MSKL divergence satisfies the lower bound*

$$D_{\text{MSKL}}(p\|q) \geq S(p)\log S(p) + S(q)\log S(q) - (S(p)+S(q))\log A(p,q),$$

where $S(p) = \int_{\mathbb{R}^m} \sqrt{p(x)}\,dx$, $S(q) = \int_{\mathbb{R}^m} \sqrt{q(x)}\,dx$, and $A(p,q) = \int_{\mathbb{R}^m} (p(x)q(x))^{1/4}dx$. We assume $0 < S(p), S(q) < \infty$ and $0 < A(p,q) < \infty$.

Proof. Denote $r_p(x) = \sqrt{p(x)}$ and $r_q(x) = \sqrt{q(x)}$. These are positive measures that do not necessarily integrate to unity. We define their normalized profiles as $\alpha(x) = \frac{r_p(x)}{S(p)}$ and $\beta(x) = \frac{r_q(x)}{S(q)}$, such that $\int \alpha(x)dx = \int \beta(x)dx = 1$.

To rigorously handle unnormalized measures, we employ the Generalized Kullback-Leibler divergence (Definition 3.1)

$$D_{\text{Gen}}(u\|v) = \int \left(u(x)\log\frac{u(x)}{v(x)} - u(x) + v(x) \right) dx. \tag{5.1}$$

Applying this to r_p and r_q

$$\begin{aligned}
D_{\text{Gen}}(r_p\|r_q) &= \int \left(r_p \log\frac{r_p}{r_q} - r_p + r_q \right) dx && (5.2)\\
&= \int S(p)\alpha \log\left(\frac{S(p)\alpha}{S(q)\beta}\right) dx - S(p) + S(q) && (5.3)\\
&= S(p)\log\frac{S(p)}{S(q)} \underbrace{\int \alpha\, dx}_{1} + S(p) \underbrace{\int \alpha \log\frac{\alpha}{\beta}\, dx}_{D_{\text{KL}}(\alpha\|\beta)} - S(p) + S(q) && (5.4)\\
&= S(p)\log\frac{S(p)}{S(q)} + S(p)D_{\text{KL}}(\alpha\|\beta) - S(p) + S(q). && (5.5)
\end{aligned}$$

Similarly, for the reverse direction

$$D_{\text{Gen}}(r_q\|r_p) = S(q)\log\frac{S(q)}{S(p)} + S(q)D_{\text{KL}}(\beta\|\alpha) - S(q) + S(p). \tag{5.6}$$

The MSKL divergence is defined as the average of these two generalized divergences

$$D_{\text{MSKL}}(p\|q) = \frac{1}{2}\left[D_{\text{Gen}}(r_p\|r_q) + D_{\text{Gen}}(r_q\|r_p)\right]. \tag{5.7}$$

Substituting the expanded forms, the mass terms $(-S(p)+S(q))$ and $(-S(q)+S(p))$ sum to zero and vanish. We are left with

$$\begin{aligned}
D_{\text{MSKL}}(p\|q) = \frac{1}{2}&\left[S(p)\log\frac{S(p)}{S(q)} + S(q)\log\frac{S(q)}{S(p)}\right]\\
&+ \frac{1}{2}\left[S(p)D_{\text{KL}}(\alpha\|\beta) + S(q)D_{\text{KL}}(\beta\|\alpha)\right]. && (5.8)
\end{aligned}$$

Using the property $\log(x/y) = -\log(y/x)$, the logarithmic terms combine to

$$\frac{1}{2}(S(p)-S(q))\log\frac{S(p)}{S(q)}. \tag{5.9}$$

Next, we bound the standard KL terms using $D_{\mathrm{KL}}(u\|v) \geq -2\log\int\sqrt{u(x)v(x)}dx$

$$\int\sqrt{\alpha(x)\beta(x)}\,dx = \int\sqrt{\frac{\sqrt{p(x)}}{S(p)}\frac{\sqrt{q(x)}}{S(q)}}\,dx = \frac{\int(p(x)q(x))^{1/4}\,dx}{\sqrt{S(p)S(q)}} = \frac{A(p,q)}{\sqrt{S(p)S(q)}}. \tag{5.10}$$

Therefore:

$$S(p)D_{\mathrm{KL}}(\alpha\|\beta) \geq -2S(p)\log\left(\frac{A(p,q)}{\sqrt{S(p)S(q)}}\right), \tag{5.11}$$

$$S(q)D_{\mathrm{KL}}(\beta\|\alpha) \geq -2S(q)\log\left(\frac{A(p,q)}{\sqrt{S(p)S(q)}}\right). \tag{5.12}$$

Substituting these bounds back into 5.8 completes the proof. □

Theorem 5.2. *Let $p(x) = \sum_{i=1}^{K}\pi_i\phi_i(x)$ be a GMM with $\phi_i = \mathcal{N}(\mu_i,\Sigma_i)$, $\pi_i \geq 0$, $\sum_{i=1}^{K}\pi_i = 1$. Then $S(p) \leq \bar{S}(p)$, where $\bar{S}(p) = (8\pi)^{m/4}\sum_{i=1}^{K}\sqrt{\pi_i}\,|\Sigma_i|^{1/4}$.*

Proof. By concavity of square root,

$$\sqrt{\sum_{i=1}^{K}\pi_i\phi_i(x)} \leq \sum_{i=1}^{K}\sqrt{\pi_i}\sqrt{\phi_i(x)}. \tag{5.13}$$

Integrating both sides,

$$S(p) = \int_{\mathbb{R}^m}\sqrt{p(x)}\,dx \leq \sum_{i=1}^{K}\sqrt{\pi_i}\int_{\mathbb{R}^m}\sqrt{\phi_i(x)}\,dx. \tag{5.14}$$

For $\phi_i = \mathcal{N}(x;\mu_i,\Sigma_i)$,

$$\int_{\mathbb{R}^m}\sqrt{\mathcal{N}(x;\mu_i,\Sigma_i)}\,dx = \int_{\mathbb{R}^m}\left[\frac{1}{(2\pi)^{m/2}|\Sigma_i|^{1/2}}\exp\left(-\frac{1}{4}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)\right)\right]dx \tag{5.15}$$

$$= \frac{1}{(2\pi)^{m/4}|\Sigma_i|^{1/4}}\int_{\mathbb{R}^m}\exp\left(-\frac{1}{4}y^T\Sigma_i^{-1}y\right)dy \quad (y = x-\mu_i) \tag{5.16}$$

$$= \frac{1}{(2\pi)^{m/4}|\Sigma_i|^{1/4}}\cdot(4\pi)^{m/2}|\Sigma_i|^{1/2} = (8\pi)^{m/4}|\Sigma_i|^{1/4} \tag{5.17}$$

