GRASSMANNIAN DIMENSIONALITY REDUCTION USING TRIPLET MARGIN LOSS FOR UME CLASSIFICATION OF 3D POINT CLOUDS

Yuval Haitman[⋆] *Joseph M. Francos*[⋆] *Louis L. Scharf*[†]

[⋆]Electrical & Computer Engineering Department, Ben-Gurion University, Beer-Sheva, Israel
[†]Department of Mathematics, Colorado State University, Fort Collins, CO, USA

ABSTRACT

We consider the problem of classifying 3-D objects undergoing rigid transformations. It has been shown that the rigid transformation universal manifold embedding (RTUME) provides a mapping from the orbit of observations on some object to a single low-dimensional linear subspace of Euclidean space. This linear subspace is invariant to the geometric transformations. In the classification problem the RTUME subspace extracted from an experimental observation is tested against a set of subspaces representing the different object manifolds, in search for the nearest class. We elaborate on the design problem of the RTUME operator in the case where the point cloud sampled from the object is sparse, noisy, and non-uniformly sampled. By introducing metric learning and negative-mining techniques into the framework of Grassmannian dimensionality reduction for universal manifold embedding, we improve classification performance for these challenging sampling conditions.

1. INTRODUCTION

Detection and classification of whole objects or parts thereof are elementary building blocks in solving 3-D vision problems, from object classification, to part segmentation, to keypoint detection and matching for point cloud registration. However, an object to be detected may present itself under an unknown geometric transformation.

Universal manifold embedding (UME) [8, 14, 18], is a methodology for constructing a matrix representation of an observation such that it is covariant with the transformation, and then using this representation to identify a linear subspace that is invariant to affine coordinate transformations of the observation. This framework has been recently expanded to address the registration and classification of 3D point clouds related by a rigid transformation, [6, 9]. Practical application of the method for classification requires a high-quality estimate of the invariant subspace for each of K objects, and each of these subspaces must be estimated from one or more versions of an object, imperfectly observed in one or more of its poses.

This research was supported by NSF-BSF Computing and Communication Foundations (CCF) grants, CCF-2016667 and BSF-2016667.

Since the classifier performance depends on the choice of the set of functions composing the RTUME operator, we wish to design a set of functions, that best separates the UME matrix representation of each class (object) from those of the other classes, while minimizing the distance between observations from the same class. In this paper we address these issues by extending the theory of universal manifold embedding and Grassmannian dimensionality reduction, employing ideas from metric learning and negative mining to refine the Universal Manifold Embedding using Grassmannian Dimensionality Reduction based on Triplet Margin Loss (TL-GDRUME). In the presence of observation noise and non-uniform random sampling patterns of the point clouds TL-GDRUME is shown to provide better performance than existing methods.

2. PROBLEM FORMULATION

Consider a 3-D object $s \in \{s_1, \cdots, s_K\}$, and an *orbit* of equivalent objects formed by the action of the transformation group $G = SE(3)$. An observation $X(s_k)$ on object s_k will be denoted X_k and the set $\psi_k = \{\alpha \circ s_k, \alpha \in G\}$ will denote the manifold (orbit) of possible appearances of s_k turned out by the group G. There exists one such orbit for each object s_k. Our aim is to nonlinearly map each observation X_k, taken from the orbit ψ_k, to a matrix representation $\mathbf{T}(X_k)$. This matrix is to be linearly covariant with the parametrization of G; Its column space, which we denote by $\langle \mathbf{T}(X_k)\rangle$ is to be G-invariant. In other words, the orbit ψ_k is mapped into a linear subspace $\langle \mathbf{T}(X_k)\rangle$, such that the mapping is G-invariant, [14], [18]. That is, the set of *all* possible observations on an object under group action G is mapped by the UME operator into a *single* linear subspace which is invariant to the geometric transformation. Since the Special Euclidean group is a subgroup of the Affine group, in this paper we adapt this general framework for classification and detection of 3-D point clouds.

2.1. The RTUME Descriptor

The *universal manifold embedding* (UME) maps every observation $X \in \mathbb{R}^n$ from the orbit of s to a matrix $\mathbf{T}(X) \in$

$\mathcal{T}(M, n+1)$, such that $\mathbf{T}(X)$ is covariant with the geometric transformation, and where $\mathcal{T}(M, n+1)$ is the space of $M \times (n+1)$ real-valued matrices, and M the dimension of the embedding Euclidean space. The map $\mathcal{Q} : \mathcal{T}(M, n+1) \to \mathrm{Gr}(M, n+1)$, where $\mathrm{Gr}(M, n+1)$ is the Grassmann manifold of $n+1$-dimensional linear subspaces of M-dimensional Euclidean space, maps $\mathbf{T}(X)$ to its column space $\langle \mathbf{T}(X) \rangle$. Thus, the UME maps the orbit of s into the G-invariant subspace $\langle \mathbf{T}(X) \rangle \in \mathrm{Gr}(M, n+1)$.

