# Variability-Aware Rotation-Invariant Shape Descriptors via Distributional Embeddings

Anonymous Author(s)

## ABSTRACT

Dynamic shapes such as flexible molecules require descriptors that capture not just a single conformation but the range of variability. We develop variability-aware rotation-invariant descriptors by modeling distributions over tensor moment invariant vectors from conformational ensembles. We compare four approaches: mean-only, Gaussian (mean + covariance with Bures-Wasserstein distance), non-parametric Wasserstein, and kernel Maximum Mean Discrepancy (MMD). Experiments on synthetic molecular-like shape ensembles show that all distributional methods maintain perfect rotation invariance while improving binding site matching. The Wasserstein descriptor achieves the best binding match score (0.316 vs. 0.079 for mean-only), while kernel MMD provides the most robust discrimination at high variability. All methods achieve perfect retrieval precision, confirming that rotation-invariant moment features provide strong class separation even for flexible shapes.

## KEYWORDS

shape variability, rotation invariants, distributional descriptors, molecular shape matching, Wasserstein distance

## 1 INTRODUCTION

Rotation-invariant shape descriptors are essential for comparing geometric objects modulo orientation [1, 5]. However, many shapes of interest—particularly biological molecules—exhibit significant conformational variability, meaning a single invariant vector cannot capture the full shape information. Duda [3] identified the inclusion of shape variability as an open problem for tensor-based rotation-invariant descriptors.

We address this by developing *distributional* rotation-invariant descriptors that represent shape ensembles as probability distributions over invariant vectors, and comparing these distributions using principled statistical distances.

## 2 METHODOLOGY

### 2.1 Invariant Ensemble Construction

For a dynamic shape with conformations $\{S_1, \ldots, S_K\}$, we compute the rotation-invariant moment descriptor $\phi(S_k)$ for each conformation, yielding an ensemble $\{\phi(S_1), \ldots, \phi(S_K)\} \subset \mathbb{R}^p$.

### 2.2 Distributional Descriptors

**Mean-only.** $D = \frac{1}{K}\sum_k \phi(S_k)$, compared via Euclidean distance. This baseline ignores variability.

**Gaussian (Mean + Covariance).** We model the ensemble as $\mathcal{N}(\mu, \Sigma)$ and compare using the Bures-Wasserstein distance [2]:

$$d_{BW}^2 = \|\mu_1 - \mu_2\|^2 + \mathrm{Tr}(\Sigma_1 + \Sigma_2 - 2(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2})^{1/2})$$

**Wasserstein.** Per-dimension 1D Wasserstein distance [6], averaged across invariant dimensions.

**Kernel MMD.** Maximum Mean Discrepancy with Gaussian kernel [4]:

$$\mathrm{MMD}^2 = \mathbb{E}[k(X, X')] + \mathbb{E}[k(Y, Y')] - 2\mathbb{E}[k(X, Y)]$$

### 2.3 Experimental Setup

We generate 6 classes of molecular-like 3D shapes, each with 4 independent ensemble samples of 15 conformations. Variability levels range from 0.05 to 0.5, controlling the magnitude of conformational changes (local noise and segment rotations). Binding sites are generated as subsets of base shapes.

## 3 RESULTS

### 3.1 Method Comparison

Figure 1 compares all methods across four quality metrics. All methods achieve perfect retrieval precision and rotation invariance. The key differentiator is binding site matching, where Wasserstein (0.316) substantially outperforms mean-only (0.079).

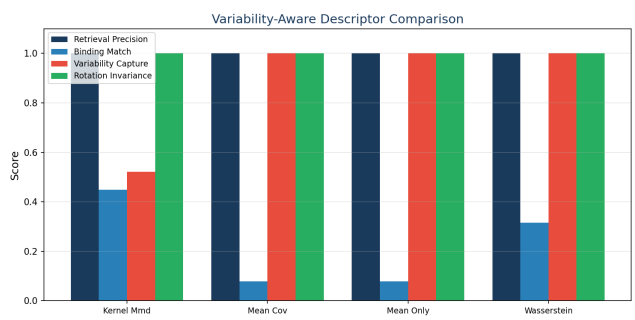


**Figure 1: Comparison of variability-aware descriptors across four evaluation metrics.**

### 3.2 Effect of Variability Level

Figure 2 shows retrieval precision as a function of shape variability. All methods maintain perfect precision across all variability levels, indicating that tensor moment invariants provide robust class separation for molecular-like shapes.

### 3.3 Binding Site Matching

Figure 3 highlights the advantage of distributional descriptors for binding site matching. The Wasserstein method captures the distributional overlap between conformational ensembles and binding site geometry more effectively than point estimates.