Substituting into (17) completes the proof. □

Theorem 5.3. *Let $p(x) = \sum_{i=1}^{K}\pi_i\phi_i(x)$ and $q(x) = \sum_{j=1}^{L}\rho_j\psi_j(x)$ be two GMMs on $\mathbb{R}^m$ with $\phi_i = \mathcal{N}(\mu_i,\Sigma_i)$, $\psi_j = \mathcal{N}(\nu_j,\Lambda_j)$, and mixing weights satisfying $\sum_i\pi_i = \sum_j\rho_j = 1$. Then,*

$$A(p,q) \leq \bar{A}(p,q).$$

where, $\bar{A}(p,q)$ denotes $\sum_{i=1}^{K}\sum_{j=1}^{L}\pi_i^{1/4}\rho_j^{1/4}I_{ij}$ where, $I_{ij} = \int_{\mathbb{R}^m}\phi_i(x)^{1/4}\psi_j(x)^{1/4}\,dx$ can be computed in closed form as

$$I_{ij} = (2\pi)^{m/4}\frac{|P_{ij}|^{-1/2}}{|\Sigma_i|^{1/8}|\Lambda_j|^{1/8}}\exp\left(-\tfrac{1}{2}Q_{ij}\right),$$

with $P_{ij} = \frac{1}{4}(\Sigma_i^{-1} + \Lambda_j^{-1})$, $h_{ij} = \frac{1}{4}(\Sigma_i^{-1}\mu_i + \Lambda_j^{-1}\nu_j)$, $R_{ij} = \frac{1}{8}(\mu_i^\top \Sigma_i^{-1}\mu_i + \nu_j^\top \Lambda_j^{-1}\nu_j)$, *and* $Q_{ij} = 2R_{ij} - h_{ij}^\top P_{ij}^{-1} h_{ij}$.

Proof. Jensen's inequality gives

$$p(x)^{1/4} \le \sum_i \pi_i^{1/4}\phi_i(x)^{1/4}, \quad q(x)^{1/4} \le \sum_j \rho_j^{1/4}\psi_j(x)^{1/4}.$$

Therefore,

$$A(p,q) = \int (p(x)q(x))^{1/4} dx \le \sum_{i,j} \pi_i^{1/4}\rho_j^{1/4} \int \phi_i(x)^{1/4}\psi_j(x)^{1/4} dx = \sum_{i,j} \pi_i^{1/4}\rho_j^{1/4} I_{ij}.$$

To compute I_{ij}, we have

$$\phi_i(x)^{1/4}\psi_j(x)^{1/4} = \frac{1}{(2\pi)^{m/4}|\Sigma_i|^{1/8}|\Lambda_j|^{1/8}} \times \exp\Big(-\tfrac{1}{8}(x-\mu_i)^\top \Sigma_i^{-1}(x-\mu_i) - \tfrac{1}{8}(x-\nu_j)^\top \Lambda_j^{-1}(x-\nu_j)\Big).$$

Expanding the quadratic forms and collecting terms, this equals

$$\frac{1}{(2\pi)^{m/4}|\Sigma_i|^{1/8}|\Lambda_j|^{1/8}} \exp\left(-\frac{1}{2}x^\top P_{ij} x + x^\top h_{ij} - R_{ij}\right)$$

where $P_{ij} = \frac{1}{4}(\Sigma_i^{-1} + \Lambda_j^{-1})$, $h_{ij} = \frac{1}{4}(\Sigma_i^{-1}\mu_i + \Lambda_j^{-1}\nu_j)$, and $R_{ij} = \frac{1}{8}(\mu_i^\top \Sigma_i^{-1}\mu_i + \nu_j^\top \Lambda_j^{-1}\nu_j)$. We have $-\frac{1}{2}x^\top P_{ij}x + x^\top h_{ij} = -\frac{1}{2}(x - P_{ij}^{-1}h_{ij})^\top P_{ij}(x - P_{ij}^{-1}h_{ij}) + \frac{1}{2}h_{ij}^\top P_{ij}^{-1}h_{ij}$.

Substituting this back and integrating we get

$$\begin{aligned} I_{ij} &= \int_{\mathbb{R}^m} \phi_i(x)^{1/4}\psi_j(x)^{1/4} dx \\ &= \frac{1}{(2\pi)^{m/4}|\Sigma_i|^{1/8}|\Lambda_j|^{1/8}} \exp\Big(\tfrac{1}{2}h_{ij}^\top P_{ij}^{-1}h_{ij} - R_{ij}\Big) \\ &\quad \int_{\mathbb{R}^m} \exp\Big(-\tfrac{1}{2}(x - P_{ij}^{-1}h_{ij})^\top P_{ij}(x - P_{ij}^{-1}h_{ij})\Big)\, dx. \end{aligned}$$

The Gaussian integral evaluates to $(2\pi)^{m/2}|P_{ij}|^{-1/2}$. Then,

$$I_{ij} = \frac{1}{(2\pi)^{m/4}|\Sigma_i|^{1/8}|\Lambda_j|^{1/8}} \exp\left(\frac{1}{2}h_{ij}^\top P_{ij}^{-1}h_{ij} - R_{ij}\right) \cdot (2\pi)^{m/2}|P_{ij}|^{-1/2}$$

which gives

$$I_{ij} = (2\pi)^{m/4} \frac{|P_{ij}|^{-1/2}}{|\Sigma_i|^{1/8}|\Lambda_j|^{1/8}} \exp\left(\frac{1}{2}h_{ij}^\top P_{ij}^{-1}h_{ij} - R_{ij}\right).$$

Since $(2\pi)^{m/4} = 2^{m/4}\pi^{m/4}$ and $|P_{ij}| = 4^{-m}|\Sigma_i^{-1} + \Lambda_j^{-1}|$, we have $|P_{ij}|^{-1/2} = 2^m|\Sigma_i^{-1} + \Lambda_j^{-1}|^{-1/2}$. Let $Q_{ij} = 2R_{ij} - h_{ij}^\top P_{ij}^{-1}h_{ij}$. This gives the stated formula for I_{ij}. □

Theorem 5.4. *Let p, q be GMMs. Let $\bar{S}(p), \bar{S}(q)$ be the bounds from Theorem 5.2 and let $\bar{A}(p,q)$ be the bound from Theorem 5.3. Then*

$$D_{\mathrm{MSKL}}(p\|q) \ \ge\ \max\left\{0,\ S(p)^* \log \frac{S(p)^*}{\bar{A}(p,q)} + S(q)^* \log \frac{S(q)^*}{\bar{A}(p,q)}\right\},$$

where

$$S(p)^* := \min\left\{\bar{S}(p), \frac{\bar{A}(p,q)}{e}\right\}, \qquad S(q)^* := \min\left\{\bar{S}(q), \frac{\bar{A}(p,q)}{e}\right\}.$$

Proof. By Theorem 5.1,

$$D_{\mathrm{MSKL}}(p\|q) \geq S(p)\log S(p) + S(q)\log S(q) - (S(p)+S(q))\log A(p,q). \tag{5.18}$$

Let us denote the right-hand side of (5.18) by $F(S(p), S(q), A)$. We have

$$0 < S(p) \leq \bar{S}(p), \qquad 0 < S(q) \leq \bar{S}(q), \qquad 0 < A(p,q) \leq \bar{A}(p,q).$$

For fixed positive $S(p), S(q)$,

$$\frac{\partial}{\partial A}F(S(p), S(q), A) = -\frac{S(p)+S(q)}{A} < 0 \qquad (A > 0).$$

Hence $F(S(p), S(q), A) \geq F(S(p), S(q), a)$ for all $A \in (0, a]$, where $a = \bar{A}(p,q)$. Therefore, from (5.18),

$$D_{\mathrm{MSKL}}(p\|q) \;\geq\; F(S(p), S(q), a) = \big[S(p)\log S(p) - S(p)\log a\big] + \big[S(q)\log S(q) - S(q)\log a\big]. \tag{5.19}$$

Define, for fixed $a > 0$,

$$g(S) := S\log S - S\log a \qquad (S > 0).$$

Then $g'(S) = \log S + 1 - \log a$ and $g''(S) = 1/S > 0$, so g is strictly convex on $(0, \infty)$ with unique stationary point at $g'(S) = 0$, i.e.,

$$\log\left(\tfrac{S}{a}\right) = -1 \iff S = \tfrac{a}{e}.$$

Consequently, $S = \frac{a}{e}$ is the global minimizer of g on $(0, \infty)$.