For rigid transformations of 3-D objects, the Rigid Transformation UME (RTUME) [6] is a mapping of functions to matrices. It is covariant with rigid transformations of coordinates, *i.e.* , the RTUME matrices of functions on $\mathbb{R}^3$ related by a rigid transformation of coordinates are related by the same rigid transformation: Given a function $h : \mathbb{R}^3 \mapsto \mathbb{R}$, the RTUME matrix representation of $h(\mathbf{x})$ is given by

$$\mathbf{T}(h) = \begin{bmatrix} \int\limits_{\mathbb{R}^3} w_1 \circ h(\mathbf{x}) d\mathbf{x} & \int\limits_{\mathbb{R}^3} x_1 w_1 \circ h(\mathbf{x}) d\mathbf{x} & \dots & \int\limits_{\mathbb{R}^3} x_3 w_1 \circ h(\mathbf{x}) d\mathbf{x} \\ & & \vdots & \\ \int\limits_{\mathbb{R}^3} w_M \circ h(\mathbf{x}) d\mathbf{x} & \int\limits_{\mathbb{R}^3} x_1 w_M \circ h(\mathbf{x}) d\mathbf{x} & \dots & \int\limits_{\mathbb{R}^3} x_3 w_M \circ h(\mathbf{x}) d\mathbf{x} \end{bmatrix} \quad (1)$$

where w_i, $i = 1, ..., M$ are measurable functions aimed at generating many compandings of the observation.

Let $h(\mathbf{x})$ and $g(\mathbf{x})$ be two functions related by a rigid transformation of coordinates such that $h(\mathbf{x}) = g(\mathbf{R}\mathbf{x} + \mathbf{t})$. The RTUME matrices $\mathbf{T}(h)$ and $\mathbf{T}(g)$ constructed from $h(\mathbf{x})$ and $g(\mathbf{x})$ as in (1) are related as follows

$$\mathbf{T}(h) = \mathbf{T}(g)\mathbf{D}^{-1}(\mathbf{R}, \mathbf{t}) , \quad (2)$$

where $\mathbf{D}(\mathbf{R}, \mathbf{t})$ is given by

$$\mathbf{D}(\mathbf{R}, \mathbf{t}) = \begin{bmatrix} 1 & \mathbf{t}^T \\ \mathbf{0} & \mathbf{R}^{-1} \end{bmatrix} \quad (3)$$

Since $\mathbf{T}(h)$ and $\mathbf{T}(g)$ are related by an invertible transformation that is a re-expression of the rigid transformation relating the observations, we say that the basis $\mathbf{T}(g)\mathbf{D}^{-1}(\mathbf{R}, \mathbf{t})$ is *covariant* with the rigid transformation. Hence it provides a method for estimating the transformation that relates any two observations. Furthermore, since $\mathbf{T}(h)$ and $\mathbf{T}(g)$ are related by a right invertible linear transformation, the column space of $\mathbf{T}(g)$ and the column space of $\mathbf{T}(h)$ are identical. Their bases are different, but their range spaces are identical.

2.2. The Distance Between Equivalence Classes

In the RTUME framework the problem of detection and classification of geometrically deformed objects is formalized as follows: Given an observation Z, in the form of a sampled point cloud of an object whose geometric deformation is unknown, the problem is to determine whether $Z = \alpha \circ X$ or $Z = \beta \circ Y$, for some $\alpha, \beta \in G$, and X, Y some reference observations of known objects.

Since the detection and classification are to be G-invariant, we propose to compute $\mathbf{T}(Z)$ using (1) and measure the distance between the subspace $\langle \mathbf{T}(Z) \rangle$ and the subspaces $\langle \mathbf{T}(X) \rangle$ and $\langle \mathbf{T}(Y) \rangle$. That is, the observation Z is determined to belong to the orbit ψ_s if the distance from $\langle \mathbf{T}(Z) \rangle$ to $\langle \mathbf{T}(X) \rangle$, where X is some representative observation on object s, is smaller than its distance to $\langle \mathbf{T}(Y) \rangle$, where Y is some representative observation on a different object.

