Intrinsic Mean Shift for Clustering on Stiefel and Grassmann Manifolds

Hasan Ertan Çetingül René Vidal
Center for Imaging Science, Johns Hopkins University, Baltimore MD 21218, USA
{ertan,rvidal}@cis.jhu.edu

Abstract

The mean shift algorithm, which is a nonparametric density estimator for detecting the modes of a distribution on a Euclidean space, was recently extended to operate on analytic manifolds. The extension is extrinsic in the sense that the inherent optimization is performed on the tangent spaces of these manifolds. This approach specifically requires the use of the exponential map at each iteration. This paper presents an alternative mean shift formulation, which performs the iterative optimization "on" the manifold of interest and intrinsically locates the modes via consecutive evaluations of a mapping. In particular, these evaluations constitute a modified gradient ascent scheme that avoids the computation of the exponential maps for Stiefel and Grassmann manifolds. The performance of our algorithm is evaluated by conducting extensive comparative studies on synthetic data as well as experiments on object categorization and segmentation of multiple motions.

1. Introduction

The mean shift (MS) algorithm is a nonparametric kernel density estimator that analyzes multimodal feature spaces directly from data points [5, 7, 14]. Given a collection of points distributed according to an unknown distribution on a Euclidean space, the MS is designed to iteratively locate the underlying modes together with the points that belong to the cluster associated with each mode. The algorithm is successfully employed in different computer vision problems, such as image segmentation [7, 21] and tracking [3, 9, 12, 22]. Furthermore, its characterization as an optimization problem are investigated in [13, 19, 4, 30] and references therein.

The success of the mean shift algorithm inspired many researchers from the computer vision community to develop different variants of the standard version. For instance, there exist a plethora of works that focus on improving its performance in terms of 1) speed (see [21] and references therein) and 2) accuracy via adaptive bandwidths [15] and asymmetric kernels [29]. It is also worth mentioning recent modifications of the MS for manifold clustering. In [24, 27], the MS algorithm was extended to two analytic manifolds, Grassmann manifolds and Lie groups, in order to address the problems of motion segmentation and multibody factorization. Following the introduction of this nonlinear extension, the medoid shift [23] and the quick shift [28] algorithms are designed to cluster data on non-Euclidean spaces and employed for image segmentation and categorization. Specifically, by constraining the points traversed towards a mode to pass through the actual data points, the medoid shift eliminates the definition of a stopping criteria and performs clustering on both linear and curved spaces. The quick shift was recently proposed to efficiently eliminate the over-fragmentation problem of the medoid shift.

1.1. Motivation and Contributions

Our work draws inspiration from the works of Tuzel *et al*. [27] and Subbarao and Meer [24], where the Euclidean MS formulation was extended to two particular analytic manifolds, Grassmann manifolds and Lie groups, and named as *nonlinear* MS. The basic idea behind the nonlinear MS is to compute the mean shift as a weighted sum of tangent vectors and map the resulting vector back to the manifold. In other words, the MS iterations are still performed on a linear space, namely the tangent space to the manifold.

At this point, it is also worth noting that there exist well-founded statistical tools, *e.g*., probability density functions (pdfs) and kernels, for the analysis of analytical manifolds [2, 6]. In fact, several problems in computer vision and in biological sciences involve the analysis of data points on particular analytic manifolds. Typical examples include motion segmentation [24], object recognition and classification [2, 17], activity recognition [26], text categorization [1, 17], and gene expression analysis [1, 17, 20]. From this perspective, it is natural to see efforts for developing iterative clustering methods that are guaranteed to operate "on" the manifold of interest. For instance, Oba *et al*. presented a hyperspherical mean shift algorithm with the von Mises-Fisher kernel to analyze gene expression profiling data [20]. Although they perform density estimation "on" a particular analytic manifold, *i.e*., unit hyperspheres, the method is not generalizable in the sense that it cannot be applied to

other analytic manifolds because of the specificity of the inherent kernel function. In addition, in the case of complex non-analytic manifolds, one can employ dimensionality reduction techniques that map the data to its intrinsic parameter manifold (see [16] and references therein). Therefore, it would be interesting to *intrinsically* reformulate the MS algorithm on analytic manifolds, investigate its convergence, and evaluate its clustering performance and efficiency.

We thereby propose an intrinsic mean shift algorithm that is designed to operate on two particular manifolds, *i.e*., Stiefel and Grassmann manifolds, using generic kernels. Specifically, we focus on clustering directional data, which constitute a particular case of the Stiefel manifold (unit hyperspheres), and segmentation of multiple motions under the affine camera model, which involves clustering subspaces, *i.e*., points on the Grassmann manifold. The idea is to perform kernel density estimation on the manifold and locate the fixed point(s) of a mapping via iterative evaluations. The resulting points are identified as the modes of the underlying density along with the memberships of the data points.

The organization of the paper is as follows: In §2, we revisit the fundamentals of the MS algorithm along with its nonlinear (extrinsic) extension to analytic manifolds. In §3, we present the mathematical details of our intrinsic MS formulation and briefly discuss its convergence. In §4, we provide synthetic experiments to compare our method with its extrinsic counterpart. In addition, we apply the intrinsic MS on object categorization and segmentation of multiple motions, and compare it with other clustering techniques for directional data. Finally, §5 outlines the conclusions along with future research directions.

