Geometry and Radiometry Invariant Matched Manifold Detection

Ran Sharon, Joseph M. Francos, and Rami R. Hagege

***Abstract*— Consider a set of deformable objects undergoing geometric and radiometric transformations. As a result of the action of these transformations, the set of different realizations of each object is generally a manifold in the space of observations. Assuming the geometric deformations an object undergoes, belong to some finite dimensional family, it has been shown that the universal manifold embedding (UME) provides 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 distinct linear subspace of an Euclidean space. In this paper, we generalize this framework to the case where the observed object undergoes both an affine geometric transformation, and a monotonic radiometric transformation, and present a novel framework for the detection and recognition of the deformable objects. Applying to each of the observations an operator that makes it invariant to monotonic amplitude transformations, but is geometry-covariant with the affine transformation, the set of *all* possible observations on that object is mapped by the UME into a *single* linear subspace—invariant with respect to both the geometric and radiometric transformations. The embedding of the space of observations is independent of the specific observed object; hence it is universal. The invariant representation of the object is the basis of a matched manifold detection and tracking framework of objects that undergo complex geometric and radiometric deformations: the observed surface is tessellated into a set of tiles such that the deformation of each one is well approximated by an affine geometric transformation and a monotonic transformation of the measured intensities. Since each tile is mapped by the radiometry invariant UME to a distinct linear subspace, the detection and tracking problems are solved by evaluating distances between linear subspaces. Classification in this context becomes a problem of determining which labeled subspace in a Grassmannian is closest to a subspace in the same Grassmannian, where the latter has been generated by radiometry invariant UME from an unlabeled observation.**

***Index Terms*— Distance learning, image analysis, matched filters, object detection.**

I. INTRODUCTION

ANALYZING and understanding different appearances of an object is an elementary problem in various fields. Since acquisition conditions vary (*e.g.*, pose, illumination), the set of possible observations on a particular object is immense. We consider a problem where, in general, we are given a set of observations (for example, images) of different objects, each undergoing different geometric and radiometric deformations. As a result of the action of the deformations, the set of different realizations of each object is generally a manifold in the space of observations. Therefore, the detection and recognition problems are 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.*, [7]. 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.

The problem of characterizing the manifold created by the multiplicity of appearances of a *single* object in some general setting is studied intensively in the field of non linear dimensionality reduction. As indicated in [6], linear methods for dimensionality reduction such as PCA and MDS generate the desired projections when the observations are mainly confined to a *single* low dimensional linear subspace, while they fail in case the inputs lie on a low dimensional non-linear manifold. Hence, a common approach among existing non-linear dimensionality reduction methods is to expand the principles of the linear spectral methods to low-dimensional structures that are more complex than a single linear subspace. This is achieved, for example, by assuming the existence of a smooth and invertible locally isometric mapping from the original manifold to some other manifold which lies in a lower dimensional space, [1]–[3].

An additional family of widely adopted methods aims at piecewise approximating the manifold or a set of manifolds, as a union of linear subspaces, in what is known as the subspace clustering problem, see [20], [13], and the references therein. The challenge here is to simultaneously cluster the data into multiple linear subspaces and to fit a low-dimensional linear subspace to each set of observations. 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], [12] leads to the framework of linear dictionary approximations of the manifolds. Geometrically, this assumption implies that the manifold can be well approximated by its tangent plane, with the quality of this approximation depending on the local curvature of the manifold.

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. In [29], we presented an alternative

Manuscript received June 12, 2016; revised January 8, 2017; accepted April 13, 2017. Date of publication April 27, 2017; date of current version July 6, 2017. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Christopher Wyatt. *(Corresponding author: Joseph M. Francos.)*

The authors are with the Department of Electrical and Computer Engineering, Ben-Gurion University, Beer Sheva 84105, Israel (e-mail: francos@ee.bgu.ac.il).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2017.2698921

to the direct methods for learning the manifold, that is both natural to the problem as it exploits the available a-priori knowledge of the type of expected deformations, and is computationally highly efficient. We concentrated on the case where the geometric deformations are the major source for the variability in the appearances of the object. Assuming the geometric deformations an object undergoes, are invertible and belong to some known finite dimensional family, it has been shown, that the universal manifold embedding (UME) 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 *distinct* linear subspace of a low dimensional vector space. The universal manifold embedding is implemented by constructing a set of non linear functionals. As such, the mapping itself is nonlinear, and no local linear approximations of the manifold are involved. The universal manifold embedding provides an *exact* characterization of the manifold in contrast with existing dimensionality reduction methods in which local approximations of the manifold structure are produced. The evaluation of the universal manifold embedding for each object requires the knowledge of the group of transformations it undergoes and a single observation on the object.