Now for the constraint $0 < S \leq \bar{S}$:

- If $\bar{S} \geq a/e$, then a/e is admissible and for all $S \in (0, \bar{S}]$, $g(S) \geq g(a/e)$.
- If $\bar{S} < a/e$, then for $S \in (0, \bar{S}]$ we have $S/a < 1/e$, hence $g'(S) = \log(S/a) + 1 < 0$, so g is strictly decreasing on $(0, \bar{S}]$ and $g(S) \geq g(\bar{S})$.

In both cases,

$$g(S) \;\geq\; g\big(\min\{\bar{S}, \tfrac{a}{e}\}\big).$$

Applying this with $(S, \bar{S}) = (S(p), \bar{S}(p))$ and with $(S, \bar{S}) = (S(q), \bar{S}(q))$ gives

$$S(p)\log S(p) - S(p)\log a \;\geq\; S(p)^*\log S(p)^* - S(p)^*\log a,$$

$$S(q)\log S(q) - S(q)\log a \;\geq\; S(q)^*\log S(q)^* - S(q)^*\log a.$$

Combining these two inequalities with (5.19) gives

$$D_{\mathrm{MSKL}}(p\|q) \;\geq\; S(p)^*\log\frac{S(p)^*}{a} + S(q)^*\log\frac{S(q)^*}{a} = S(p)^*\log\frac{S(p)^*}{\bar{A}(p,q)} + S(q)^*\log\frac{S(q)^*}{\bar{A}(p,q)}.$$

Since $D_{\mathrm{MSKL}}(p\|q) \geq 0$ always, we may write

$$D_{\mathrm{MSKL}}(p\|q) \;\geq\; \max\left\{0,\; S(p)^*\log\frac{S(p)^*}{\bar{A}(p,q)} + S(q)^*\log\frac{S(q)^*}{\bar{A}(p,q)}\right\}.$$

□

5.1. **Upper Bound via the Bhattacharyya Coefficient.**

Theorem 5.5. *Let p and q be GMMs. Then for any $\epsilon > 0$, there exists a real number $T(p,q)$ depending on the mixture parameters such that*

$$D_{\mathrm{MSKL}}(p\|q) \;\le\; -\frac{\bar{S}(p)+\bar{S}(q)}{2}\,\log\big(\max\{T(p,q),\epsilon\}\big).$$

Proof. Denote $r_p = \sqrt{p}$ and $r_q = \sqrt{q}$. These are positive measures with $S(p) = \int r_p\,dx$ and $S(q) = \int r_q\,dx$. From Section 3.2

$$D_{\mathrm{MSKL}}(p\|q) = \frac{1}{2}\left[\int r_p \log\frac{r_p}{r_q}\,dx + \int r_q \log\frac{r_q}{r_p}\,dx\right].$$

By Jensen's inequality for the concave function log,

$$\frac{1}{S(p)}\int r_p \log r_p\,dx \le \log\left(\frac{\int r_p^2\,dx}{S(p)}\right) = -\log S(p),$$

$$\frac{1}{S(p)}\int r_p \log r_q\,dx \le \log\left(\frac{\int r_p r_q\,dx}{S(p)}\right).$$

Let $\mathrm{BC}(p,q)$ denote the Bhattacharyya coefficient $\int \sqrt{p(x)q(x)}\,dx = \int r_p(x) r_q(x)\,dx$. Then

$$\int r_p \log\frac{r_p}{r_q}\,dx = \int r_p \log r_p\,dx - \int r_p \log r_q\,dx \le -S(p)\log \mathrm{BC}(p,q).$$

Similarly, $\int r_q \log\frac{r_q}{r_p}\,dx \le -S(q)\log\mathrm{BC}(p,q)$, hence

$$D_{\mathrm{MSKL}}(p\|q) \le -\frac{S(p)+S(q)}{2}\,\log \mathrm{BC}(p,q). \tag{5.20}$$

For Gaussian components $\phi_i = \mathcal{N}(\mu_i,\Sigma_i)$ and $\psi_j = \mathcal{N}(\nu_j,\Lambda_j)$, the Bhattacharyya coefficient has the closed form

$$\mathrm{BC}(\phi_i,\psi_j) = \frac{|\Sigma_i|^{1/4}|\Lambda_j|^{1/4}}{\left|\frac{\Sigma_i+\Lambda_j}{2}\right|^{1/2}} \exp\left(-\frac{1}{8}(\mu_i-\nu_j)^\top\left(\frac{\Sigma_i+\Lambda_j}{2}\right)^{-1}(\mu_i-\nu_j)\right).$$

By the Cauchy–Schwarz inequality,

$$\sum_{i=1}^{K}\sum_{j=1}^{L}\sqrt{\pi_i\phi_i(x)\,\rho_j\psi_j(x)} \le \sqrt{\left(\sum_{i=1}^{K}\pi_i\phi_i(x)\right)\left(\sum_{j=1}^{L}\rho_j\psi_j(x)\right)} = \sqrt{p(x)q(x)}.$$

Integrating both sides yields

$$\mathrm{BC}(p,q) \;\ge\; T(p,q) := \sum_{i=1}^{K}\sum_{j=1}^{L}\pi_i\rho_j\,\mathrm{BC}(\phi_i,\psi_j),$$

which is explicit and computable from the mixture parameters.

Substituting into (5.20) and applying the upper bound $S(p) \le \bar{S}(p)$ from Theorem 5.2,

$$D_{\mathrm{MSKL}}(p\|q) \;\le\; -\frac{\bar{S}(p)+\bar{S}(q)}{2}\,\log T(p,q).$$

The ϵ regularization ensures numerical stability when $T(p,q)$ is very small. $\square$

Combining the results from Theorems 5.1–5.5, we obtain the following bound for the MSKL divergence:

$$\max\left(0,\ S(p)^* \log \frac{S(p)^*}{\bar{A}(p,q)} + S(q)^* \log \frac{S(q)^*}{\bar{A}(p,q)}\right) \leq D_{\mathrm{MSKL}}(p\|q) \leq -\frac{\bar{S}(p)+\bar{S}(q)}{2} \log \max\{T(p,q), \epsilon\}. \tag{5.21}$$

These bounds are explicitly computable from the GMM parameters and ensure $0 \leq D_{\mathrm{MSKL}}(p\|q) < \infty$ for all GMMs.