Following [7], [5] we compute the distance between a pair of subspaces as the extrinsic distance, evaluated using the projection Frobenius-norm

$$d_{pF}(\langle \mathbf{T}(Z) \rangle, \langle \mathbf{T}(X) \rangle) = \frac{1}{\sqrt{2}} ||\mathbf{P}_X - \mathbf{P}_Z||_F = ||\sin \boldsymbol{\theta}||_2 \quad (4)$$

where $\sin \boldsymbol{\theta}$ is a vector of sines of principal angles between the subspaces. The matrix $\mathbf{P}_X$ denotes the orthogonal projection matrix onto the subspace $\langle \mathbf{T}(X) \rangle$.

Since the classifier performance depends on the choice of the set of w-functions composing the operator $\mathbf{T}$, we wish to find the set of w-functions, that best separates the UME matrix representation of each class (object) from those of the other classes.

3. DESIGN OF THE RTUME OPERATOR

As point clouds are sets of coordinates in 3-D with no functional relation imposed on them, a necessary step in adapting the UME framework for point cloud processing is to define a function that assigns a value to each point in the cloud, invariant to the action of the transformation group. We call this step SE(3)-invariant coloring. (See examples in Fig. 1).

3.1. Representation by Level-Sets

Assume we are given an observation $X(\mathbf{u})$, $\mathbf{u} \in \mathbb{R}^3$, where, in general, $X(\mathbf{u})$ is evaluated from the raw point cloud measurements using an SE(3)-invariant function such as the distance of $\mathbf{u}$ from the object center of mass, or alternatively is provided by the measurement device, for example as an RGB measurement at $\mathbf{u}$. Further assume that the values of X are uniformly quantized at levels $\{q_i\}_{i=1}^{Q}$, so that it may be written as $X(\mathbf{u}) = \sum_{i=1}^{Q} q_i I_i^X(\mathbf{u})$ where $I_i^X(\mathbf{u})$ is the indicator function that equals 1 on the level-set of $\mathbf{u}$ where $q_{i-1} < X(\mathbf{u}) \leq q_i$, and zero elsewhere. Since $X(\mathbf{u})$ is SE(3)-invariant, the support of $I_i^X(\mathbf{u})$ for every i should be on the same surface points regardless of the rigid transformation the object has undergone.

The w operators must be designed such that the result of their application is covariant with the geometric transformation, and hence they are not functions of the coordinates. The action of w_m on X is simply to map the levels q_i into levels

$w_m(q_i)$, leaving the indicator functions I_i^X unchanged. Then, each term in the matrix $\mathbf{T}(X)$ may be written as

$$\begin{aligned}\mathbf{T}_{m,j} &= \int_{\mathbb{R}^3} w_m \circ X(\mathbf{u}) u_j d\mathbf{u} \\ &= \sum_{i=1}^{Q} w_m(q_i) \int_{\mathbb{R}^3} I_i^X(\mathbf{u}) u_j d\mathbf{u} = \sum_{i=1}^{Q} w_{m,i} F_{ij}^X \quad (5)\end{aligned}$$

where $w_{m,i} = w_m(q_i)$. This makes the moments $F_{ij}^X = \int_{\mathbb{R}^3} I_i^X(\mathbf{u}) u_j d\mathbf{u}$, the point cloud features of fundamental interest. Rewriting (5) in a matrix form we have

$$\mathbf{T}(X) = \mathbf{W}^T \mathbf{F}^X;\ \mathbf{W}^T = \{w_{m,i}\} \in \mathbb{R}^{M \times Q} \quad (6)$$

where $\mathbf{F}^X = \{F_{ij}^X\} \in \mathbb{R}^{Q \times 4}$ may be called the *Fundamental RTUME* representation matrix (FUME) for point cloud X. Since $M \leq Q$, the role of $\mathbf{W}$ is to transform the subspace $\langle \mathbf{F}^X \rangle \subset \mathrm{Gr}(Q, 4)$ to a subspace $\langle \mathbf{G}^X \rangle \subset \mathrm{Gr}(M, 4)$. A single $\mathbf{W}$ has to serve for all the orbits $\psi_1, \ldots, \psi_K$.

We have thus reduced the problem of finding an optimal set of RTUME representations for the K objects, to a problem of finding $\mathbf{W}$, where $\mathbf{W}$ maps the Fundamental RTUME matrix to a Grassmannian of lower dimension.