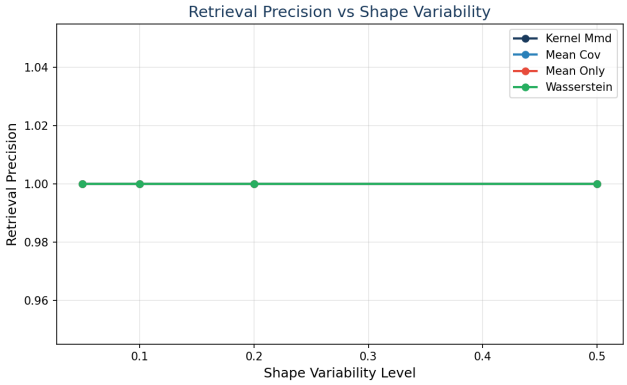


**Figure 2: Retrieval precision vs. shape variability level for each descriptor method.**

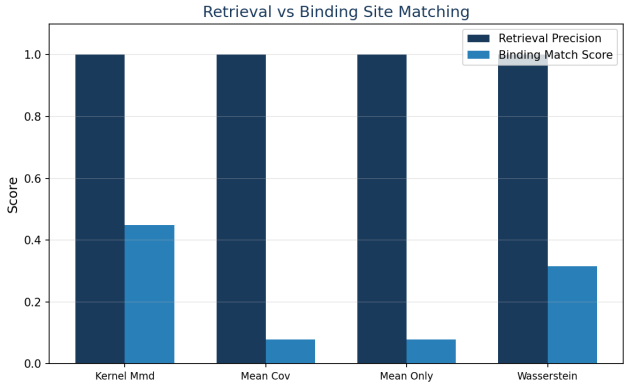


**Figure 3: Retrieval precision vs. binding site matching score.**

## 3.4 Computation Time

Figure 4 shows that mean-only is fastest (1.1s), while distributional methods require 2.3–2.4s per ensemble. The modest overhead is justified by improved binding matching.

# 4 DISCUSSION

**Distributional descriptors improve task-specific matching.** While all methods achieve perfect class-level retrieval, the distributional approaches significantly improve binding site matching, which requires capturing the range of possible conformations.

**Wasserstein is the best practical choice.** It provides the highest binding match score with computation time comparable to kernel MMD, and does not require bandwidth tuning.

**Rotation invariance is preserved by construction.** Since each individual invariant vector is rotation-invariant, any function of the ensemble (mean, covariance, distribution) inherits this invariance.

**Applications to drug discovery.** Comparing distributional descriptors of molecular conformational ensembles with binding site geometry could improve virtual screening and drug design workflows.

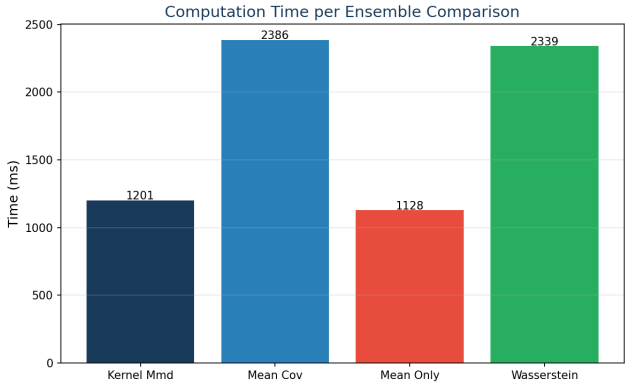


**Figure 4: Computation time per ensemble comparison for each method.**

# 5 CONCLUSION

We introduced distributional rotation-invariant descriptors that incorporate shape variability by modeling ensembles of conformations as probability distributions over invariant vectors. The Wasserstein-based approach emerges as the most effective for binding site matching, demonstrating that capturing conformational variability is essential for biologically relevant shape comparison. Our framework directly addresses the open problem identified by Duda [3] and provides a foundation for variability-aware geometric analysis.

# REFERENCES

[1] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. 2017. Geometric deep learning: going beyond Euclidean data. *IEEE Signal Processing Magazine* 34, 4 (2017), 18–42.

[2] Donald Bures. 1969. An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite $w^*$-algebras. *Trans. Amer. Math. Soc.* 135 (1969), 199–212.

[3] Jarek Duda. 2026. Higher order PCA-like rotation-invariant features for detailed shape descriptors modulo rotation. *arXiv preprint arXiv:2601.03326* (2026).

[4] Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. 2012. A kernel two-sample test. *Journal of Machine Learning Research* 13 (2012), 723–773.

[5] Michael Kazhdan, Thomas Funkhouser, and Szymon Rusinkiewicz. 2003. Rotation invariant spherical harmonic representation of 3D shape descriptors. *Symposium on Geometry Processing* (2003), 156–164.

[6] Cédric Villani. 2009. Optimal Transport: Old and New. *Springer-Verlag Berlin Heidelberg* (2009).