While these results provide theoretical guarantees, in practice the exact computation of D_{MSKL} is often intractable. Instead, empirical evaluation requires discretized approximations. In our implementation, we employ a grid-based discretization and normalize the square-root densities for numerical stability. Specifically, we compute

$$D_{\mathrm{MSKL}}^{\mathrm{discrete}}(p\|q) = \frac{1}{2}\sum_{i=1}^{N}\left[\tilde{p}_i \log \frac{\tilde{p}_i}{\tilde{q}_i} + \tilde{q}_i \log \frac{\tilde{q}_i}{\tilde{p}_i}\right], \tag{5.22}$$

where

$$\tilde{p}_i = \frac{\sqrt{p(g_i)+\epsilon}}{\sum_j \sqrt{p(g_j)+\epsilon}}, \quad \tilde{q}_i = \frac{\sqrt{q(g_i)+\epsilon}}{\sum_j \sqrt{q(g_j)+\epsilon}}$$

are the normalized square-root density values at grid points $\{g_i\}_{i=1}^N$, with $\epsilon = 10^{-10}$. This normalization ensures numerical stability and scale invariance while preserving the symmetric divergence structure established in the theoretical framework. In the following section, we describe the datasets and preprocessing steps used to validate our approach.

6. Data Overview and Preprocessing

This section discuss an overview of the datasets used in the experimental analysis, namely the MPI FAUST dataset and the G-PCD dataset.

6.1. **Human and Animal Datasets.** For the human dataset, we employ the MPI-FAUST dataset [25] (see Figure 2), which provides high-resolution 3D scans of human figures captured across diverse postural configurations. These point clouds exhibit complex surface topology making them ideal for evaluating shape similarity measures under non-rigid transformations.

For the animal dataset, we utilize the G-PCD dataset [8] (see Figure 3), which contains 3D scans of various animal species with various morphological structures. These point clouds feature rich local geometric variations, including sharp features, smooth surfaces, and complex curvature patterns.

7. Experimental Methodology

This section describes the complete computational pipeline for measuring shape similarity using the statistical manifold of GMMs. Our approach transforms 3D

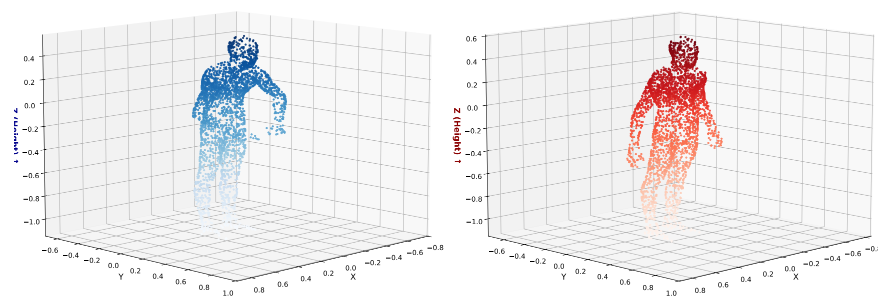


FIGURE 2. MPI-FAUST human dataset samples showing two distinct poses of a Human body scans.

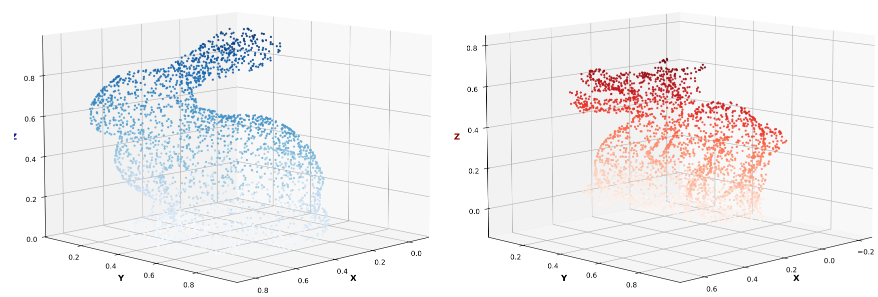


FIGURE 3. G-PCD animal dataset samples illustrating morphological diversity. Left: Rabbit Point Cloud 1. Right: Dragon Point Cloud 2.

point clouds into probability distributions on a low-dimensional statistical manifold, where divergence measures quantify shape dissimilarity. The methodology consists of five main stages: preprocessing, geometric feature extraction, manifold embedding, probabilistic modeling, and divergence and metric computation (see Figure 5).

7.1. **Data Preprocessing.** Before feature extraction and manifold-based modeling, each point cloud undergoes a preprocessing step which is designed to establish geometric consistency and ensure compatibility with the computational framework. The raw point clouds often contain several thousand points with non-uniform sampling density across their surfaces. To obtain a uniform resolution while preserving the overall geometric structure, we sample from each point cloud using the Farthest Point Sampling (FPS) [24].

FPS progressively selects points that maximize their minimal distance to the current sample set, thereby ensuring broad coverage of the underlying shape.

Starting from an initial point chosen uniformly at random, the algorithm iteratively constructs the subset by selecting

$$p_i = \arg\max_{p \in X \setminus U} \min_{j<i} \|p - p_j\|,$$

which guarantees that each new point is maximally separated from the previously selected ones. Using this approach each point cloud is downsampled to 1024 points. To ensure consistency across all scans, each point cloud is then normalized independently. This removes global translation and scale differences, placing all shapes into a comparable coordinate frame without altering their intrinsic geometry.

7.2. **Geometric Feature Extraction.** Following preprocessing, we characterize the local geometry of each point cloud by computing pointwise descriptors derived from neighborhood structure. For every point, we identify its $k = 20$ nearest neighbors and compute a set of geometric features based on eigenvalue decomposition of the local covariance matrix.

For each point $\mathbf{x}_i$, let $\mathcal{N}_i = \{\mathbf{x}_{j_1}, \dots, \mathbf{x}_{j_k}\}$ denote its k nearest neighbors. The covariance matrix of the centered neighborhood is given by

$$\mathbf{C}_i = \frac{1}{k} \sum_{\mathbf{x}_j \in \mathcal{N}_i} (\mathbf{x}_j - \bar{\mathbf{x}}_i)(\mathbf{x}_j - \bar{\mathbf{x}}_i)^\top, \tag{7.1}$$

where $\bar{\mathbf{x}}_i$ is the neighborhood centroid. The eigendecomposition yields eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ and corresponding eigenvectors. These eigenvalues characterize the local geometric structure.

From this eigenstructure, we extract eleven geometric features $\mathbf{f}_i \in \mathbb{R}^{11}$. Let $\Sigma_\lambda = \sum_{j=1}^{3} \lambda_j$ denote the sum of eigenvalues. The features include: eigenvalue sum (Σ_λ), omnivariance ($(\prod_{j=1}^{3} \lambda_j)^{1/3}$), and eigenentropy ($-\sum_{j=1}^{3} p_j \ln p_j$ where $p_j = \lambda_j/\Sigma_\lambda$). We compute three shape factors normalized by the eigenvalue sum: anisotropy ($(\lambda_1 - \lambda_3)/\Sigma_\lambda$), planarity ($(\lambda_2 - \lambda_3)/\Sigma_\lambda$), and linearity ($(\lambda_1 - \lambda_2)/\Sigma_\lambda$). Additionally, we include the three normalized eigenvalues (λ_j/Σ_λ for $j = 1, 2, 3$, where the third component represents surface variation), sphericity (λ_3/λ_1), and verticality ($1 - |v_z|$, where v_z is the z-component of the eigenvector corresponding to λ_3) [10].