In this paper we expand the framework of the universal manifold embedding, [29], to the more general case where the observed object undergoes not only geometric transformations, but also radiometric transformations. More precisely, in [29] it is assumed that the only source for the variability in the object appearance is the geometric deformations the object undergoes, while the measured image intensities remain unchanged. In practice, this is rarely the case. Therefore in this paper we consider the more realistic case where the observed object undergoes both an affine geometric transformation, and a monotonic radiometric transformation and prove by construction the existence of a radiometry invariant universal manifold embedding (RIUME). This approach is motivated by the observation that almost any observed surface can be tessellated into a set of tiles such that the deformation of each one is well approximated by an affine geometric transformation and a monotonic transformation of the measured intensities.

As could be expected from the fundamental properties of the UME operator, in the presence of radiometric transformations the geometry-induced low-dimensional manifold model for the set of possible observations on an object, [29], becomes over simplified. Thus, in order to ensure the applicability of the geometry-induced manifold model, the observations must be projected onto some "canonic" manifold by finding a transformation that makes the observation invariant to radiometric deformations while being covariant with the geometric transformation. (This "canonic" manifold can be thought of as the manifold spanned by the set of all possible observations generated by the action of the affine group on some template observation of the object). The resulting framework provides an exact description of the manifold despite using as low as a single observation. Hence the need for using large amounts of observations in order to learn the manifold or a corresponding dictionary is eliminated. This in turn, makes the method especially attractive for tracking problems, where in general, no prior observations of the object are available. Moreover, the proposed radiometry invariant universal manifold embedding does not involve any discretization of the model, nor local approximations of the manifold, as the parametrization of the manifold remains in the continuum. These new results lead to the derivation of a *radiometry-geometry invariant matched manifold detector* and consequently of a *radiometry-geometry invariant matched manifold tracker*.

It is shown in this paper that employing the RIUME provides a different type of information than existing point matching algorithms on the one hand, or global registration algorithms, on the other hand. Point matching algorithms aim at finding key points in the observed image and characterize them through the properties of small regions around them. These inherently local approaches use relatively small amounts of information (small patches) in generating the descriptor of a key point. As a consequence they frequently result in ambiguous descriptors, which in turn lead to high rates of false matches that need to be eliminated before any further processing can take place. Such verification procedures require knowledge of the global geometric transformation model (*e.g.*, RANSAC) and are computationally demanding. Moreover, such global geometric transformation model is often unknown, or difficult to estimate. On the other end, global registration algorithms may be applied only when the family of expected geometric deformations is a-priori known, and the radiometric deformations between the two observations are minimal. The RIUME based matched manifold detection scheme provides a method for efficiently combining the advantages of the local, key point based methods, and the global methods. This is achieved by employing the above mentioned observation that almost any surface can be well approximated by its tessellation into tiles, such that two observations on the same tile are related by simultaneous affine transformation of coordinates and a monotonic mapping of the intensities. Thus, the combinatorial complexity of RANSAC-like methods is eliminated by efficiently testing larger patches, defined by the triangulation procedure between key points, to their similarity to the assumed matching triangular tiles in the other image using the RIUME induced distance metric between linear subspaces. Since each such triangle is much larger than the small neighborhoods used to extract the key point descriptors, the matching procedure of triangular tiles is much more accurate than any procedure for matching local key point descriptors.

In [30] an early and short description of the concept of radiometry-geometry invariant matched manifold detection is presented.

The structure of this paper is as follows: In Section II we provide the basic definitions and the scope of the proposed radiometry invariant universal manifold embedding. Then, in Section III we extend the previously developed framework of the UME, [29], and prove by construction the existence of a radiometry invariant universal manifold embedding in the case where the set of possible geometric transformations the objects may undergo is the set of affine transformations, and the radiometric deformations are monotonic. In Section IV we employ the geometry and radiometry invariance provided by the radiometry invariant universal manifold embedding in order