2. Preliminaries

2.1. Euclidean Mean Shift

The MS algorithm is a nonparametric kernel density estimator that iteratively locates the modes of a density function by gradient ascent. Suppose that the set of data points $\mathcal{X} = \{\boldsymbol{x}_n\}_{n=1}^N \subset \mathbb{R}^m$ has been independently sampled from an unknown density function $\mathbf{f}$. The kernel density estimate of $\mathbf{f}$ at $\boldsymbol{x}$, denoted by $\hat{\mathbf{f}}(\boldsymbol{x}; h)$, is defined as

$$\hat{\mathbf{f}}(\boldsymbol{x}; h) = c_\Phi \sum_{n=1}^{N} \Phi(\boldsymbol{x}_n, \boldsymbol{x}; h), \tag{1}$$

where Φ is a kernel function and c_Φ is a normalization term that depends on the number of points N, the dimension m, and the bandwidth $h > 0$.

A natural choice for the kernel function is the class of radially symmetric kernels, such as the Gaussian kernel or the Epanechnikov kernel [7]. If this class of kernels is selected for density estimation, one can first define the profile function ϕ of the kernel Φ such that $\phi(u_n; h) = \Phi(\boldsymbol{x}_n, \boldsymbol{x}; h)$, where $u_n = \mathbf{g}(\boldsymbol{x}_n, \boldsymbol{x})$ for some function $\mathbf{g} : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^+$. In addition, if one further defines the shadow of the profile as $\psi(u_n; h) = -\frac{\mathrm{d}\phi}{\mathrm{d}u_n}(u_n; h)$ and rewrites (1) in terms of ψ, then the gradient of the density estimate is given by

$$\nabla_{\boldsymbol{x}} \hat{\mathbf{f}}(\boldsymbol{x}; h) = c_\psi \sum_{n=1}^{N} \psi(u_n; h) \underbrace{\left[\frac{\sum_{n=1}^N \psi(u_n; h) \boldsymbol{x}_n}{\sum_{n=1}^N \psi(u_n; h)} - \boldsymbol{x} \right]}_{\boldsymbol{m}(\boldsymbol{x}; h)}, \tag{2}$$

where c_ψ is the corresponding normalization term that depends on $\{N, m, h\}$. The second term in (2) is referred to as the *mean shift* $\boldsymbol{m}(\boldsymbol{x}; h)$, which yields gradient ascent iterations of the form $\boldsymbol{x}^{(i+1)} = \boldsymbol{x}^{(i)} + \boldsymbol{m}(\boldsymbol{x}^{(i)}; h)$. Under some conditions, the resulting scheme converges to the modes of the underlying distribution [7, 19]. In addition, the density estimation can be further improved by using a different bandwidth $h_n \equiv h(\boldsymbol{x}_n)$ for each data point $\boldsymbol{x}_n$ [8].

2.2. Nonlinear (Extrinsic) Mean Shift

Following the work of Tuzel *et al*. [27], which performs motion estimation by detecting modes on Lie groups, Subbarao and Meer presented a nonlinear mean shift algorithm to cluster data points lying on Lie groups or Grassmann manifolds [24]. The idea is to compute the mean shift as a weighted sum of vectors in the tangent spaces of the manifold of interest. Therefore, we refer to this algorithm as the *extrinsic* mean shift (Ext-MS). Specifically, consider an analytic manifold $\mathcal{M}$ with a metric d and the set of points $\mathcal{X} = \{X_n\}_{n=1}^N \subset \mathcal{M}$ along with the definitions in §2.1. Then the kernel density estimate can be written as

$$\hat{\mathbf{f}}(X; h) = c_\phi \sum_{n=1}^{N} \phi\big(d^2(X, X_n); h\big), \tag{3}$$

with the normalization term c_ϕ. By taking the gradient of (3) and using the shadow ψ, the mean shift is computed as

$$\boldsymbol{m}(X; h) = -\frac{\sum_{n=1}^N \nabla_X d^2(X, X_n)\ \psi\big(d^2(X, X_n); h\big)}{\sum_{n=1}^N \psi(d^2(X, X_n); h)}, \tag{4}$$

where $\forall n$, $\nabla_X d^2(X, X_n) \in T_X\mathcal{M}$, *i.e*., the tangent space of $\mathcal{M}$ at X. The algorithm proceeds by moving the point along the geodesic defined by the mean shift. Accordingly, given the exponential map at X as $\exp_X(\cdot) : T_X\mathcal{M} \to \mathcal{M}$, the update equation, *i.e*., one mean shift iteration, is of the form

$$X^{(i+1)} = \exp_{X^{(i)}}\Big(\boldsymbol{m}(X^{(i)}; h)\Big). \tag{5}$$

Due to the extrinsic nature of the formulation, one can still employ the aforementioned radially symmetric kernels. However, one also needs the exponential map (5) to obtain the resulting point on the manifold [24]. The reader is referred to [27, 24] for further details.

Extrinsic MS on Stiefel Manifolds: The Stiefel manifold $\mathcal{V}_{k,m}$ comprises the space of $k \leq m$ orthonormal vectors in $\mathbb{R}^m$, which is represented in matrix form as $\mathcal{V}_{k,m} = \{X \in \mathbb{R}^{m\times k} | X^\top X = I_k\}$, where I_k is the $k \times k$ identity matrix [6, 11]. Special cases constitute the unit $(m-1)$-sphere $\mathbb{S}^{m-1} = \{\boldsymbol{s} \in \mathbb{R}^m | \boldsymbol{s}^\top \boldsymbol{s} = 1\} = \mathcal{V}_{1,m}$ and the group of orthogonal $m \times m$ matrices $\mathcal{O}(m) = \mathcal{V}_{m,m}$.