7.3. **Manifold Embedding via Isomap.** Following preprocessing and geometric feature extraction, each point is represented jointly by its normalized coordinates and local descriptors. For a point in a cloud, we denote by $\mathbf{x}_i \in \mathbb{R}^3$ its normalized 3D position and by $\mathbf{f}_i \in \mathbb{R}^{11}$ the corresponding geometric feature vector. These are concatenated into a single augmented descriptor

$$\mathbf{y}_i = \begin{bmatrix} \mathbf{x}_i \\ \mathbf{f}_i \end{bmatrix} \in \mathbb{R}^{14}, \tag{7.2}$$

so that the point cloud is represented as a set of 14-dimensional vectors. After constructing these descriptors for both point clouds, all points are stacked into a single matrix $\mathbf{Y} \in \mathbb{R}^{(n_1+n_2)\times 14}$, which serves as the input to the manifold learning stage.

We use Isomap [32] to embed this high-dimensional augmented representation into a low-dimensional latent space. Isomap preserves geodesic distances on the underlying manifold by approximating them via shortest paths in a k-nearest neighbor graph and then applying multidimensional scaling. In this setting, the embedding map

$$\mathcal{F} : \mathbb{R}^{14} \to \mathbb{R}^2 \tag{7.3}$$

is learned from $\mathbf{Y}$, producing two-dimensional latent coordinates $\mathbf{z}_i = \mathcal{F}(\mathbf{y}_i)$ for each point.

A key design choice is to embed both point clouds jointly rather than separately. If the two preprocessed and feature-augmented clouds contain $\{\mathbf{y}_i^{(1)}\}_{i=1}^{n_1}$ and $\{\mathbf{y}_j^{(2)}\}_{j=1}^{n_2}$, respectively, we form the combined dataset

$$\mathcal{Y}_{\text{joint}} = \{\mathbf{y}_i^{(1)}\}_{i=1}^{n_1} \cup \{\mathbf{y}_j^{(2)}\}_{j=1}^{n_2}, \tag{7.4}$$

and apply Isomap once to $\mathcal{Y}_{\text{joint}}$. The resulting latent coordinates

$$\{\mathbf{z}_i^{(1)}\}_{i=1}^{n_1}, \qquad \{\mathbf{z}_j^{(2)}\}_{j=1}^{n_2} \subset \mathbb{R}^2$$

thus lie in a common coordinate system, this ensure that all the subsequent computations are performed on the same latent space.

7.3.1. *Empirical Validation of Two-Dimensional Embedding.* The choice of a two-dimensional latent space requires justification, as reducing from 14 dimensions to 2 involves significant dimensionality reduction. We provide both theoretical rationale and empirical validation.
Theoretical Justification: Human body surfaces are intrinsically two-dimensional, that is, they are curved surfaces embedded in three-dimensional space. Isomap is guaranteed to recover the true underlying manifold dimension when data lies on or near a low-dimensional manifold [32]. Although the 14-dimensional augmented representation provides rich characterization, the underlying surface structure remains two-dimensional.
Empirical Validation: We verify that the two-dimensional embedding preserves geometric structure by measuring how well distances are preserved. Specifically, we compute the coefficient of determination R^2 between geodesic distances in the original 14D space and Euclidean distances in the 2D embedded space. Higher R^2 values indicate better distance preservation. Method......

The validation protocol consists of seven steps: (1) Apply Farthest Point Sampling to obtain 1024 representative points from the original mesh, (2) Normalize to unit sphere, (3) Extract 11-dimensional geometric features for each point, (4) Construct the 14-dimensional augmented representation, (5) Approximate geodesic distances via shortest paths in a k-nearest neighbor graph ($k = 10$) on the 14D space, (6) Apply Isomap to embed into d dimensions for $d \in \{2, 3, 4, 5\}$, (7) Compute R^2 by comparing distances in the original and embedded spaces.

Table 1 and Figure 4 present results for a representative MPI-FAUST scan. The two-dimensional embedding achieves $R^2 = 0.9675$, preserving 96.75% of the distance structure. Figure 4(c) visualizes this strong correlation. Importantly, increasing to 3D yields only 0.5% improvement, while 5D provides just 1.9% improvement over 2D (Figure 4(d)). This marginal gain does not justify the

TABLE 1. Distance preservation for different embedding dimensions. The 2D embedding preserves 96.75% of the distance structure, exceeding the 95% threshold commonly used for dimensionality reduction.

Embedding Dim.	R^2	**Residual Variance**	**% Preserved**
2D	**0.9675**	**0.0325**	**96.75%**
$3D$	0.9725	0.0275	97.25%
$4D$	0.9796	0.0204	97.96%
$5D$	0.9867	0.0133	98.67%

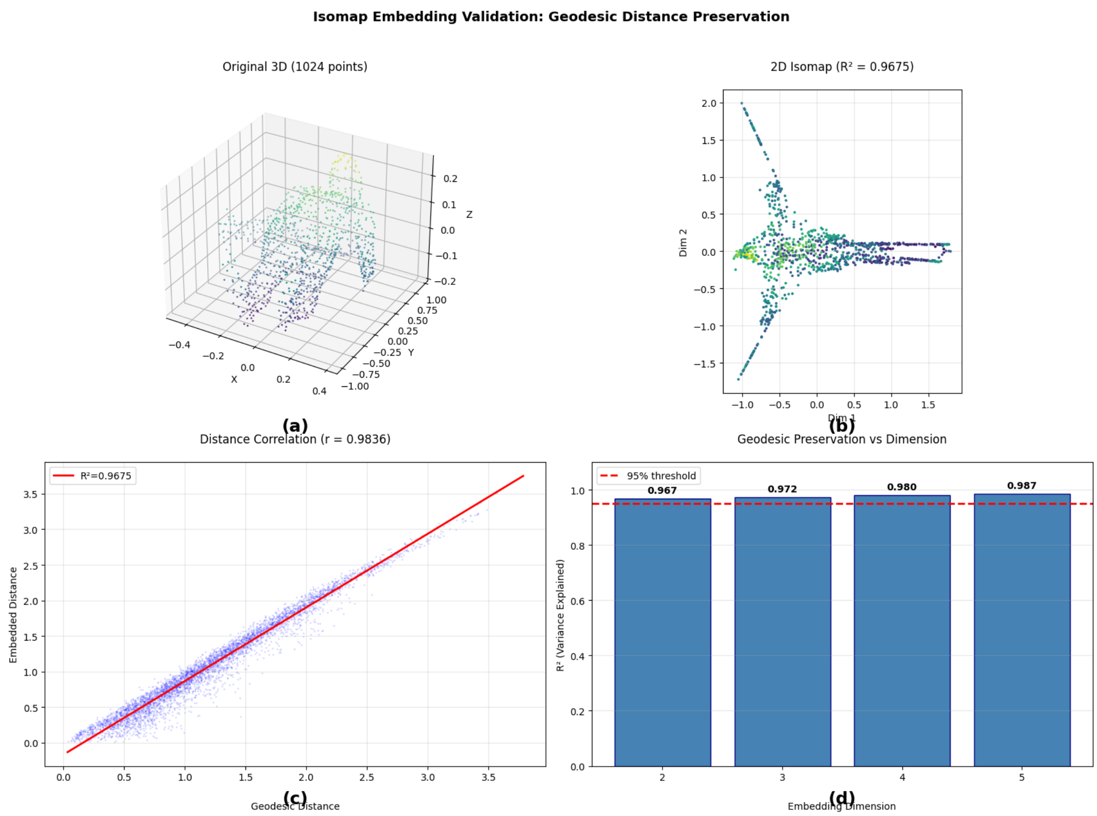


FIGURE 4. Empirical validation of 2D Isomap embedding (MPI-FAUST dataset, 1024 points). (a) Original 3D point cloud after sampling. (b) Two-dimensional embedding preserving surface structure. (c) Distance correlation plot showing strong linear relationship ($R^2 = 0.9675$) between geodesic and embedded distances. (d) Comparison across dimensions: all exceed 95% threshold (dashed line), with minimal improvement beyond 2D.

increased computational cost and model complexity, validating our choice of 2D embedding.