Grassmannian dimensionality reduction (GDR) is often employed in the framework of Grassmannian discriminant analysis (GDA), [19]. A method that employs Grassmannian optimization for GDR is the Projection Metric Learning (PML) on the Grassmann manifold [11]. PML learns a mapping $f : \mathrm{Gr}(D, q) \to \mathrm{Gr}(d, q)$ where $D > d$ by optimizing a discriminant function, designed to minimize the Grassmannian projection distances of within-class subspace pairs while maximizing the projection distances of between-class subspace pairs. Similarly, [12], employs a discriminant function with an affinity matrix for encoding the within-class and between-class information. This method is eventually reduced into finding a matrix on the Stiefel manifold. GDR for designing the UME operator, was first suggested in [9] for classification of uniformly and densely sampled point clouds.

4. RTUME DESIGN USING GRASSMANNIAN DIMENSIONALITY REDUCTION

We next find the set of w-functions, that best separates the RTUME matrix representation of each object from those of the other objects. It is assumed that we have a set of K objects, such that for each object N observations are available. Applying the RTUME operator (1), using a set of indictor functions on the level-sets of the quantization levels, *i.e.* , $\mathbf{W} = \mathbf{I}_Q$, we find the fundamental RTUME representation for each of the observations. Next, we wish to find an improved alternative for this choice of $\mathbf{W}$.

Thus, the problem of designing the optimal w-functions is the following: We are given a training set of N labeled point clouds (observations) for each one of K different objects, $\{X_i^k, i = 1, \ldots, N; k = 1, \ldots, K\}$, where observations of the same object differ by a rigid coordinate deformation, sampling pattern, and additive Gaussian noise. FUME matrix is generated for each observation $\{\mathbf{F}_i^k, i = 1, \ldots, N; k = 1, \ldots, K\}$ where $\langle \mathbf{F}_i^k \rangle \in \mathrm{Gr}(Q, 4)$. We are looking for a transformation matrix $\mathbf{W}^T \in \mathbb{R}^{M \times Q}$ that maps $\mathbf{F}_i^k$ to $\mathbf{Y}_i^k = \mathbf{W}^T \mathbf{F}_i^k$ where $\langle \mathbf{Y}_i^k \rangle \in \mathrm{Gr}(M, 4)$ and $\mathrm{Gr}(M, 4)$ is a Grassmann of a smaller ambient space. We denote by $\mathcal{P}$ the set of all positive pairs of FUME matrices, *i.e.* , pairs of FUME matrices generated from pairs of observations from the same orbit, that is $\mathcal{P} = \{\mathbf{F}_i^k, \mathbf{F}_j^k\}$. The complementary set of all negative pairs, *i.e.* , pairs of FUME matrices generated from pairs of observations from different orbits is defined by $\mathcal{N} = \{\mathbf{F}_i^k, \mathbf{F}_j^\ell, k \neq \ell\}$.

Inspired by the ideas of LPP, [10], and Grassmannian dimensionality reduction, [12, 18, 19], we next derive a companding of the level-set functions that maximizes classification accuracy. More specifically, we need to solve an optimization problem for $\mathbf{W}$ that will map FUME matrices belonging to the same orbit to close points on a Grassmannian of a reduced dimension ambient space, and map FUME matrices from different orbits to widely separated points. Following the techniques of metric learning and negative mining, [1, 4], we define the *triplet loss* function to be minimized, where the aim is to minimize the distance between positive pairs, such that the distance between negative pairs is at least a *margin* away. In particular, we employ negative mining so that the distance of the hardest-negative from each element in a positive pair is at least m, where the hardest-negative is the negative pair with minimal distance between the subspaces. Thus, we define the *hardest-negative triplet loss* as follows:

$$\begin{aligned} L(\mathbf{W}) \;=\; & \sum_{(i,j)\in\mathcal{P}} \Big\{ \Big[m + D_{i,j}(\mathbf{W}) - \min_{k\in\mathcal{N}} D_{i,k}(\mathbf{W}) \Big]_+ \\ & + \Big[m + D_{i,j}(\mathbf{W}) - \min_{k\in\mathcal{N}} D_{j,k}(\mathbf{W}) \Big]_+ \Big\} \quad (7) \end{aligned}$$

where $D_{i,j}(\mathbf{W}) = d_{pF}^2 (\langle \mathbf{Q}_i(\mathbf{W}) \rangle, \langle \mathbf{Q}_j(\mathbf{W}) \rangle)$. The orthogonal matrix $\mathbf{Q}_i(\mathbf{W})$ is determined by the *QR-decomposition* of $\mathbf{W}^T \mathbf{F}_i$ and $[\cdot]_+ = \max(0, \cdot)$. A typical value of m will be in the range of possible values of $D_{i,j}(\mathbf{W}) \in [0, 4]$.