In order to apply the extrinsic MS on $\mathcal{V}_{k,m}$, let us define a discrepancy measure between two matrices $X, X_n \in \mathcal{V}_{k,m}$ as $I_k - X_n^\top X$ [6] and the corresponding metric as

$$d^2(X, X_n) = k - \text{tr}(X_n^\top X). \tag{6}$$

Using the formula in [11] for the gradient of a differentiable function $\mathbf{f} : \mathcal{V}_{k,m} \to \mathbb{R}$, *i.e.*, $\nabla_X \mathbf{f} = \mathbf{f}_X - X\mathbf{f}_X^\top X$ where $\mathbf{f}_X \doteq \frac{\partial \mathbf{f}}{\partial X}$, the gradient of the metric becomes

$$\nabla_X d^2(X, X_n) = XX_n^\top X - X_n. \tag{7}$$

Now, given a tangent vector $\Delta \in T_X\mathcal{V}_{k,m} \subset \mathbb{R}^{m\times k}$, one can first obtain the matrices $Q \in \mathbb{R}^{m\times k}$ and $R \in \mathbb{R}^{k\times k}$, which denote the compact QR decomposition of $(I_m - XX^\top)\Delta$, to compute the exponential map [11] on $\mathcal{V}_{k,m}$ as

$$\exp_X(\Delta) = XB + QC, \tag{8}$$

where $B, C \in \mathbb{R}^{k\times k}$ are given by

$$\begin{bmatrix} B \\ C \end{bmatrix} = \exp\left(\begin{bmatrix} A & -R^\top \\ R & \mathbf{0}_k \end{bmatrix}\right)\begin{bmatrix} I_k \\ \mathbf{0}_k \end{bmatrix} \text{ with } A = X^\top\Delta.$$

Thus, by substituting (6)-(7) into (4) and using any radially symmetric kernel, the mean shift on the tangent space can be calculated. The corresponding point on $\mathcal{V}_{k,m}$ is then obtained using the exponential map (8).

Extrinsic MS on Grassmann Manifolds: The Grassmann manifold $\mathcal{G}_{k,m-k}$ comprises the space of k-dimensional linear subspaces in $\mathbb{R}^m$. Equivalently, it is also obtained by identifying the matrices in $\mathcal{V}_{k,m}$ whose columns span the quotient manifold $\mathcal{V}_{k,m}/\mathcal{O}(k)$ [6, 11]. Therefore, $\mathcal{G}_{k,m-k}$ is equivalent to $\mathcal{P}_{k,m-k}$, *i.e.*, the space of $m \times m$ orthogonal projection matrices of rank $k < m$. We will frequently use this equivalence in the following discussions.

To employ the extrinsic MS on $\mathcal{G}_{k,m-k}$, let $P = XX^\top$ and $P_n = X_nX_n^\top$ denote two points on $\mathcal{P}_{k,m-k}$ such that $X, X_n \in \mathcal{V}_{k,m}$. A discrepancy measure between P and P_n is of the form $I_k - X_n^\top XX^\top X_n$ for $k < m$, and the corresponding metric can be written as

$$d^2(X, X_n) = k - \text{tr}(X_n^\top XX^\top X_n). \tag{9}$$

Using the formula in [11] for the gradient of a function $\mathbf{f} : \mathcal{G}_{k,m-k} \to \mathbb{R}$, *i.e.*, $\nabla_X \mathbf{f} = \mathbf{f}_X - XX^\top \mathbf{f}_X$, one can write the gradient of the metric as

$$\nabla_X d^2(X, X_n) = -2(I_m - XX^\top)X_nX_n^\top X. \tag{10}$$

Finally, the exponential map of a tangent vector $\Delta \in T_X\mathcal{G}_{k,m-k}$ at X is of the form

$$\exp_X(\Delta) = [XV\cos(\Sigma) + U\sin(\Sigma)]\,V^\top, \tag{11}$$

where $U\Sigma V^\top$ represents the compact singular value decomposition (SVD) of the tangent vector Δ [11].

Substituting (9)-(10) into (4) and using any radially symmetric kernel, we obtain the mean shift, which is mapped back to the manifold using the exponential map (11).

3. Intrinsic Mean Shift on Analytic Manifolds

We now present the details of our intrinsic mean shift (Int-MS) formulation. The rationale behind our approach is to perform the kernel density estimation on the manifold of interest and locate the modes via consecutive evaluations of an appropriate mapping $\mathbf{p}$. In particular, the inherent optimization constitutes a modified gradient ascent scheme that avoids the computation of the exponential map at each iteration. Our formulation is initiated by employing two different estimators proposed in [6], which use a kernel function of the form $\Phi(T) = \text{etr}(-T) \doteq \exp(\text{tr}(-T))$ for some $T \in \mathbb{R}^{k\times k}$, in order to estimate unknown density functions on Stiefel or Grassmann manifolds.

3.1. Formulation on Stiefel Manifolds $\mathcal{V}_{k,m}$

Recall that a discrepancy measure between two matrices $X, X_n \in \mathcal{V}_{k,m}$ is $I_k - X_n^\top X$. Given a set of data points $\mathcal{X} = \{X_n\}_{n=1}^N$ that are independently sampled from a density $\mathbf{f}$ on $\mathcal{V}_{k,m}$, its estimate $\hat{\mathbf{f}}$ at X can be computed as

$$\hat{\mathbf{f}}(X; M) = c_1 \sum_{n=1}^{N} \Phi(M^{-\frac{1}{2}}(I_k - X_n^\top X)M^{-\frac{1}{2}}), \tag{12}$$

where $M \in \mathbb{R}^{k\times k}$ is a symmetric positive definite smoothing parameter matrix (which generalizes the bandwidth h) and c_1 is the normalization constant [6]. We subsequently compute the gradient of (12), which is given by

$$\nabla_X \hat{\mathbf{f}}(X; M) = \beta_1(\bar{X} - X\bar{X}^\top X), \tag{13}$$

where $\beta_1 = c_1\text{etr}(-M^{-1})$ is a constant term and

$$\bar{X} = \mathbf{b}_1(X; \mathcal{X}) = \sum_{n=1}^{N} X_n M^{-1}\text{etr}(M^{-1}X_n^\top X). \tag{14}$$