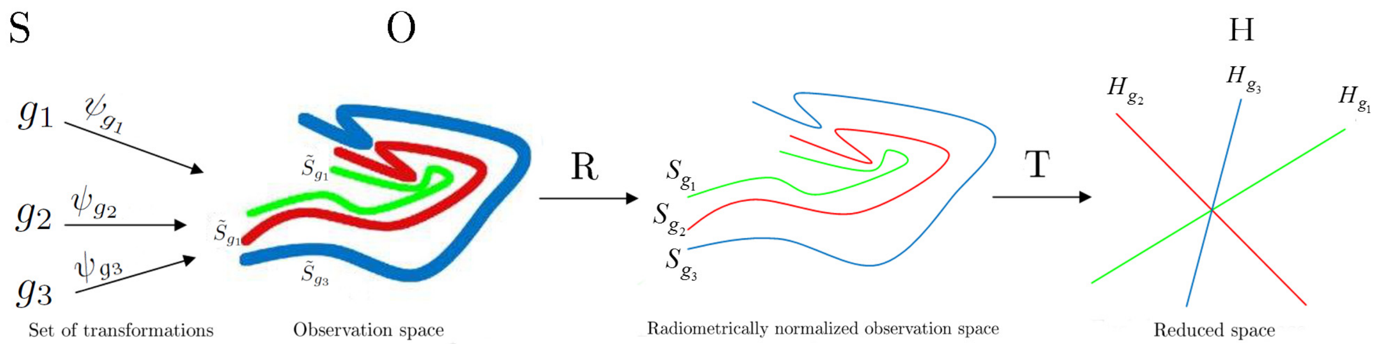

Fig. 1. The Radiometry Invariant Universal Manifold Embedding framework (from left to right): The physical model that generates the observations - applying the set of possible radiometric and geometric deformations to some object g produces $\tilde{S}_g$ which is the set of all possible observations on g. $\tilde{S}_g$ is a subset of the space of observations O. The operator R maps each manifold to a radiometry normalized manifold, S_g. The UME - all observations in S_g are nonlinearly mapped by T to a single linear subspace $H_g = T(S_g)$.

to define and evaluate distances between object manifolds. In Section V we detail several experiments that illustrate the strength of the radiometry invariant universal manifold embedding and help understand the interplay between its different parameters. Finally, in Section VI the derived RIUME is employed as the basis of a matched manifold detection and tracking framework of objects that undergo complex geometric and radiometric deformations: The observed surface is tessellated into a set of tiles such that the deformation of each one is approximated by an affine geometric transformation and a monotonic transformation of the measured intensities.

II. Problem Definition

Let us begin by informally stating the problem studied in this paper. Suppose we are given two observations g and h, on the same object. The two observations are related by an affine transformation of the geometry of the object and a monotonic transformation of the measured amplitudes, *i.e.*,

$$h(\mathbf{x}) = U(g(\mathcal{A}(\mathbf{x}))), \tag{1}$$

where U is strictly monotonic and $\mathcal{A}$ is affine. The right-hand composition of g with $\mathcal{A}$ represents the spatial affine deformation, while the left-hand composition with U represents the radiometric transforation applied to the signal's amplitude.

More precisely, let $\mathcal{O}$ be the space of observations (for example, images), let $\tilde{\mathcal{A}}$ be the set of possible affine deformations of the object geometry, $\mathcal{U}$ the set of monotonic one-dimensional transformations, and let S be a set of known objects, where each object is modeled as a compact support, bounded, and Lebesgue measurable (or more simply, integrable) function from $\mathbb{R}^n$ to $\mathbb{R}$.

We assume that the observations are the result of the following procedure: We choose an object $g \in S$ and some geometric-radiometric transformation pair $(\mathcal{A}, U)$ in $\tilde{\mathcal{A}} \times \mathcal{U}$. An operator $\psi : S \times \tilde{\mathcal{A}} \times \mathcal{U} \to \mathcal{O}$ acts on the chosen object g such that it jointly undergoes an affine geometric deformation $\mathcal{A}$, and a monotonic radiometric transformation U, producing an observation such as h (or g, in case both the radiometric and geometric deformations are the identity transformation) above. For a specific object $g \in S$ we will denote by ψ_g the restriction of the map to this object. For any object (function) $g \in S$ the set of all possible observations on this particular function is denoted by $\tilde{S}_g$. We refer to this subset as the orbit of g under the direct product $\tilde{\mathcal{A}} \times \mathcal{U}$. In general, this subset is a non linear manifold in the space of observations. The orbit of each function forms a different manifold.

The space $\mathcal{O}$ has a very high dimension (*e.g.*, the number of pixels in an image). It is composed of the union of orbits, $\tilde{S}_g$, of the different objects g in S such that each orbit $\tilde{S}_g$ is the result of the action of $\tilde{\mathcal{A}} \times \mathcal{U}$ on the object g. As indicated earlier, existing non-linear dimensionality reduction methods rely on dense sampling of $\tilde{S}_g$ to obtain its description using local linear approximations, or alternatively, provide only an approximate description of the manifold.