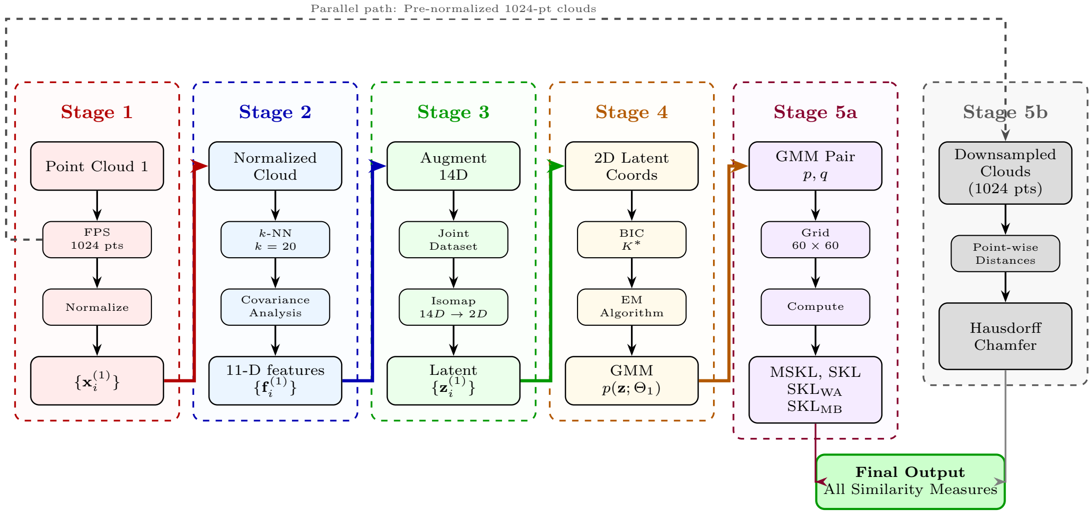


FIGURE 5. Five-stage computational pipeline for point cloud shape analysis on the statistical manifold of GMMs.

7.4. **Statistical Manifold Modeling.** With the two point clouds embedded in a common two-dimensional latent space, we model each shape using a Gaussian Mixture Model (GMM). The theoretical formulation and statistical manifold structure of GMMs are established in Section 3.1; here we describe the practical implementation.

To determine the number of mixture components K, we evaluate the Bayesian Information Criterion (BIC) for $K \in \{1, 2, \ldots, 10\}$ and select

$$K^* = \arg\min_K \mathrm{BIC}(K). \tag{7.5}$$

The GMM parameters $\Theta = \{(\boldsymbol{\mu}_k, \Sigma_k, \pi_k)\}_{k=1}^{K^*}$ are then estimated using the EM algorithm. For computational efficiency, we employ diagonal covariance matrices

$$\Sigma_k = \mathrm{diag}(\sigma_{k,1}^2,\ \sigma_{k,2}^2), \tag{7.6}$$

where $\sigma_{k,d}^2$ denotes the variance for component k in dimension $d \in \{1, 2\}$. matching the two-dimensional latent representation. A regularization term of 10^{-6} is added to each diagonal element for numerical stability, and EM is terminated after at most 200 iterations. The resulting GMMs provide smooth probabilistic representations $p(\mathbf{z}; \Theta_1)$ and $q(\mathbf{z}; \Theta_2)$ for the two point clouds on the latent manifold, which are used directly for divergence computation.

7.5. **Divergence Computation.** The divergence measures are evaluated on the GMMs defined over the 2D latent space. We approximate them via a regular grid discretization. For a given pair of latent embeddings, first compute the bound range $[\mathbf{z}_{\min}, \mathbf{z}_{\max}]$ enclosing all points, expand it by a 5% margin in each direction,

$$\mathbf{z}_{\min} \leftarrow \mathbf{z}_{\min} - 0.05(\mathbf{z}_{\max} - \mathbf{z}_{\min}), \qquad \mathbf{z}_{\max} \leftarrow \mathbf{z}_{\max} + 0.05(\mathbf{z}_{\max} - \mathbf{z}_{\min}),$$

and then construct a uniform grid of resolution 60×60 over this domain. This yields $N = 3600$ evaluation points $\{\mathbf{g}_i\}_{i=1}^{N}$. The corresponding cell area ΔA is used when required for mass-normalization.

1. Modified Symmetric KL Divergence. We approximate the MSKL divergence $D_{\text{MSKL}}(p\|q)$ by evaluating the GMM densities $p(\mathbf{g}_i)$ and $q(\mathbf{g}_i)$ at all grid points, applying the square-root transformation, and normalizing to obtain discrete probability distributions $\{\tilde{p}_i\}_{i=1}^N$ and $\{\tilde{q}_i\}_{i=1}^N$. The discrete MSKL is then computed as

$$D_{\text{MSKL}}^{\text{disc}}(p\|q) = \frac{1}{2}\sum_{i=1}^{N}\left[\tilde{p}_i \log\frac{\tilde{p}_i}{\tilde{q}_i} + \tilde{q}_i \log\frac{\tilde{q}_i}{\tilde{p}_i}\right], \tag{7.7}$$

with regularization $\epsilon = 10^{-8}$ added before normalization to ensure numerical stability.

2. Baseline Distances. For comparison, Hausdorff and Chamfer distances are computed on the original downsampled point clouds, using the definitions from Section 2. These classical geometric distances provide complementary baselines that operate directly on Euclidean coordinates.

3. Baseline KL Divergences Between GMMs. We also evaluate the symmetric KL divergence (SKL) along with the symmetric weighted Average (SKL_{WA}) and symmetric matching-Based (SKL_{MB}) approximations defined in Section 3.2.3, all computed on the same latent-space GMMs.

8. Experimental Results

The proposed geometric method is applied to two different cases and it is compared with other traditional measures. Each case demonstrates the effectiveness of the proposed method for the point cloud comparison.

8.1. **Human Pose Discrimination.**

We analyze four scans from a single subject in the MPI-FAUST dataset, representing various postures in which three similar standing poses with minor arm variations (arms down, hands on hips, arms extended) and one walking pose. Table 2 presents the pairwise similarity measures for all combinations.

MSKL divergence demonstrates stable, monotonic behavior with values ranging from 0.0004 (nearly identical standing poses) to 0.1105 (standing versus walking), providing clear discrimination that reflects the postural variation. In contrast, Hausdorff distance exhibits extreme sensitivity also showing poor discrimination for similar poses (0.20–0.26 for most standing comparisons). Chamfer distance offers better stability (0.034–0.213).