Our intrinsic formulation involves the definition of a mapping $\mathbf{p} \doteq (\mathbf{a}_1 \circ \mathbf{b}_1) : \mathcal{V}_{k,m} \to \mathcal{V}_{k,m}$, which is iteratively evaluated as $X^{(i+1)} = \mathbf{p}\left(X^{(i)}\right)$ so that $\nabla_{X^{(i)}}\hat{\mathbf{f}}(X^{(i)}; M) \to \mathbf{0}$ as $i \to \infty$. For this purpose, an intuitive choice for $\mathbf{a}_1$ is to take the Q-part of the compact QR-decomposition of $\bar{X}$, and set it as the new point on the manifold. The resulting

iterative scheme can be summarized as follows: At the i-th iteration, given the set $\mathcal{X}$ and the current mean $X^{(i)}$, one can then get an estimate for the new mean as

$$X^{(i+1)} = \mathbf{p}(X^{(i)}) = (\mathbf{a}_1 \circ \mathbf{b}_1)(X^{(i)}; \mathcal{X}) = Q_{\mathcal{X}}^{(i)}, \quad (15)$$

where $Q_{\mathcal{X}}^{(i)} \in \mathcal{V}_{k,m}$ denotes the Q-part of the the compact QR-decomposition of $\bar{X}^{(i)}$, *i.e.*, $\mathbf{a}_1(\bar{X}^{(i)}) = Q_{\mathcal{X}}^{(i)}$. The validity of this iterative scheme will be discussed next.

Note on the convergence of the Int-MS: The analysis of convergence of the Euclidean MS was presented by Comaniciu and Meer in [7], followed by the work of Li *et al*. [19] where the proof in [7] was corrected. Although our method is conceptually similar to the Euclidean MS by being a kernel density estimation framework, the inherent optimization is reformulated in terms of iterative evaluations of a specific mapping. Nevertheless, at each iteration, if the current point moves along a direction "consistent" with the gradient at that point, the proposed scheme converges to the modes of the underlying distribution. Specifically, consecutive evaluations of the mapping $\mathbf{p}$ constitute a modified Riemannian gradient ascent scheme and locate a fixed point if

$$\left\langle \log_{X^{(i)}}(X^{(i+1)}), \nabla_{X^{(i)}} \hat{\mathbf{f}}(X^{(i)}; M) \right\rangle_{X^{(i)}} > 0, \quad (16)$$

where $\langle \cdot, \cdot \rangle_X : T_X\mathcal{M} \times T_X\mathcal{M} \to \mathbb{R}$ is the Riemannian metric, and $\log_X(\cdot) : \mathcal{M} \to T_X\mathcal{M}$ is the logarithm map. In the case of Stiefel manifolds $\mathcal{V}_{k,m}$ for $k \neq 1$, proving (16) is not straightforward, though the theoretical result for the case of unit hypersphere, which is elaborated next, gives us a strong expectation that this will be so. This expectation is borne out by empirical experience.

Case on $\mathcal{V}_{1,m}$: We now present the analysis of the convergence of our intrinsic formulation on the unit hypersphere $\mathbb{S}^{m-1} \equiv \mathcal{V}_{1,m}$. Note that, in a slight abuse of notation, the gradient of the function $\hat{\mathbf{f}}$ at $\boldsymbol{x} = \boldsymbol{x}^{(i)}$ is given by

$$\begin{aligned} \nabla\hat{\mathbf{f}}(\boldsymbol{x}) &= \beta_1 \sum_{n=1}^{N} \left(\frac{1}{M}\boldsymbol{x}_n - \frac{\boldsymbol{x}_n^\top \boldsymbol{x}}{M}\boldsymbol{x} \right) \text{etr}\left(\frac{\boldsymbol{x}_n^\top \boldsymbol{x}}{M} \right) \\ &\sim \bar{\boldsymbol{x}} - (\boldsymbol{x}^\top \bar{\boldsymbol{x}})\boldsymbol{x}. \end{aligned} \quad (17)$$

Assuming that $\boldsymbol{x}$ and $\boldsymbol{y}$ are not antipodal, the logarithm map of $\boldsymbol{y} = \boldsymbol{x}^{(i+1)}$ at $\boldsymbol{x}$ has the following form

$$\begin{aligned} \log_{\boldsymbol{x}}(\boldsymbol{y}) &= \frac{\boldsymbol{y} - (\boldsymbol{x}^\top \boldsymbol{y})\boldsymbol{x}}{\|\boldsymbol{y} - (\boldsymbol{x}^\top \boldsymbol{y})\boldsymbol{x}\|} \arccos(\boldsymbol{x}^\top \boldsymbol{y}) \\ &\sim \boldsymbol{y} - (\boldsymbol{x}^\top \boldsymbol{y})\boldsymbol{x}. \end{aligned} \quad (18)$$

Note that $\boldsymbol{y} = \boldsymbol{x}^{(i+1)}$ is the Q-part of the compact QR decomposition of $\bar{\boldsymbol{x}}^{(i)}$, hence $\boldsymbol{y} \sim \bar{\boldsymbol{x}}$. Therefore, (16) is immediately satisfied because $\log_{\boldsymbol{x}^{(i)}}(\boldsymbol{x}^{(i+1)})$ and $\nabla\hat{\mathbf{f}}(\boldsymbol{x}^{(i)})$ are parallel. As a consequence, our method and the nonlinear MS on the hypersphere coincide, up to a scale factor.