In this paper we show, by construction, that under the above assumptions there exists a linear space H which we call the *reduced space*, and a pair of maps $R : \mathcal{O} \to \mathcal{O}$ and $T : \mathcal{O} \to H$ such that the composite map $T \circ R \circ \psi_g : \tilde{\mathcal{A}} \times \mathcal{U} \to H$ is invariant to the action of $\mathcal{U}$ and is linear in the parameters defining $\tilde{\mathcal{A}}$. This construction holds for every object $g \in S$, and any monotonic transformation of its amplitudes. The maps T and R are *independent* of the object. We call the map $L = T \circ R$, *radiometry invariant universal manifold embedding* (RIUME) as it universally maps each of the different manifolds, where each manifold corresponds to a single object, into a *different* linear subspace of H. The map $\psi_g : \tilde{\mathcal{A}} \times \mathcal{U} \to O$ maps $\tilde{\mathcal{A}} \times \mathcal{U}$ non-linearly and represents the physical relation between the object and the observations on it. For any fixed pose of an object, the map R projects the entire set of observations related by monotonic amplitude transformations, to a unique point on the canonic manifold which represents the orbit of geometry-only deformations of the object. The map T then maps the result non-linearly so that the overall map $L \circ \psi_g : \tilde{\mathcal{A}} \times \mathcal{U} \to H$ is such that all observations from $\tilde{S}_g$ are mapped to the same distinct linear subspace H_g of H. We next provide a formal definition of the RIUME:

Definition 1: A radiometry invariant universal manifold embedding $L : \mathcal{O} \to H$ is a map from the space of functions (observations) into a low dimensional Euclidean space, H, such that the set $L(\tilde{S}_g)$ is a linear subspace of H for any g, and invariant w.r.t. any monotonic transformation of the amplitudes of g.

Figure 1 schematically illustrates the concept of the radiometry invariant universal manifold embedding. Our aim is to find a map L such that $L(\tilde{S}_g)$ is a linear subspace of some linear space $H \subset \mathbb{R}^{\mathrm{M}}$, universally for every $g \in S$, such that the map is independent of the deformation parameters.

The existence of the radiometry invariant universal embedding $L : \mathcal{O} \to H$ implies that many problems concerning the multiplicity of appearances of an object can be directly solved in H using classical *linear theory*, eliminating the need to employ nonlinear analysis. Thus for example, in order to characterize the mapped manifold of some object in the linear space H all that is required is a single sample from the set of appearances of the object so that the linear subspace in H can be evaluated.

III. RADIOMETRY INVARIANT UNIVERSAL MANIFOLD EMBEDDING FOR MULTI-DIMENSIONAL AFFINE TRANSFORMATIONS

In this section we extend the framework of the universal manifold embedding, [29], to the more general case where the observed object undergoes not only geometric transformations, but also radiometric transformations. More precisely, in [29] it is assumed that the variability in the object appearance is only due the geometric deformations the object undergoes, while the measured image intensities remain unchanged. In practice, this is rarely the case. Therefore in this section we generalize the previous results and consider the more realistic case where the observed object undergoes both an affine geometric transformation, and a monotonic radiometric transformation. We prove by construction the existence of a radiometry invariant UME.

More specifically let $\mathbb{R}^n$ be the n-dimensional Euclidean space and let $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^n$ be an affine transformation of coordinates, that is, $\mathbf{y} = \mathcal{A}(\mathbf{x})$ where $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, *i.e.*, $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$, $\mathbf{y} = [y_1, y_2, \ldots, y_n]^T$ such that $\mathbf{y} = \mathcal{A}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{c}$ and $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y} + \mathbf{b}$, where $\mathbf{A} \in GL_n(\mathbb{R})$, $\mathbf{b}$, $\mathbf{c} \in \mathbb{R}^n$. 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 $\mathbf{D} = \left[\mathbf{b}\ \mathbf{A}^{-1}\right]$ that represents the geometric deformation. Hence, in this case the set of possible geometric transformations $\mathcal{A}$ is parameterized by $\mathbf{D}$ (or equivalently by $\mathbf{A}$ and $\mathbf{c}$). Let $U : \mathbb{R} \to \mathbb{R}$ be an invertible function, representing the monotonic radiometric deformation.