The KL-based approximations show major problems. SKL_{WA} ranges from 1.5–4.9 that fails to distinguish different poses well, while SKL_{MB} shows unstable behavior with values changing from 0.0014 to 1.627 for similarly posed standing configurations. The SKL shows reasonable ability to distinguish poses 0.0012–0.2158 but has no guaranteed limits for numerical stability and boundedness.

8.2. **Animal Data Comparison.**

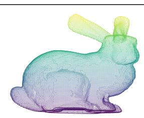
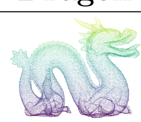
To evaluate divergence behavior on shapes with fundamentally different global topologies, we compare the Rabbit and dragon models from the G-PCD dataset. Table 3 presents the results.

MSKL yields 0.2100 for these morphologically distinct animals. Among manifold-based measures, MSKL demonstrates bounded numerical behavior: while SKL_{WA} produces an extreme value of 28.139 MSKL remains within a stable range due to

Table 2. Pairwise similarity measures for MPI-FAUST human body scans (same subject, different poses). Lower values indicate greater similarity.

Comparison	MSKL	SKL	SKL_{WA}	SKL_{MB}	Hausdorff	Chamfer
 	0.0333	0.1023	3.056	0.6835	0.2616	0.0571
 	0.0004	0.0012	2.531	0.0014	0.2000	0.0340
 	0.1065	0.2158	4.901	1.627	0.2616	0.0517
 	0.1105	0.1778	1.502	0.4361	0.8803	0.2127
 	0.0094	0.0268	2.017	0.0766	0.4208	0.1207

Table 3. Similarity measures for cross-category comparison (bunny vs dragon from G-PCD dataset).

Bunny	Dragon	MSKL	SKL	SKL_{WA}	SKL_{MB}	Hausdorff	Chamfer
		0.2100	0.4207	28.139	1.3163	0.4569	0.1411

its theoretical upper and lower bounds. SKL produces 0.4207 and SKL_{MB} yields 1.3163, both operating without guaranteed bounds. Classical distances Hausdorff (0.4569) and Chamfer (0.1411) provide point-wise proximity measures but lack the statistical manifold structure that captures distributional shape properties.

9. Conclusion and future work

In this paper, the point clouds are interpreted as the samples from the underlying probability distribution-GMM. Then, the statistical manifold structure is given to the space of point clouds which enables to employ geometric invariants like length, angle, divergence measure, connection and curvature for the analysis of the point clouds. In the present work, we use only divergence measures for the comparison of manifolds of point clouds. Modified Symmetric KL divergence is used for comparison of manifolds of point clouds. The advantages of the proposed method are (i) it offers a mathematically rigorous framework based on information geometry, (ii) it is a comprehensive method that can be used for different types of datasets and (iii) the method's balanced sensitivity and robustness to variations in geometry make it suitable for many practical applications.

Another key theoretical contribution of this work is the derivation of computable upper and lower bounds for the MSKL divergence. These bounds make the divergence theoretically well-posed and allow for practical computation without sacrificing mathematical rigor.

The proposed geometric method consistently exhibits a balanced sensitivity to shape variations and geometry in all the different types of datasets showing its ability to capture the geometry of complex datasets. In the context of 3D shape analysis, the proposed geometric method is effective in differentiating topologically similar shapes, such as human body scans in different postures. Thus, the proposed method provides a new and effective way to compare the manifold of point clouds which could be very useful in the areas like computer vision, 3D modeling, pattern recognition, Biomedical research etc.

The current framework has certain limitations that require future investigation. First, the computational complexity of the proposed framework increases significantly with point cloud size and the number of mixture components, particularly during the EM algorithm and divergence computation on high-resolution grids. While the use of diagonal covariance matrices reduces computational cost, it may not fully capture complex local geometric correlations. Second, the choice of dimensionality for the Isomap embedding is currently fixed at two dimensions for computational efficiency, though higher-dimensional latent representations might preserve more intricate geometric structure for certain shape categories.

The statistical manifold framework can be extended to other geometric invariants beyond divergence measures, including the Riemannian metric, geodesic distances on the manifold, and sectional curvature computations etc. These geometric invariants could provide shape descriptors that capture different aspects of shape similarity.

In conclusion, this work establishes an information-geometric framework for point cloud analysis that combines mathematical framework with practical effectiveness. By representing point clouds as points on a statistical manifold of GMMs and introducing the bounded MSKL divergence, we provide a stable and discriminative measure for shape comparison that outperforms existing approximations and baseline metrics. The framework opens new possibilities for using information geometry in 3D shape analysis, with potential applications spanning computer vision, medical imaging, robotics.

Data availability. The study utilized several publicly available datasets for the analysis of point clouds. The MPI FAUST dataset containing human body shapes can be accessed through https://faust-leaderboard.is.tuebingen.mpg.de. For analyzing animal shapes, we used the EPFL Geometry Point Cloud Dataset (G-PCD) which is available at https://www.epfl.ch/labs/mmspg/downloads/geometry-point-cloud-dataset/.

Code availability. The code used in this study will be made available upon reasonable request.

Author contribution. Both authors contributed equally to this work.

References

1. Akagic, A., Krivić, S., Dizdar, H., & Velagić, J. (2022). Computer Vision with 3D Point Cloud Data: Methods, Datasets and Challenges. In *Proc. 2022 XXVIII Int. Conf. Inf., Commun. and Autom. Technol. (ICAT)* (pp. 1–8). doi: 10.1109/ICAT54566.2022.9811120
2. Amari, S.-i., & Nagaoka, H. (2000). *Methods of Information Geometry* (Translations of Mathematical Monographs, vol. 191). Providence, RI: American Mathematical Society.
3. Amari, S. (2013). Information Geometry and Its Applications: Survey. In F. Nielsen & F. Barbaresco (Eds.), *Geometric Science of Information* (Lecture Notes in Computer Science, Vol. 8085). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-40020-9_1
4. Bian, N. H., Kanga, S.-H. A., Okou, G. C., & Hili, O. (2025). Robust goodness-of-fit test for copulas by the maximum mean discrepancy estimator. *Gulf Journal of Mathematics*, *20*(1), 1–21. https://doi.org/10.56947/gjom.v20i.2857
5. Chen, M., Hu, Q., Yu, Z., Thomas, H., Feng, A., Hou, Y., McCullough, K., Ren, F., & Soibelman, L. (2022). STPLS3D: A Large-Scale Synthetic and Real Aerial Photogrammetry 3D Point Cloud Dataset. arXiv preprint arXiv:2203.09065. https://arxiv.org/abs/2203.09065
6. Cover, T. M., & Thomas, J. A. (1991). *Elements of Information Theory.* John Wiley & Sons. https://doi.org/10.1002/0471200611
7. Cramér, H., & Wold, H. O. A. (1936). Some Theorems on Distribution Functions. *Journal of the London Mathematical Society*, 290–294. https://doi.org/10.1112/jlms/s1-11.4.290
8. EPFL Geometry Point Cloud Dataset. (2023). https://www.epfl.ch/labs/mmspg/downloads/geometry-point-cloud-dataset/. Accessed: Sep 20, 2025.
9. Goldberger, J., Gordon, S., & Greenspan, H. (2003). An efficient image similarity measure based on approximations of KL-divergence between two Gaussian mixtures. In *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV)* (Vol. 1, pp. 487–493). https://doi.org/10.1109/ICCV.2003.1238387
10. Hackel, T., Wegner, J. D., & Schindler, K. (2016). Contour Detection in Unstructured 3D Point Clouds. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1610–1618. https://doi.org/10.1109/cvpr.2016.178
11. Harsha, K. V., & Moosath, K. S. S. (2014). F-Geometry and Amari's α-Geometry on a Statistical Manifold. *Entropy*, 16(5), 2472–2487. https://doi.org/10.3390/e16052472
12. Hasanbelliu, E., Giraldo, L. S., & Príncipe, J. C. (2014). Information Theoretic Shape Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36, 2436–2451. https://doi.org/10.1109/TPAMI.2014.2324585
13. Hausdorff, F. (2008). *Felix Hausdorff – Gesammelte Werke Band III: Mengenlehre (1927, 1935) Deskriptive Mengenlehre und Topologie* (pp. 1–1005). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-76807-4
14. Hershey, J. R., & Olsen, P. A. (2007). Approximating the Kullback–Leibler divergence between Gaussian mixture models. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Vol. 4, pp. IV-317–IV-320). https://doi.org/10.1109/ICASSP.2007.366913
15. Huang, X., Mei, G., Zhang, J., & Abbas, R. (2021). A comprehensive survey on point cloud registration. https://arxiv.org/abs/2103.02690
16. Jian, B., & Vemuri, B. C. (2005). A Robust Algorithm for Point Set Registration Using Mixture of Gaussians. In *Proc. Tenth IEEE Int. Conf. Computer Vision (ICCV'05)* (Vol. 2, pp. 1246–1251). https://doi.org/10.1109/ICCV.2005.17
17. Kim, I., & Singh, S. K. (2025). 3D point cloud denoising via Gaussian processes regression. *The Visual Computer*, 41(12), 10481–10496. https://doi.org/10.1007/s00371-025-04049-7