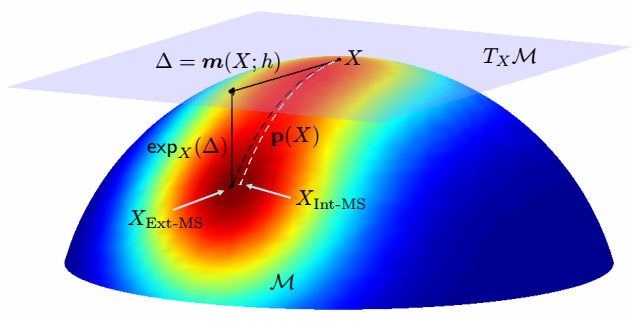

Figure 1. Illustration of the extrinsic and intrinsic MS iterations.

Figure 1 illustrates the difference between the iterations of the extrinsic and intrinsic mean shift. Specifically, the Ext-MS follows the geodesic defined by the mean shift at the expense of computing the exponential map, whereas the Int-MS follows a path on the manifold $\mathcal{M}$ consistent with the (local) gradients.

3.2. Formulation on Grassmann Manifolds $\mathcal{G}_{k,m-k}$

Suppose now that we are given a set of N points $\mathcal{R} = \{P_n | P_n = X_n X_n^\top, X_n \in \mathcal{V}_{k,m}, k < m\}$, which are independently sampled from a density $\mathbf{f}$ on the manifold $\mathcal{P}_{k,m-k} \equiv \mathcal{G}_{k,m-k}$. Recall that a discrepancy measure between two matrices P and P_n on $\mathcal{P}_{k,m-k}$ is $I_k - X_n^\top P X_n$. In this case, using the aforementioned kernel function Φ, the density estimate $\hat{\mathbf{f}}$ at $P = XX^\top$ becomes

$$\hat{\mathbf{f}}(P; M) = c_2 \sum_{n=1}^{N} \Phi(M^{-\frac{1}{2}}(I_k - X_n^\top P X_n)M^{-\frac{1}{2}}), \quad (19)$$

with the symmetric positive definite smoothing parameter matrix $M \in \mathbb{R}^{k \times k}$ and the normalization constant c_2 [6]. After replacing P with $XX^\top$, we compute the gradient of (19) with respect to X as

$$\nabla_X \hat{\mathbf{f}}(X; M) = \beta_2(\bar{P}X - XX^\top \bar{P}X), \quad (20)$$

with the constant term $\beta_2 = 2c_2 \text{etr}(-M^{-1})/N$ and

$$\bar{P} = \mathbf{b}_2(P; \mathcal{R}) = \sum_{n=1}^{N} X_n M^{-1} X_n^\top \text{etr}(M^{-1} X_n^\top P X_n). \quad (21)$$

In order to obtain an iterative scheme $P^{(i+1)} = \mathbf{p}\left(P^{(i)}\right)$, where $\mathbf{p} \doteq (\mathbf{a}_2 \circ \mathbf{b}_2) : \mathcal{G}_{k,m-k} \to \mathcal{G}_{k,m-k}$, we choose a mapping $\mathbf{a}_2$ that takes the first k columns of U from the SVD of $\bar{P} = U\Sigma U^\top$ and computes $UU^\top$. Therefore, at the i-th iteration, given the set $\mathcal{R}$ and the current mean $P^{(i)}$, the new mean $P^{(i+1)} = X^{(i+1)} X^{(i+1)\top}$ can be computed as

$$P^{(i+1)} = \mathbf{p}(P^{(i)}) = (\mathbf{a}_2 \circ \mathbf{b}_2)(P^{(i)}; \mathcal{R}) = U_{\mathcal{R}}^{(i)} U_{\mathcal{R}}^{(i)\top}, \quad (22)$$

where $U_{\mathcal{R}}^{(i)} \in \mathcal{V}_{k,m}$ denotes the first k columns of the U matrix in the SVD of $\bar{P}^{(i)}$, *i.e.*, $\mathbf{a}_2(\bar{P}^{(i)}) = U_{\mathcal{R}}^{(i)} U_{\mathcal{R}}^{(i)\top}$.

4. Experimental Results

The performance of our algorithm is first evaluated by synthetic experiments. Specifically, we compare the intrinsic mean shift (Int-MS) with its extrinsic counterpart (Ext-MS) in terms of clustering accuracy and speed. Next, we cluster directional data on unit hyperspheres and compare the Int-MS with other unsupervised hyperspherical clustering techniques for object categorization. Meanwhile, we also extract a different feature representation to further assess this problem on Grassmann manifolds. Finally, we employ our intrinsic method for segmentation of multiple motions.

4.1. Simulations on Synthetic Data

We first compare the clustering performance of our intrinsic MS formulation with that of the extrinsic counterpart in terms of the dimensions $\{m, k\}$ of the manifolds and the runtimes. For this purpose, pseudo-random matrices on the manifolds of interest ($\mathcal{V}_{k,m}$ or $\mathcal{G}_{k,m-k}$) are generated using the method described in [6]. Specifically, notice that any orthogonal matrix $S \in \mathcal{O}(m)$ can be represented as a product of $\frac{m(m-1)}{2}$ orthogonal matrices of the form

$$R_m^\nu(\theta) = \begin{bmatrix} I_{\nu-1} & 0 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) & 0 \\ 0 & \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 0 & I_{m-\nu-1} \end{bmatrix}, \quad (23)$$

for $1 \leq \nu \leq m-1$. If an auxiliary matrix S_m^ν is defined as $S_m^\nu = \prod_{j=\nu}^{m-1} R_m^j(\theta_{\nu,j})$, the orthogonal matrix $S \in \mathcal{O}(m)$ can be written as $S = \prod_{\nu=1}^{m-1} S_m^\nu$.