The manifold of the function $g \in S$ is given by $\tilde{S}_g = \{U \circ g \circ \mathcal{A} | \mathcal{A} \in \tilde{\mathcal{A}}, U \in \mathcal{U}\}$ and our aim is to find a map L such that $L(\tilde{S}_g)$ is a linear subspace of some linear space $H \subset \mathbb{R}^{\mathrm{M}}$, universally for every $g \in S$, such that the map is independent of the parameters of U and $\mathcal{A}$.

A. Radiometry-Invariant, Geometry-Covariant Representation

Consider the following mapping $R : \mathcal{O} \to \mathcal{O}$ of an observation $g(\mathbf{x})$ to a new and "normalized" observation $\mathcal{G}(\mathbf{x})$, such that

$$\mathcal{G}(\mathbf{u}) = \frac{\lambda[\mathbf{x} : g(\mathbf{x}) \leq g(\mathbf{u})]}{\lambda[\mathrm{supp}\{\mathrm{g}(\mathbf{x})\}]} \tag{2}$$

where λ is the Lebesgue measure on $\mathbb{R}^n$, and supp{g} denotes the support of the function g, *i.e.*, the closure of the set where g does not vanish. Since $h = U \circ g \circ \mathcal{A}$ we have following [11] that applying R to both h and g, the remapped observations $\mathcal{G}(\mathbf{x})$ and $\mathcal{H}(\mathbf{x})$ are related by the affine relation

$$\mathcal{H}(\mathbf{x}) = [\mathcal{G} \circ \mathcal{A}](\mathbf{x}) = \mathcal{G}(\mathcal{A}(\mathbf{x})). \tag{3}$$

Thus, (3) represents an *affine, geometric only,* relation between the induced transformations of h and g by the operator R. In other words by applying the operator R, the dependence of the relation between h and g in the unknown radiometric deformation U, has been removed, while the resulting observations, $\mathcal{H}(\mathbf{x})$ and $\mathcal{G}(\mathbf{x})$ are two "new" observations related by an affine transformation $\mathcal{A}$. Thus, the operator R provides a radiometry-invariant representation of the manifold, and hence a radiometry-invariant-geometry-covariant relation between the products of applying it to any two objects $h, g \in \tilde{S}_g$. We next derive new representations for $\mathcal{H}(\mathbf{x})$ and $\mathcal{G}(\mathbf{x})$ that are invariant to the unknown affine transformation.

B. Radiometry Invariant Universal Manifold Embedding

Define an auxiliary function space, W, such that every function $w : \mathbb{R} \to \mathbb{R}$ in W is bounded, Lebesgue measurable, and vanishes at zero. Next, we define the mapping from the space of observations to itself induced by the functions in W. More specifically, we define a mapping such that a function $g(\mathbf{x})$ is mapped by $w \in W$ to some function $w(g(x))$. This operator is, in general, non linear.

Lemma 1 [10]: Let $\mathcal{G}, \mathcal{H} \in S_g$ be two observations on the same object such that $\mathcal{H} = \mathcal{G} \circ \mathcal{A}$ (i.e., U, the radiometric transformation, is the identity transformation). Let $M \in \mathbb{N}$ and let $w_\ell \in W$ $\ell = 1, \ldots, M$ be a set of bounded, Lebesgue measurable functions $w_\ell : \mathbb{R} \to \mathbb{R}$, where $w_\ell(0) = 0$. Let $\mathbf{D}_k$ denote the kth row of the matrix $\mathbf{D}$. Then, linear constraints (up to a scale factor) on the parametrization of $\mathcal{A}$ are found by applying functionals of the forms $\int_{\mathbb{R}^n} x_k w_\ell \circ \mathcal{H}(\mathbf{x}) d\mathbf{x}$ for some $w_\ell \in W$. These constraints take the form

$$\int_{\mathbb{R}^n} x_k w_\ell \circ \mathcal{H}(\mathbf{x}) d\mathbf{x} = \left|\mathbf{A}^{-1}\right| \int_{\mathbb{R}^n} (\mathbf{D}_k \tilde{\mathbf{y}}) w_\ell \circ \mathcal{G}(\tilde{\mathbf{y}}) d\tilde{\mathbf{y}}\ . \tag{4}$$

Inspecting equation (4) we conclude that application of each of the non linear functionals to the known relation $\mathcal{H}(\mathbf{x}) = \mathcal{G}(\mathcal{A}(\mathbf{x}))$ amounts to obtaining a linear constraint (up to a constant scale factor) on the affine transformation $\mathcal{A}$ between the centers of mass of the functions $w_\ell(\mathcal{H}(\mathbf{x}))$ and $w_\ell(\mathcal{G}(\mathbf{x}))$ obtained by an identical nonlinear operation on the amplitudes of $\mathcal{H}$ and $\mathcal{G}$. Thus, the proposed method employs stable "features" (centers of mass) rigorously extracted such that the correspondence between them is explicitly known.

