Detection and Recognition of Deformable Objects Using Structured Dimensionality Reduction

Ran Sharon, Rami R. Hagege and Joseph M. Francos
Electrical & Computer Engineering Department
Ben Gurion University, Israel

***Abstract*—We present a novel framework for detection and recognition of deformable objects undergoing geometric deformations. Assuming the geometric deformations belong to some finite dimensional family, it is shown that there exists a set of nonlinear operators that universally maps each of the different manifolds, where each manifold is generated by the set all of possible appearances of a single object, into a unique linear subspace. In this paper we concentrate on the case where the deformations are affine. Thus, all affine deformations of some object are mapped by the above universal manifold embedding into the same linear subspace, while any affine deformation of some other object is mapped by the above universal manifold embedding into a different subspace. It is therefore shown that the highly nonlinear problems of detection and recognition of deformable objects can be formulated in terms of evaluating distances between linear subspaces. The performance of the proposed detection and recognition solutions is evaluated in various settings.**

I. Introduction

Solutions to many problems in image and signal analysis have to cope with the effects of the multiplicity of appearances of objects. In the problem of object recognition the "same" object may have a huge family of different appearances, and the first problem one needs to confront with, is the understanding of the set of all possible appearances of that *single* object. One of the main reasons for the huge variability in the appearance of an object is due to changes in its underlying geometry.

In general, we are given a set of observations (for example, images) of different objects, each undergoing a different geometric deformation. As a result of the action of the deformations, the set of different realizations for each object is generally a manifold in the space of observations. Therefore, the recognition problem is strongly related to the problems of manifold learning and dimensionality reduction of high dimensional data that have attracted considerable interest in recent years, see *e.g.*, [6]. The common underlying idea unifying existing manifold learning approaches is that although the data is sampled and presented in a high-dimensional space, for example because of the high resolution of the camera sensing the scene, in fact the intrinsic complexity and dimensionality of the observed physical phenomenon is very low.

As indicated in [7] linear methods for dimensionality reduction such as PCA and MDS generate faithful projections when the inputs are mainly confined to a *single* low dimensional subspace, while they fail in case the inputs lie on a manifold. Hence, one of the dominant approaches among existing dimensionality reduction methods is to expand the principles of the linear spectral methods to more complex low-dimensional structures than a single subspace, by assuming the existence of a smooth and invertible isometric mapping from the original manifold to some other manifold which lies in a lower dimensional space, [1]-[3]. These dimensionality reduction methods make very modest assumptions on the reasons for the variability in the appearances of the object. The common general assumption is that the degrees of freedom act continuously on the objects and therefore the set of appearances of a single object is some continuous entity – the manifold. Additional assumptions on its smoothness and local properties are well explored [4], [5]. As a result of the very mild assumptions made, the only way to determine the structure of the manifold generated by a single object is to densely sample it such that any other appearance of the object can be approximated locally and linearly by the collected samples. In many cases this implies the collection of a very large number of samples. An additional family of widely adopted methods aims at piecewise approximating, the manifold or a set of manifolds, as a union of linear subspaces [10] in what is known as the subspace clustering problem. A different assumption, namely that the data has a sufficiently sparse representation as a linear combination of the elements of an *a-priori* known basis or of an over-complete dictionary [8], [9] leads to the framework of linear dictionary approximations of the manifolds, as well as to the widely used framework of compressed sensing, [11].

Indeed, there are many cases where no prior knowledge on the reasons for the variability in the appearances of an object is available. On the other hand, there are many scenarios in which such information is inherently available, and hence can be efficiently exploited. One of the simplest examples is the case of a three dimensional object undergoing rigid motions in space. Here, one clearly knows the source of the variability, and this knowledge can be exploited in order to understand the structure of the manifolds before any sample is being collected. In this work we present a method that exploits this type of a priori knowledge in order to enable efficient detection and recognition of *multiple and deformable* objects.