We conduct our experiments for clustering $N = 200$ points from 4 classes (50 points per class). For each class, we generated $\frac{m(m-1)}{2}$ angles $\{\theta_{\nu,j}\}$, each of which is randomly drawn from one of $\frac{m(m-1)}{2}$ bins in $[0, \pi)$. We then corrupted these angles with a uniform random noise $\eta \in [-\pi/9, \pi/9]$. Following the generation of the set of matrices $\{S\}$ for each class, we take the first k orthonormal columns of each S to obtain the matrices $X \in \mathcal{V}_{k,m}$ and $P = XX^\top \in \mathcal{G}_{k,m-k}$. Note that the rank of P should be at most $m-1$ so that (9) remains as a valid metric. Finally, we set the bandwidth $h = 0.1$ in Ext-MS and the smoothing parameter matrix $M = 0.1 I_k$ in Int-MS.

Table 1 shows the clustering rates of the Ext-MS and the Int-MS on particular Stiefel manifolds as well as the average of the ratio of the runtimes $t_{\text{Int-MS}}/t_{\text{Ext-MS}}$ after 50 trials. The maximum clustering rates are indicated in parenthesis. We observe that both methods achieve comparable rates in all 7 cases, and the Int-MS slightly outperforms the Ext-MS in terms of clustering accuracy in 5 cases. Furthermore, in the case of hyperspherical data on $\mathcal{V}_{1,m}$, $m = 3, 10, 50$, the runtimes of our method are less than those of the Ext-MS, whereas it becomes slower than the Ext-MS for $k > 1$.

Table 1. Clustering rates (%) of Ext-MS and Int-MS and their runtime ratio in the case of 4-class clustering on Stiefel $\mathcal{V}_{k,m}$.

$\mathcal{V}_{k,m}$		Performance		Runtime Ratio
m	k	Ext-MS	Int-MS	$t_{\text{Int-MS}}/t_{\text{Ext-MS}}$
3	1	70.94 (100)	71.22 (100)	0.77
3	2	78.37 (100)	82.81 (100)	1.11
3	3	79.43 (100)	78.55 (100)	1.03
5	3	90.58 (100)	91.93 (100)	1.78
10	3	96.00 (100)	96.00 (100)	1.58
10	1	82.28 (100)	83.21 (100)	0.74
50	1	80.74 (100)	84.09 (100)	0.66

Table 2 shows the clustering rates of the aforementioned algorithms on particular Grassmann manifolds as well as the average of the ratio of the runtimes after 50 trials. The maximum clustering rates are indicated in parenthesis. We observe that the clustering rates of the intrinsic and extrinsic formulations are comparable but relatively lower than the previous results. The Int-MS slightly outperforms the Ext-MS in terms of clustering accuracy in 5 cases, whereas it converges slower than the Ext-MS in all cases.

Table 2. Clustering rates (%) of Ext-MS and Int-MS and their runtime ratio in the case of 4-class clustering on Grassmann $\mathcal{G}_{k,m-k}$.

$\mathcal{G}_{k,m-k}$		Performance		Runtime Ratio
m	k	Ext-MS	Int-MS	$t_{\text{Int-MS}}/t_{\text{Ext-MS}}$
3	1	56.26 (99.50)	60.37 (99.50)	1.12
3	2	53.25 (87.50)	59.86 (96.50)	1.05
5	3	74.33 (100)	71.23 (100)	1.25
5	4	47.43 (92)	49.54 (83)	1.14
10	4	60.91 (100)	59.67 (100)	1.94
20	4	56.68 (100)	61.41 (100)	1.67
20	1	66.22 (100)	67.08 (100)	1.08

4.2. Object Categorization

Object categorization refers to the task of grouping similar objects of the same class in an image or video sequence. In our experiments, we select a subset of the ETH-80 data set [18], which contains images from 3 different object categories (see Figure 2). We conduct two separate experiments (Exp-80 and Exp-150) for which each category contains either 80 or 150 images, and extract features on different analytic manifolds as elaborated next.

Figure 2. Selected object categories from the ETH-80 data set.

i. The Unit Hypersphere: Quantitative analysis of directional data can be performed using either supervised (probabilistic) learning algorithms (see [17] and references therein) or hyperspherical versions of unsupervised techniques such as k-means (HSp-kM) [10] and expectation-maximization (HSp-EM) [1]. In that case, unit-norm feature vectors should be extracted from the object image. As described in [17], such a typical feature vector is obtained from the magnitude of the image gradient and the Laplacian at three different scales. Specifically, the 32-bin histograms of each of the six resulting images are computed and concatenated as a feature vector of length $m = 192$, which is then normalized. Therefore, the problem boils down to clustering points on the Stiefel manifold $\mathcal{V}_{1,192} \equiv \mathbb{S}^{191}$.

ii. The Grassmann Manifold: In the procedure described above, the normalization of the feature vector after concatenation may corrupt the class separability information embedded in the individual histograms. To overcome this problem, we use the square-root representation of probability mass functions. Specifically, assuming that the ℓ_1-norm of the each histogram is 1, we can take the square root of each entry to make their ℓ_2-norms equal to 1. Now, if we form a feature matrix by stacking the aforementioned six 32-bin histograms as columns and then taking the SVD of the resulting 32×6 matrix, its singular vectors span a subspace of dimension $k = 6$ in $\mathbb{R}^{m=32}$. Thus the new feature representation is a point on $\mathcal{G}_{6,32-6}$.