Let f be some observation on a deformable object and let

$$\mathbf{T}^{f,1} = \begin{bmatrix} \int_{\mathbb{R}^n} w_1 \circ f(\mathbf{y}) & \int_{\mathbb{R}^n} y_1\ w_1 \circ f(\mathbf{y}) & \cdots & \int_{\mathbb{R}^n} y_n w_1 \circ f(\mathbf{y}) \\ \vdots & & \ddots & \vdots \\ \int_{\mathbb{R}^n} w_M \circ f(\mathbf{y}) & \int_{\mathbb{R}^n} y_1\ w_M \circ f(\mathbf{y}) & \cdots & \int_{\mathbb{R}^n} y_n w_M \circ f(\mathbf{y}) \end{bmatrix} \tag{5}$$

where in general, the notation $\mathbf{T}^{f,j}$ indicates that only moments of order less or equal to j, of $w_\ell \circ f$ are employed. Since in this subsection we consider only moments up to the first order, in the following we shall use the shorter notation $\mathbf{T}^f$ instead of $\mathbf{T}^{f,1}$, when we refer to $\mathbf{T}^{f,1}$.

We next provide a constructive proof of existence of the RIUME:

Theorem 1: Let $g \in \tilde{S}_g$ *and let* $\{1, x_1, \ldots, x_n\}$ *be the set of basis functions spanning the space of n-dimensional affine transformations. Let h be some other function in* $\tilde{S}_g$ *such that* $h = U \circ g \circ \mathcal{A}$*, and* $\mathcal{A} \in \tilde{\mathcal{A}}$*,* $U \in \mathcal{U}$*. Then, there exist a linear space* $H \subset \mathbb{R}^{\mathrm{M}}$ *and a pair of maps* $R : \mathcal{O} \to \mathcal{O}$ *and* $T : \mathcal{O} \to H$ *such that the restriction of the composite map to* $\tilde{S}_g$*, denoted by* $L^g : \tilde{\mathcal{A}} \times \mathcal{U} \to H$*, where* $L^g = L \circ \psi_g = T \circ R \circ \psi_g$*, is such that* $L^g = T \circ R \circ \psi_g$ *is a linear map (up to a fixed scale factor) from* $\tilde{\mathcal{A}}$ *to H, invariant of* $\mathcal{U}$*. Denote by* $T^{\mathcal{G}}$ *the restriction of T to* $R \circ \psi_g$*. Then,* $T^{\mathcal{G}}$ *is a linear map (up to a fixed scale factor) from* $\tilde{\mathcal{A}}$ *to H. As a linear operator on the parametrization of* $\tilde{\mathcal{A}}$*,* $T^{\mathcal{G}}$ *admits a matrix representation,* $\mathbf{T}^{\mathcal{G}}$ *of the form (5). The operator* L^g *is invertible up to a monotonic amplitude transformation, if and only if* $\mathbf{T}^{\mathcal{G}}$ *is of rank* $n + 1$*. The operator* L^g *is independent of the parameters of* $\mathcal{A}$ *and U.*

Proof: Since $h = U \circ g \circ \mathcal{A}$, we first apply the operator R defined in (2) to the observations h and g in order to obtain their radiometry-invariant-geometry-covariant representations $\mathcal{H}(\mathbf{x})$ and $\mathcal{G}(\mathbf{x})$, respectively. Following (3) we have that these representations are related by the affine relation $\mathcal{H}(\mathbf{x}) = \mathcal{G}(\mathcal{A}(\mathbf{x}))$.

Since the vectors in H are M-dimensional and since $L = T \circ R$, then in order to satisfy the required properties of L, we have that T has to be composed of M components $\{T_\ell\}_{\ell=1}^{M}$, where the ℓ-th component of T is such that the ℓ-th component of $T^{\mathcal{G}}$, $T_\ell^{\mathcal{G}} = T_\ell \circ (R \circ \psi_g)$ is a linear map from $\tilde{\mathcal{A}}$ to $\mathbb{R}$. In this case, L^g is also a linear map from $\tilde{\mathcal{A}}$ to $\mathbb{R}$. Since by the previous step we have that $\mathcal{H} = \mathcal{G} \circ \mathcal{A}$, we are looking for functionals $T_\ell(\mathcal{H})$ that are linear expressions (up to a scale factor) in the entries of $\mathbf{D}$ for any $\mathcal{G}$. We next construct T_ℓ and $T_\ell^{\mathcal{G}}$. Using the notation of Lemma 1, we have by fixing k and by taking $\ell = 1, \ldots, M$, that $T_\ell(\mathcal{H}) = \int_{\mathbb{R}^n} x_k w_\ell \circ \mathcal{H}(\mathbf{x}) d\mathbf{x}$. Hence, (4) can be rewritten in the following form