II. Object Recognition Under Affine Deformations: Problem Definition

The basic problem addressed in this paper is the following: Given two bounded, Lebesgue measurable functions g, h with compact support such that $h : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}$, find if there is a matrix $\mathbf{A} \in GL_n(\mathbb{R})$ and a vector $\mathbf{c} \in \mathbb{R}^n$ such that:

$$g(\mathbf{x}) = h(\mathbf{A}\mathbf{x} + \mathbf{c}) \quad \forall \mathbf{x} \in \mathbb{R}^n \tag{1}$$

In the following, we show that the problem of determining whether such $\mathbf{A}, \mathbf{c}$ exist, whose direct solution requires a highly complex search in the product space of $GL_n(\mathbb{R}) \times \mathbb{R}^n$, can be formulated as a problem of classifying subspaces of a low dimensional linear space.

More specifically, let $\mathcal{O}$ be the space of observations (for example, images), let Q be the set of possible geometric deformations with N degrees of freedom, and let S be a set of known objects. We assume that the observations are constructed by the following procedure: we first choose an object $s \in S$ and an arbitrary geometric deformation $\varphi \in Q$. Next, we define an operator $\psi : S \times Q \rightarrow \mathcal{O}$ that acts on an object and a geometric deformation, producing an observation. The observation is $o = \psi(s, \varphi)$. For a specific object $s \in S$ we will denote by $\psi_s : Q \rightarrow \mathcal{O}$ the restriction of the map to this object. We assume that the N parameters characterizing Q are embedded in some linear space. For example, if Q is the set of functions describing invertible two dimensional affine deformations then Q is of dimension 6 and each geometric deformation is given by $\varphi(x, y) = (ax + by + c, dx + ey + f)$. For any object (function) $g \in S$ the set of all possible observations on this particular function is denoted by S_g. We refer to this subset as the orbit of g under Q. In general, since ψ is not linear, this subset is a non linear manifold in the space of observations. The orbit of each function forms a different manifold. Since in general $\mathcal{O}$ has a very high dimension (the number of pixels), one must find an accurate description of S_g in order to enable any further analysis of it. Dimensionality reduction methods rely on dense sampling of S_g to achieve this description using low dimensional patches. We next show that under the above assumptions and for some specific choices of Q there exists a map $T : \mathcal{O} \rightarrow H$ such that H is a linear space, which we call the *reduced space*. Moreover, the map $T \circ \psi_s : Q \rightarrow H$ is *linear* and *invertible*. These properties hold for every object $s \in S$ and the map T is *independent* of the object. We call such a map T, *universal manifold embedding* as it universally maps each of the different manifolds, where each manifold corresponds to a single object, into a *unique* linear subspace such that the overall map $T \circ \psi_s : Q \rightarrow H$ is linear. In other words, each manifold is uniquely mapped into a linear subspace of H. The map $\psi_s : Q \rightarrow \mathcal{O}$ maps Q non-linearly and represents the physical relation between the object and the observations on it. The map $T : \mathcal{O} \rightarrow H$, maps $\mathcal{O}$ non-linearly such that the overall map $T \circ \psi_s : Q \rightarrow H$ is linear. This universal map allows us to represent the (mapped) observation in a space where the action of Q is linear.

A universal manifold embedding T is a map from the space of observations into a low dimensional linear space, such that the set $T(S_g)$ is linearly dependent with a low dimensional basis for *any* $g \in S$ and the restriction of T to the manifold S_g, which we denote by $T|_{S_g}$, is invertible from S_g to $T(S_g)$. In those cases where the universal embedding $T : \mathcal{O} \rightarrow H$ exists, one can solve many problems concerning the multiplicity of appearances of an object directly in H using classical *linear theory*, instead of being forced to employ non-linear analysis. Thus for example, in order to characterize the mapped manifold of some object in the reduced linear space H all one needs to do is to take a few samples of its appearances on the linear subspace in H, and to find the desired manifold by describing this linear subspace. Figure 1 schematically illustrates the concept of the universal manifold embedding. We next describe the implementation of the concept of universal manifold embedding for the case where the geometric deformations are affine.