We formulate the constrained optimization problem as

$$\min_{\mathbf{W} \in \mathbb{R}^{Q \times M}} \quad L(\mathbf{W}) \quad \text{subject to} \quad \mathbf{W^T W} = \mathbf{I}_M \quad (8)$$

Due to the constraint on $\mathbf{W}$ this optimization problem is solved over the Stiefel manifold $\mathrm{St}(Q, M)$. The constraint guarantees that we avoid degeneration to a trivial solution, and that with probability 1, for all i, the dimension of the subspace $\langle \mathbf{W}^T \mathbf{F}_i \rangle$ remains 4 despite the reduction in the dimensionality of the ambient space. To solve (8) we employ the framework of conjugate gradient minimization on the Stiefel manifold [2, 5, 15, 16], implemented using Manopt [3].

5. EXPERIMENTAL RESULTS

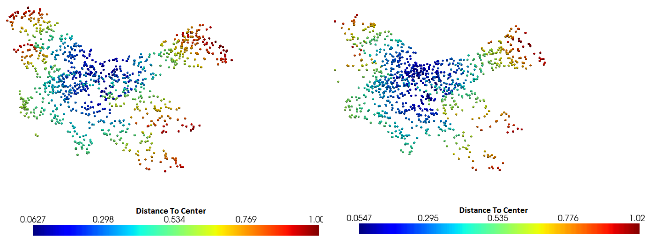

Fig. 1. SE(3)-invariant coloring example of a sparsely and uniformly sampled point cloud (left), and SE(3)-invariant coloring example of a sparsely and non-uniformly sampled point cloud of the same object (right).

We evaluate TL-GDRUME performance for object classification on the ModelNet40 dataset, and compare the results to those obtained by the FUME method derived in [9]. ModelNet40 contains CAD models of objects from 40 different classes given as triangular meshes. We used a point cloud representation of each object by sampling its mesh at 1024 points. Since one of the most difficult challenges in point cloud registration and classification is encountered when the point cloud is sparsely and non-uniformly sampled, and transformations are large, we evaluate the performance of the analyzed methods for large transformations, and for both uniform and non-uniform sparse sampling scenarios. We normalized each point cloud to the unit sphere to avoid potential scale differences between the models. A subset of ModelNet40 was employed, *i.e.* , we took 10 different models from each class. For each model we generated 200 different observations (100 for the training phase and 100 for testing), each with different rigid transformation, random sub-sampling and additive Gaussian noise. The rigid transformations are composed of a random roll and yaw in the range of $[0, 180]$ degrees and random pitch in the range of $[0, 90]$ degrees. Two noise levels are investigated: noise with std of 0.5 mesh resolution and noise with std of 0.8 mesh resolution (where mesh resolution of a point cloud is evaluated as the average distance between a point and its nearest neighbor). In order to employ the RTUME framework, we need to define on each point cloud an $SE(3)$-invariant function. However, since the two point clouds are differently sampled and noisy, every function which is $SE(3)$-invariant in the noise-free case, will be noisy, as well. See Fig. 1 for an illustrative example. In the experiments we employ the Euclidean distance from center of mass. In all the experiments we set the triplet-loss margin value to $m = 1$. For every point cloud, the FUME matrix is evaluated using (1) with $Q = 128$. The desired reduced dimension is $M = 90$. We use a nearest neighbor classifier between the test set and the training set, *i.e.* , the class of each observation in the test is determined to be the label of its nearest neighbor in the training set. We tested the classification

Sampling	Method	0.5 MR noise	0.8 MR noise
Uniform	FUME	0.85	0.83
	TLGDRUME	**0.93**	**0.91**
NonUniform	FUME	0.83	0.81
	TLGDRUME	**0.92**	**0.90**

Table 1. Accuracy comparison of FUME and TL-GDRUME on deformed ModelNet40 observations, uniformly and non-uniformly sampled, for large transformations, in the presence of noise.

performance under two different noise statistics. For each we also tested the effect of the sampling method (uniform/non-uniform) on the classifier performance. The results are evaluated using the estimated accuracy computed as the ratio of the number of correct decisions to the total number of trials. The results are summarized in Table 1. While the performance of base line methods such as PointNet, [13], deteriorates in the presence of large transformations, observation noise and non-uniform sampling (see, *e.g.* , [13, 17, 9]), both the FUME and TL-GDRUME methods achieve high accuracy, with TL-GDRUME consistently achieving better performance and robustness to the effects of noise and non-uniform sampling.