$$\mathbf{T}_\ell^{\mathcal{G}} \mathbf{D}_k^T = |\mathbf{A}|\, \mathbf{T}_{\ell,k}^{\mathcal{H}} \tag{6}$$

where $\mathbf{T}_\ell^{\mathcal{G}}$ is the ℓth row of $\mathbf{T}^{\mathcal{G}}$ and $\mathbf{T}_{\ell,k}^{\mathcal{H}}$ is the $(\ell, k + 1)$ element of $\mathbf{T}^{\mathcal{H}}$. Thus, we have that by choosing a family of linearly independent functions $\{w_\ell\}_{\ell=1}^{M} \in W$, we can construct an operator $T^{\mathcal{G}}$, identical for all k, that for every k imposes linear constraints on the elements of $\mathbf{D}_k$, which are the parameters defining $\tilde{\mathcal{A}}$. The operator T is constructed by stacking a sequence of M components $\{T_\ell\}_{\ell=1}^{M}$, and similarly $T^{\mathcal{G}}$ is constructed by stacking a sequence of M components $\{T_\ell^{\mathcal{G}}\}_{\ell=1}^{M}$. Since each operator T_ℓ is such that the composed map $T_\ell^{\mathcal{G}} = T_\ell \circ (R \circ \psi_g)$ is linear in our parametrization of $\tilde{\mathcal{A}}$ by the elements of $\mathbf{D}_k$, $T^{\mathcal{G}}$ is also linear in the elements of $\mathbf{D}_k$, for every $k = 1, \ldots, n$, and hence also L^g. Moreover, using (6) we have that since $T^{\mathcal{G}}$ is a linear operator from $\tilde{\mathcal{A}}$ to $\mathbb{R}^M$ it admits an $M \times (n + 1)$ matrix representation, given by $\mathbf{T}^{\mathcal{G}}$. Thus, $T^{\mathcal{G}}$ is invertible if and only if $\mathbf{T}^{\mathcal{G}}$ is of rank $n + 1$. Therefore, L^g is invertible up to a monotonic amplitude transformation, if and only if $\mathbf{T}^g$ is invertible.

Finally, L has the required properties to be a radiometry invariant universal manifold embedding, as it is independent of the specific parameters of $\mathcal{A}$ and U, while its structure is identical for every g. ∎

Remark 1: The proof of Theorem 1 shows by construction that there exists a linear space H which we call the reduced space*, and a pair of maps* $R : \mathcal{O} \to \mathcal{O}$ *and* $T : \mathcal{O} \to H$ *such that the overall map* $T \circ R \circ \psi_g : \tilde{\mathcal{A}} \times \mathcal{U} \to H$ *is invariant to the action of* $\mathcal{U}$ *and is linear in the parameters defining* $\tilde{\mathcal{A}}$*. This construction holds for every object* $g \in S$*, and any monotonic transformation of its amplitudes. The maps T and R are independent of the object. We call the map* $L = T \circ R$ radiometry invariant universal manifold embedding *(RIUME) as it universally maps each of the different manifolds, where each manifold corresponds to a single object, into a different linear subspace of H. The map* $\psi_g : \tilde{\mathcal{A}} \times \mathcal{U} \to O$ *maps* $\tilde{\mathcal{A}} \times \mathcal{U}$ *non-linearly and represents the physical relation between the object and the observations on it. For any fixed pose of an object, the map R projects the entire set of observations related by monotonic amplitude transformations, to a unique point on the canonic manifold which represents the orbit of geometry-only deformations of the object. The map T then maps the result non-linearly so that the composite map* $L \circ \psi_g : \tilde{\mathcal{A}} \times \mathcal{U} \to H$ *is such that all observations from* $\tilde{S}_g$ *are mapped to the same distinct linear subspace* H_g *of H.*

Denote $\tilde{\mathbf{D}} = [\mathbf{e}_1\ \mathbf{D}^T]$ where $\mathbf{e}_1 = [1, 0, \ldots, 0]^T$.