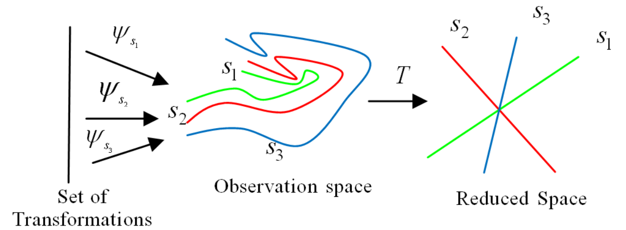

Fig. 1. Universal manifold embedding: Observations from the same manifold are mapped to the same subspace.

Consider the case where h is an observation of g undergoing an affine transformation. Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, *i.e.*, $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$, $\mathbf{y} = [y_1, y_2, \ldots, y_n]^T$, such that $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{c}$, $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y} + \mathbf{b}$. Let $\tilde{\mathbf{y}} = [1, y_1, \ldots, y_n]^T$. Thus, $\mathbf{x} = \mathbf{D}\tilde{\mathbf{y}}$ where $\mathbf{D}$ is an $n \times (n+1)$ matrix given by $\mathbf{D} = \left[\mathbf{b} \ \mathbf{A}^{-1}\right]$.

Let $m \in \mathbb{N}$ and let $w_l \ \ l = 1, \ldots, m$ be a set of bounded, Lebesgue measurable functions $w_l : \mathbb{R} \rightarrow \mathbb{R}$. Let $\mathbf{D}_k$ denote the kth row of the matrix $\mathbf{D}$. Then, [12],

$$\int_{\mathbb{R}^n} x_k w_\ell \circ h(\mathbf{x}) d\mathbf{x} = \left|\mathbf{A}^{-1}\right| \int_{\mathbb{R}^n} (\mathbf{D}_k \tilde{\mathbf{y}}) w_\ell \circ g(\tilde{\mathbf{y}}) d\tilde{\mathbf{y}} \tag{2}$$

Let f be some observation on a deformable object and let

$$\mathbf{T}(f) = \begin{bmatrix} \int_{\mathbb{R}^n} w_1 \circ f(\mathbf{y}) & \int_{\mathbb{R}^n} y_1 w_1 \circ f(\mathbf{y}) & \cdots & \int_{\mathbb{R}^n} y_n w_1 \circ f(\mathbf{y}) \\ \vdots & \ddots & \vdots & \\ \int_{\mathbb{R}^n} w_m \circ f(\mathbf{y}) & \int_{\mathbb{R}^n} y_1 w_m \circ f(\mathbf{y}) & \cdots & \int_{\mathbb{R}^n} y_n w_m \circ f(\mathbf{y}) \end{bmatrix} \tag{3}$$

Denote $\tilde{\mathbf{D}} = [\mathbf{e}_1 \ \ \mathbf{D}^T]$ where $\mathbf{e}_1 = [1, 0, \ldots, 0]^T$. Then, if h is an observation of g undergoing an affine deformation represented by the matrix $\mathbf{D}$, then from (2) we get:

$$\mathbf{T}(g) \left|\mathbf{A}^{-1}\right| \tilde{\mathbf{D}} = \mathbf{T}(h) \tag{4}$$

Since the deformations at hand are invertible, this implies that the column space of $\mathbf{T}(g)$ and the column space of $\mathbf{T}(h)$ are the same subspace. Thus, after applying the mapping $\mathbf{T}$ to the space of observations, the problem of object recognition becomes a problem of classifying subspaces.