Finally, we evaluated the performance of TL-GDRUME as a function of the targeted reduced dimension of the Grassmannian, when the initial dimension of the Grassmannian is $Q = 128$. The performance of the TL-GDRUME for $M = 32, 64, 90$ is compared with the performance of the FUME evaluated for $Q = M$. In this test, noise std is 0.5 mesh resolution. The results indicate that performance gain of 2 to 4 percents is obtained for TL-GDRUME when Grassmannian dimension is reduced from $M = 90$ to $M = 32$, while no significant effect on the FUME performance is found.

6. CONCLUSIONS

We have presented a novel approach for designing the universal manifold embedding of 3D point clouds for optimizing detection and classification performance. In the presence of observation noise and non-uniform random sampling patterns, the observations do not lie strictly on the manifold and the resulting RTUME subspaces are noisy. The proposed method employs Grassmannian dimensionality reduction and metric learning to optimize the design of universal manifold embedding. It is shown that in the challenging conditions of non-uniform sampling, observation noise, and large transformations, the proposed design of the TL-GDRUME operator provides robust classification performance.

7. REFERENCES

[1] Vassileios Balntas, Edgar Riba, Daniel Ponsa, and Krystian Mikolajczyk. Learning local feature descriptors with triplets and shallow convolutional neural networks. In *BMVC*, 2016.

[2] Nicolas Boumal. An introduction to optimization on smooth manifolds. *Available online, Aug*, 2020.

[3] Nicolas Boumal, Bamdev Mishra, P-A Absil, and Rodolphe Sepulchre. Manopt, a matlab toolbox for optimization on manifolds. *The Journal of Machine Learning Research*, 15(1):1455–1459, 2014.

[4] Christopher Choy, Jaesik Park, and Vladlen Koltun. Fully convolutional geometric features. In *ICCV*, 2019.

[5] Alan Edelman, Tomás A Arias, and Steven T Smith. The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.

[6] Amit Efraim and Joseph M Francos. The universal manifold embedding for estimating rigid transformations of point clouds. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5157–5161. IEEE, 2019.

[7] Gene H Golub and Charles F Van Loan. Matrix computations. edition, 1996.

[8] Rami R Hagege and Joseph M Francos. Universal manifold embedding for geometrically deformed functions. *IEEE Transactions on Information Theory*, 62(6):3676–3684, 2016.

[9] Yuval Haitman, Joseph M Francos, and Louis L Scharf. Grassmannian dimensionality reduction for optimized universal manifold embedding representation of 3d point clouds. In *ICCV*, 2021.

[10] Xiaofei He and Partha Niyogi. Locality preserving projections. *Advances in neural information processing systems*, 16(16):153–160, 2004.

[11] Zhiwu Huang, Ruiping Wang, Shiguang Shan, and Xilin Chen. Projection metric learning on grassmann manifold with application to video based face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 140–149, 2015.

[12] Tianci Liu, Zelin Shi, and Yunpeng Liu. Joint normalization and dimensionality reduction on grassmannian: a generalized perspective. *IEEE Signal Processing Letters*, 25(6):858–862, 2018.

[13] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017.

[14] Ran Sharon, Joseph M Francos, and Rami R Hagege. Geometry and radiometry invariant matched manifold detection. *IEEE Transactions on Image Processing*, 26(9):4363–4377, 2017.

[15] Steven Thomas Smith. *Geometric optimization methods for adaptive filtering*. Harvard University, 1993.

[16] Steven T Smith. Optimization techniques on riemannian manifolds. *Fields institute communications*, 3(3):113–135, 1994.

[17] Chenxi Xiao and Juan Wachs. Triangle-net: Towards robustness in point cloud learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 826–835, 2021.

[18] Ziv Yavo, Yuval Haitman, Joseph M Francos, and Louis L. Scharf. Matched manifold detection for group-invariant registration and classification of images. *IEEE Transactions on Signal Processing*, 69:4162–4176, 2021.

[19] Jiayao Zhang, Guangxu Zhu, Robert W Heath Jr, and Kaibin Huang. Grassmannian learning: Embedding geometry awareness in shallow and deep learning. *arXiv preprint arXiv:1808.02229*, 2018.