18. Zhang, H., Zhong, J., & Chen, C. (2024). An Algorithm for Extracting Edge Feature Points of Point Clouds based on Plane Segmentation and Dimensionality Reduction. In *Proceedings of the 2024 6th International Conference on Telecommunications and Communication Engineering (ICTCE '24)* (pp. 59–64). https://doi.org/10.1145/3705391.3705401
19. Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22, 79–86. https://doi.org/10.1214/aoms/1177729694
20. Lee, J. M. (2012). *Introduction to Smooth Manifolds*. New York: Springer. https://doi.org/10.1007/978-1-4419-9982-5
21. Lee, Y., Kim, S., Choi, J., & Park, F. (2022). A Statistical Manifold Framework for Point Cloud Data. In *Proceedings of the International Conference on Machine Learning (ICML)* (pp. 12378–12402).
22. Li, F., Fujiwara, K., & Matsushita, Y. (2021). Toward a Unified Framework for Point Set Registration. In *Proc. 2021 IEEE Int. Conf. Robotics and Automation (ICRA)* (pp. 12981–12987). https://doi.org/10.1109/ICRA48506.2021.9561594
23. Mémoli, F., & Sapiro, G. (2004). Comparing Point Clouds. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing (SGP '04)* (pp. 32–40). https://doi.org/10.1145/1057432.1057436
24. Moenning, C., & Dodgson, N. A. (2003). Fast Marching Farthest Point Sampling. *Tech. Rep., University of Cambridge, Computer Laboratory*. https://doi.org/10.2312/egp.20031024
25. MPI-FAUST Dataset. (2023). https://faust-leaderboard.is.tuebingen.mpg.de/. Accessed: Sep 20, 2025.
26. Myronenko, A., & Song, X. (2010). Point Set Registration: Coherent Point Drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32, 2262–2275. https://doi.org/10.1109/TPAMI.2010.46
27. Nagar, R. (2025). Robust extrinsic symmetry estimation in 3D point clouds. *The Visual Computer*, 41, 115–128. https://doi.org/10.1007/s00371-024-03313-6
28. Nielsen, F. (2020). An Elementary Introduction to Information Geometry. *Entropy*, 22(10), 1100. https://doi.org/10.3390/e22101100
29. Reynolds, D. (2009). Gaussian Mixture Models. In S. Z. Li & A. Jain (Eds.), *Encyclopedia of Biometrics* (pp. 659–663). Boston, MA: Springer US. doi: 10.1007/978-0-387-73003-5_196
30. Rubner, Y., Tomasi, C., & Guibas, L. J. (2000). The Earth Mover's Distance as a Metric for Image Retrieval. *International Journal of Computer Vision*, 40, 99–121. https://doi.org/10.1023/A:1026543900054
31. Teicher, H. (1963). Identifiability of finite mixtures. *The Annals of Mathematical Statistics*, 34(4), 1265–1269. https://doi.org/10.1214/aoms/1177703862
32. Teenbaum, J., Silva, D. D., & Langford, J. A. (2000). Global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500), 2319–2323. https://doi.org/10.1126/science.290.5500.2319
33. Vishwakarma, A., & Moosath, K. S. (2025). Enhanced 3D Shape Analysis via Information Geometry. *arXiv preprint arXiv:2512.16213*. https://arxiv.org/abs/2512.16213
34. Vishwakarma, A., & Subrahamanian Moosath, K. S. (2025). Distance Measure Based on an Embedding of the Manifold of K-Component Gaussian Mixture Models into the Manifold of Symmetric Positive Definite Matrices. arXiv:2501.07429. https://arxiv.org/abs/2501.07429
35. Wang, F., Vemuri, B. C., & Rangarajan, A. (2006). Groupwise Point Pattern Registration Using a Novel CDF-Based Jensen–Shannon Divergence. In *Proc. 2006 IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR'06)* (pp. 1283–1288). https://doi.org/10.1109/CVPR.2006.131
36. Wang, Q., & Kim, M.-K. (2019). Applications of 3D point cloud data in the construction industry: A fifteen-year review from 2004 to 2018. *Advanced Engineering Informatics*, 39(C), 306–319. doi: 10.1016/j.aei.2019.02.007

37. Yakowitz, S. J., & Spragins, J. D. (1968). On the identifiability of finite mixtures. *The Annals of Mathematical Statistics*, 39(1), 209–214. https://doi.org/10.1214/AOMS/1177698520
38. Yang, Y., Feng, C., Shen, Y., & Tian, D. (2018). FoldingNet: Point Cloud Auto-Encoder via Deep Grid Deformation. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)* (pp. 206–215). https://doi.org/10.1109/CVPR.2018.00029
39. Zhang, Y.-X., Gui, J., Yu, B., Cong, X., Gong, X., Tao, W., & Tao, D. (2025). Deep Learning-Based Point Cloud Registration: A Comprehensive Survey and Taxonomy. arXiv preprint arXiv:2404.13830. https://arxiv.org/abs/2404.13830
40. Zhu, H., & Rohwer, R. (1997). Measurements of Generalisation Based on Information Geometry. In S. W. Ellacott, J. C. Mason, & I. J. Anderson (Eds.), *Mathematics of Neural Networks* (Operations Research/Computer Science Interfaces Series, Vol. 8). Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-6099-9_69

[1] Department of Mathematics, Indian Institute of Space Science and Technology, Trivandrum, Kerala 695547, India.

Email address: amitvishwakarma.22@res.iist.ac.in

Email address: smoosath@iist.ac.in