In the case of using the unit-norm feature vector representation, for the sake of a fair comparison, we solely focus on unsupervised learning techniques and compare Int-MS not only with Ext-MS, but also with HSp-kM and HSp-EM. On the other hand, if the points are on the Grassmann manifold, we only employ the Ext-MS (with bandwidth $h = 0.1$) or the Int-MS (with smoothing matrix $M = 0.1I_k$) for clustering. Table 3 shows the clustering performances of the aforementioned techniques using both representations ($\mathcal{V}_{k,m}$ and $\mathcal{G}_{k,m-k}$) for 3-class object categorization. It is observed that the Ext-MS and the Int-MS achieve comparable rates and outperform other techniques. Considering the fact that both HSp-kM and HSp-EM require an estimate for the number of classes in advance, the clustering performances ($>99\%$) of the MS formulations on $\mathcal{V}_{k,m}$ are significant.

Table 3. Clustering performances (%) on object categorization for selected objects in ETH-80

Methods	Performance			
	$\mathcal{V}_{1,192}$		$\mathcal{G}_{6,32-6}$	
	Exp-80	Exp-150	Exp-80	Exp-150
HSp-kM	82.52	82.96	n/a	n/a
HSp-EM	86.38	88.77	n/a	n/a
Ext-MS	100.00	99.77	82.08	88.00
Int-MS	100.00	99.11	80.42	89.34

4.3. Segmentation of Multiple Motions

Segmentation of multiple motions refers to the problem of identifying regions, *i.e.*, collection of image points, of consistent motions in video sequences. Consider the positions of a point tracked over F frames under the affine camera model and the feature vector of trajectories in $\mathbb{R}^{2F}$. Specifically, for the points that move with respect to the same motion, the corresponding feature vectors lie in a 4-dimensional subspace of $\mathbb{R}^{2F}$ defined by that motion. Thus, in the presence of multiple motions, each motion becomes a point on the Grassmann manifold. The trajectories of 4 points sharing the same motion are adequate to find the bases via the SVD.

In our experiments, we select 10 video sequences (see Figure 3) from the Hopkins 155 data set [25], which contains several sequences along with automatically extracted feature vectors. In order to cluster the point trajectories $\mathcal{A} = \{\boldsymbol{a}_t\}$ in a particular sequence into different motion classes, we need to first generate several motion candidates by randomly sampling 4 point trajectories $\{\boldsymbol{a}_n\}$ and forming the matrix $A = \begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3 & \boldsymbol{a}_4\end{bmatrix}$. We then compute the corresponding motion candidate as the U matrix of the SVD of $A = U\Sigma V^\top$. We generate 1,000 candidates and prune the poor ones via $\|(I_{2F} - UU^\top)\boldsymbol{a}_t\| > \tau, \forall \boldsymbol{a}_t \in \mathcal{A}\backslash\{\boldsymbol{a}_n\}$ for a low threshold τ.

Figure 3. Selected video sequences from the Hopkins 155 data set.

Table 4 shows the clustering performances of the Ext-MS (with bandwidth $h = 0.1$) and the Int-MS (with smoothing matrix $M = 0.1I_k$) in terms of segmentation accuracy and number of motions identified. The segmentation accuracy is quantified via the average and the maximum clustering rates (in parenthesis), whereas the accuracy in estimating the number of classes is quantified via the rate of estimating the correct number of motions over 20 trials. We observe that the Int-MS achieves higher average clustering rates in 8 cases. Specifically, it achieves its lowest and highest (average) performances at 64.71% and at 92.91%, respectively, whereas the Ext-MS achieves those performances at 57.29% and at 94.32%, respectively. In addition, in the cases when the average clustering rates are comparable, the Int-MS estimates the correct number of motions in all sequences with equal (in 3 cases) or higher (in 7 cases) identification rates than the Ext-MS. Notice also that the Ext-MS fails to identify the number of motions in one sequence (cars8) and achieves a low clustering rate of 57.29%, whereas the Int-MS successfully segments multiple motions in that sequence.

Table 4. Segmentation of multiple motions for selected sequences from Hopkins 155: Clustering rates (%) and percentages of trials that correctly identify the number of motions. The numbers in parenthesis in the first column are the true number of motions.

Sequence	Performance		Estimation	
	Ext-MS	Int-MS	Ext-MS	Int-MS
arm (2)	69.35 (100)	72.27 (100)	30	30
articulated (3)	69.83 (100)	75.50 (100)	20	25
cars1 (2)	79.93 (100)	77.00 (100)	20	20
cars2 (2)	88.10 (99.80)	90.92 (99.59)	15	35
cars4 (2)	78.40 (100)	88.06 (100)	85	85
cars5 (3)	80.06 (85.68)	80.59 (87.72)	25	30
cars6 (2)	94.32 (96.98)	92.91 (100)	5	30
cars8 (2)	57.29 (57.29)	64.71 (100)	0	25
truck1 (2)	71.44 (100)	86.76 (100)	60	90
2RT3RC (3)	87.48 (98.73)	92.60 (98.91)	80	95

5. Conclusions and Future Work

We have presented an alternative mean shift formulation that intrinsically performs unsupervised clustering of data points on analytic manifolds. Specifically, our algorithm employs a local density estimator based on generic kernels on the manifold of interest and locates the modes of the underlying distribution via iterative evaluations of a mapping. These evaluations constitute a modified gradient ascent scheme on Stiefel and Grassmann manifolds. We obtained promising results on object categorization and segmentation of multiple motions. In particular, even though hyperspherical versions of k-means and EM have an estimate for the number of groups in advance, the intrinsic mean shift outperforms these methods. For segmentation of multiple motions, we observe that the intrinsic method, apart from achieving higher clustering rates, outperforms its extrinsic counterpart in identifying the correct number of motions. Our future work includes using an adaptive smoothing matrix in the formulation and applying the method for the classification of dynamic textures.