III. Classifying Subspaces

Measuring the distance between two subspaces of a larger subspace is a well explored problem. One way of measuring distance between subspaces is by principal angles. Let $\mathbf{A} \in \mathbb{R}^{m \times p}$ and $\mathbf{B} \in \mathbb{R}^{m \times q}$ be real matrices with the same number of rows and assume for convenience that $\mathbf{A}, \mathbf{B}$ have full column rank and that $p \geq q$. We denote the range (column space) of $\mathbf{A}$ by $range\left(\mathbf{A}\right)$.

The q principal angles $\theta_k \in \left[0, \frac{\pi}{2}\right]$ between $range\left(\mathbf{A}\right)$ and $range\left(\mathbf{B}\right)$ are recursively defined for $k = 1, 2, \ldots, q$ as

$$\begin{aligned} \cos\left(\theta_k\right) &= \max_{\substack{x \in \mathbb{R}^p \\ y \in \mathbb{R}^q}} \frac{\left|x^T \mathbf{A}^T \mathbf{B} y\right|}{\left\|\mathbf{A}x\right\|_2 \left\|\mathbf{B}y\right\|_2} \\ &= \frac{\left|x_k^T \mathbf{A}^T \mathbf{B} y_k\right|}{\left\|\mathbf{A}x_k\right\|_2 \left\|\mathbf{B}y_k\right\|_2} \end{aligned} \quad (5)$$

subject to $x_i^T \mathbf{A}^T \mathbf{A} x_k = 0$ and $y_i^T \mathbf{B}^T \mathbf{B} y_k = 0$ for $i = 1, \ldots, k-1$. Note that all of the principal angles are in the interval $[0, \frac{\pi}{2}]$

A. Using Principal Angles for Classification

Since principal angles measure how "close" two subspaces are to each other, we employ them in the universal manifold embedding framework to distinguish between two hypotheses: Given two observations on affine deformed objects, one hypothesis is that the observations are on the same object (regardless of what the object is) while the alternative hypothesis is that the observations come from different objects. More specifically we wish to design a linear classifier to discriminate between the distance of a deformed object from other deformed observations on it, and its distances from observations on other objects. In order to design the classifier we first collect observations of the different objects. We next calculate the empirical distributions of the principal angles when any two observations are on the same object, and when any two observations are on different objects. These distributions are then employed to design a linear classifier (discriminant function), as illustrated in the next section.

B. Projection Matrices for Subspace Classification

Another way for classifying subspaces is to use the fact that projection matrices have a one-to-one correspondence to subspaces. That is, given two matrices $\mathbf{A}, \mathbf{B}$ whose column spaces are the same, we get that the projection matrix onto the column space of $\mathbf{A}$ is identical to the projection matrix onto the column space of $\mathbf{B}$. This enables us to measure the distance between subspaces by measuring the norm of the difference between the projection matrices onto the different object subspaces.

IV. Classification of Deformed Objects

In this section we provide a concrete example detailing how the concept of deformed object classification using the universal manifold embedding can be implemented in practice. Consider a set of non-linear operators w_i $i = 1, \ldots, I$, (as defined in (2)) such that:

$$w_i\left(x\right) = \begin{cases} 1 & x \in \left(\frac{i-1}{I}, \frac{i}{I}\right] \\ 0 & x \notin \left(\frac{i-1}{I}, \frac{i}{I}\right] \end{cases} \quad (6)$$

with $I = 10$. The collection of objects for the classification test is a set of 250 images of objects. In the experiment, one object was chosen at random and 1000 observations of the object were created (using random affine deformations). Additional 1000 observations of other objects were also generated. All of the observations were of resolution 400×400. The operator $\mathbf{T}(f)$ defined in (3) was applied to each of the 2000 generated images in order to map each observation into a subspace. Note that since images are two-dimensional, the column space of $\mathbf{T}(f)$ is of dimension 3.

A. Classification Based on the Frobenius Norm Between Projection Matrices

In this implementation we consider two types of distances: One is the set of pairwise distances between all the subspaces corresponding to the observations. The second is the set of distances between a subspace and its nearest subspace.

For each of the 1000 observations on the deformed target object (each under a different deformation), we record its distance from all other observations on the target (meaning one measurement for each pair of observations, and a total of $\frac{1000 \cdot 999}{2}$ measurements). Also, for each observation on the target object we record its minimal distance from all of the other 999 observations on the target (1000 measurements in total). Similarly, for each observation on a non-target object, we collected its distance from all of the 1000 observations of the target object (a total of $1000 \cdot 1000$ measurements) and also its minimal distance from observations on the target object (1000 measurements in total). Here, we used the Frobenius norm of the difference between projection matrices to measure distance between subspaces. The experiment was repeated 20 times for randomly selected objects from the set of 250 objects. Figure 2 shows a typical result from this experiment.

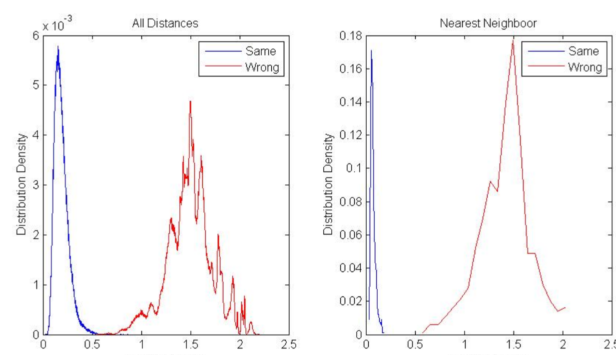

Fig. 2. Frobenius norm distribution. Blue: Distribution where each pair of observations is on deformations of the same object. Red: Distribution where each pair of observations is on deformations of different objects.

It can be seen that while there is a small overlap between the distributions when taking all of the pairwise distances into consideration, when using the distance to the nearest subspace, the supports of the distributions are disjoint and perfect classification is achievable.

B. Principle Angles Based Classification

In Section III-A we described the approach of directly using principle angles for subspace classification. Next, we adapt this approach for the classification of observations on deformed objects, using the universal manifolds embedding. In this experiment we evaluate the performance of a classifier that employs principal angles between the subspaces computed by the universal manifold embedding of each image. The results are compared to those of an alternative procedure where we estimate the affine deformation that relates the observations, apply it to the appropriate image and calculate the 2-norm of the difference image. An observation was created from an object and then its distances (using both methods described above) were calculated from both its own template and from those of the other objects. Figure 3 shows scatter plots of each of the principal angles in comparison with the 2-norm distance between the observation and the estimated deformed template. It can be seen that although classification is possible using both methods, the principal angles method is more stable.

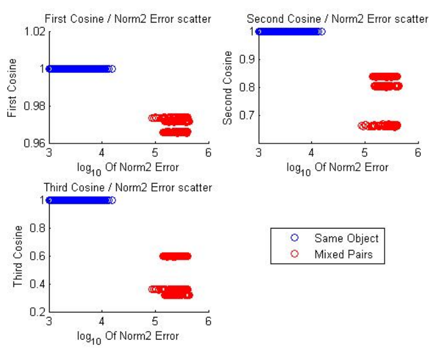

Fig. 3. Classification based on evaluating the principal angles in comparison with a direct deformation estimation

V. Detection of a Deformed Object

In this section, we evaluate the performance of the proposed approach in the context of an object detection problem, where the observation is an unknown affine deformation of the object of interest. In real-world applications, the object we are trying to detect is surrounded by background. Thus, the object detector is implemented by sliding a window over the observed image and testing for each position of the window for the existence of the object, using the universal manifold embedding approach described above. Clearly, as the amount of pixels in the observation that do not belong to the object