Acknowledgements

The authors thank Dr. Onur C. Hamsici for providing various types of directional data. This work has been funded by Johns Hopkins WSE startup funds and by grants ONR N00014-05-10836 and NSF CAREER IIS-0447739.

References

[1] A. Banerjee, I. S. Dhillon, J. Ghosh, and S. Sra. Clustering on the unit hypersphere using von Mises-Fisher distributions. *Journal of Machine Learning Research*, 6:1345–1382, 2005.

[2] E. Begelfor and M. Werman. How to put probabilities on homographies. *IEEE Trans. on PAMI*, 27(10):1666–1670, 2005.

[3] S. Birchfield and S. Rangarajan. Spatiograms vs histograms for region-based tracking. In *CVPR*, volume II, pages 1158–1163, 2005.

[4] M. Carreira-Perpiñán. Gaussian mean-shift is an EM algorithm. *IEEE Trans. on PAMI*, 29(5):767–776, 2007.

[5] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Trans. on PAMI*, 17(8):790–799, 1995.

[6] Y. Chikuse. *Statistics on Special Manifolds*. Lecture Notes in Statistics. Springer, New York, 2003.

[7] D. Comaniciu and P. Meer. Mean Shift: A robust approach toward feature space analysis. *IEEE Trans. on PAMI*, 24(5):603–619, 2002.

[8] D. Comaniciu, V. Ramesh, and P. Meer. The variable bandwidth mean shift and data-driven scale selection. In *ICCV*, volume 1, pages 438–445, 2001.

[9] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *IEEE Trans. on PAMI*, 25(5):564–577, 2003.

[10] I. Dhillon and D. Modha. Concept decompositions for large sparse text data using clustering. *Machine Learning*, 42(1):143–175, 2001.

[11] A. Edelman, T. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal of Matrix Analysis Applications*, 20(2):303–353, 1998.

[12] A. Elgammal, R. Duraiswami, and L. Davis. Efficient kernel density estimation using the fast gauss transform with applications to color modeling and tracking. *IEEE Trans. on PAMI*, 25(11):1499–1504, 2003.

[13] M. Fashing and C. Tomasi. Mean shift is a bound optimization. *IEEE Trans. on PAMI*, 27(3):471–474, 2005.

[14] K. Fukunaga and L. Hostetler. The estimation of the gradient of a density function with application in pattern recognition. *IEEE Trans. on Information Theory*, pages 32–40, 1975.

[15] B. Georgescu, I. Shimshoni, and P. Meer. Mean shift based clustering in high dimensions: A texture classification example. In *ICCV*, volume 1, pages 456–463, 2003.

[16] H. Gong, C. Pan, Q. Yang, H. Lu, and S. Ma. A semi-supervised framework for mapping data to the intrinsic manifold. In *ICCV*, volume 1, pages 98–105, 2005.

[17] O. Hamsici and A. Martinez. Spherical-homoscedastic distributions: The equivalency of spherical and normal distributions in classification. *Journal of Machine Learning Research*, 8:1583–1623, 2007.

[18] B. Leibe and B. Schiele. Analyzing appearance and contour based methods for object categorization. In *CVPR*, volume 2, pages 409–415, 2003.

[19] X. Li, Z. Hu, and F. Wu. A note on the convergence of the mean shift. *Pattern Recognition*, 40:1756–1762, 2007.

[20] S. Oba, K. Kato, and S. Ishii. Multi-scale clustering for gene expression profiling data. In *IEEE Symposium on Bioinformatics and Bioengineering*, pages 210–217, 2005.

[21] S. Paris and F. Durand. A topological approach to hierarchical segmentation using mean shift. In *CVPR*, 2007.

[22] M. Park, Y. Liu, and R. Collins. Efficient mean shift belief propagation for vision tracking. In *CVPR*, 2008.

[23] Y. Sheikh, E. Khan, and T. Kanade. Mode-seeking by medoidshifts. In *ICCV*, 2007.

[24] R. Subbarao and P. Meer. Nonlinear mean shift for clustering over analytic manifolds. In *CVPR*, volume 1, pages 1168–1175, 2006.

[25] R. Tron and R. Vidal. A benchmark for the comparsion of 3-D motion segmentation algorithms. In *CVPR*, 2007.

[26] P. Turaga, A. Veeraraghavan, and R. Chellappa. Statistical analysis on Stiefel and Grassmann manifolds with applications in computer vision. In *CVPR*, 2008.

[27] O. Tuzel, R. Subbarao, and P. Meer. Simultaneous multiple 3D motion estimation via mode finding on Lie groups. In *ICCV*, volume 1, pages 18–25, 2005.

[28] A. Vedaldi and S. Soatto. Quick shift and kernel methods for mode seeking. In *ECCV*, pages 705–718, 2008.

[29] A. Yilmaz. Object tracking by asymmetric kernel mean shift with automatic scale and orientation selection. In *CVPR*, 2007.

[30] X. Yuan and S. Li. Half quadratic analysis for mean shift: with extension to a sequential data mode-seeking method. In *ICCV*, 2007.