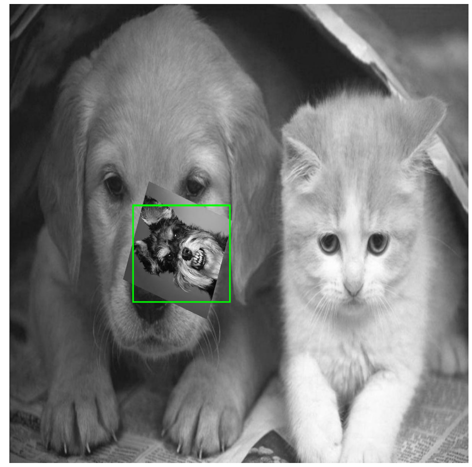

Fig. 4. Object embedded in background, and the detection result (green square): window position with minimal distance to the target template

decreases, the background effect declines, and vice-versa. In each test, two objects from the set of 250 objects were chosen at random. One was used as background and the other one was deformed by a random affine deformation and placed on the background image (see for example, Figure 4). For each position of the sliding window, the distance from the image bounded by the window to the known template was measured. The output of the procedure is the bounding window that bounds the square image patch that is closest to the template. The Frobenius norm between the projection matrices was used to measure distance between subspaces, with the nonlinear operators defined as in (6).

VI. Conclusions

We have presented a novel framework for detection and recognition of deformable objects undergoing geometric deformations. Assuming the geometric deformations belong to some finite dimensional family, it is shown that there exists a set of nonlinear operators that universally maps each of the different manifolds, where each manifold is generated by the set all of possible appearances of a single object, into a unique linear subspace. It has been shown that the highly nonlinear problems of detection and recognition of deformable objects can be formulated in terms of evaluating distances between linear subspaces. Two classification schemes, one based on evaluating the principal angles between the subspaces generated by the universal manifold embedding of the objects to be classified, and one based on evaluating the Frobenius norm between the projection matrices on these subspace have been derived and their performance has been experimentally demonstrated. Similar recognition and detection procedures, to those derived here for the case where the deformations are affine, can be adopted when dealing with elastic deformations, using the universal manifold embedding for the framework derived in [13].

References

[1] P. Dollar, V. Rabaud and S. Belongie, "Learning to Traverse Image Manifolds", *Proc. NIPS*, 2006.

[2] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," *Science*, vol. 290, pp. 2319–2323, 2000.

[3] S. T. Roweis and L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, vol. 290, pp. 2323-2326, 2000.

[4] C. Walder and B. Schölkopf, "Diffeomorphic Dimensionality Reduction" *Proc. NIPS*, 2008.

[5] Z. Zhang and H. Zha, "Principal Manifolds and Nonlinear Dimension Reduction via Local Tangent Space Alignment," *SIAM Jou. Sci. Comp.* vol. 26 , pp. 313 - 338, 2005.

[6] "Special Issue on Dimensionality Reduction Methods," *Signal Process. Mag.*, Mar. 2011.

[7] K. Weinberger and L. Saul, "Unsupervised Learning of Image Manifolds by Semidefinite Programming," *Int. J. Comput. Vision*, pp. 77-90, 2006.

[8] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Info. Theory*, vol. 50, pp. 2231-2242, 2004.

[9] I. Tosic and P. Frossard, "Dictionary Learning," *Signal Process. Mag.*, pp. 27-38, Mar. 2011.

[10] R. Vidal, "Subspace Clustering," *Signal Process. Mag.*, pp. 52-67, Mar. 2011.

[11] R. Baraniuk and M. Wakin, "Random Projections of Smooth Manifolds," *Found. Comput. Math.*, pp. 51-77, 2009.

[12] R. Hagege and J. M. Francos, "Parametric Estimation of Affine Transformations: An Exact Linear Solution," *Jou. of Math. Imaging and Vision*., pp. 1-16, Jan. 2010.

[13] R. Hagege and J. M. Francos, "Linear Estimation of Time Warped Signals," *IEEE Trans. Info. Theory*, pp. 4423 - 4439, July 2013.