Corollary 1: Let $g \in \tilde{S}_g$ *and let h be some other function in* $\tilde{S}_g$ *such that* $h = U \circ g \circ \mathcal{A}$*, where* $\mathcal{A} \in \tilde{\mathcal{A}}$*,* $U \in \mathcal{U}$*. Then rewriting (6) in a matrix form we have*

$$\mathbf{T}^{\mathcal{G}} \left|\mathbf{A}^{-1}\right| \tilde{\mathbf{D}} = \mathbf{T}^{\mathcal{H}} \tag{7}$$

Since $\mathcal{A}$ *is invertible, so is its matrix representation* $\tilde{\mathbf{D}}$*, and hence the column space of* $\mathbf{T}^{\mathcal{G}}$ *and the column space of* $\mathbf{T}^{\mathcal{H}}$ *are identical subspaces of H. Hence all choices of the representative function of* $\tilde{S}_g$ *(g, or h, or any other function in the manifold) are equivalent.*

Remark 2: Corollary 1 implies that the functions in the manifold $\tilde{S}_g$ *form an equivalence class (with respect to producing the linear subspace by the radiometry invariant universal manifold embedding). Hence, any function from the manifold can be chosen as its representative. Any such arbitrary selection would yield the same linear subspace to which the entire manifold is mapped by the radiometry invariant universal manifold embedding.*

C. Radiometry Invariant Universal Manifold Embedding Using High Order Moments

We next extend the results of the previous section where zero- and first-order moments of the result of applying the non-linear operators $\{w_k\}_{k=1}^{M}$, to the radiometry normalized observations were considered, and employ higher order moments. As we show next, these high order moments yield *linear* constraints on corresponding higher order moments of the transformation parameters, and hence provide a more detailed linear representation of the linear subspace onto which the manifold is projected. We begin by providing a detailed analysis of the results when second-order moments are employed. The extension to higher orders is immediate, along the same lines.

Let $d_{i,j}$ denote the (i, j) element of $\mathbf{D}$. Thus, using the previously defined notations we have $\mathbf{x}_k = \sum_{i=0}^{n} d_{k,i} \tilde{\mathbf{y}}_i$. Following a procedure similar to the one in (4), linear constraints (up to a scale factor) on the moments of the parametrization of $\mathcal{A}$ are found by applying functionals of the forms $\int\limits_{\mathbb{R}^n} \mathbf{x}_k \mathbf{x}_p w_\ell(\mathcal{H}(\mathbf{x}))d\mathbf{x}$ for some $w_\ell \in W$. These constraints take the form

$$\begin{aligned}
&\int\limits_{\mathbb{R}^n} \mathbf{x}_k \mathbf{x}_p w_\ell(\mathcal{G}(\mathbf{x}))d\mathbf{x} \\
&\quad = \left|\mathbf{A}^{-1}\right| \int\limits_{\mathbb{R}^n} \sum_{i=0}^{n} d_{k,i}\tilde{\mathbf{y}}_i \sum_{j=0}^{n} d_{p,j}\tilde{\mathbf{y}}_j w_\ell(\mathcal{G}(\tilde{\mathbf{y}}))d\tilde{\mathbf{y}} \\
&\quad = \left|\mathbf{A}^{-1}\right| \sum_{i=0}^{n}\sum_{j=0}^{n} d_{k,i}d_{p,j} \int\limits_{\mathbb{R}^n} \tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_j w_\ell(\mathcal{G}(\tilde{\mathbf{y}}))d\tilde{\mathbf{y}} \ . \qquad (8)
\end{aligned}$$

Thus let

$$\mathbf{T}^{\mathcal{G},2} = \left[\mathbf{T}^{\mathcal{G},1} \left| \begin{matrix} \int\limits_{R^n} y_1 y_1 \ w_1(\mathcal{G}(\tilde{\mathbf{y}}))d\tilde{\mathbf{y}} & \cdots & \int\limits_{R^n} y_n y_n w_1(\mathcal{G}(\tilde{\mathbf{y}}))d\tilde{\mathbf{y}} \\ \vdots & & \vdots \\ \int\limits_{R^n} y_1 y_1 \ w_M(\mathcal{G}(\tilde{\mathbf{y}}))d\tilde{\mathbf{y}} & \cdots & \int\limits_{R^n} y_n y_n w_M(\mathcal{G}(\tilde{\mathbf{y}}))d\tilde{\mathbf{y}} \end{matrix} \right. \right]. \qquad (9)$$

Let us denote by Q^2 the set of the (inverse) affine transformation parameters and their distinct pairwise products, *i.e.*, $Q^2 = \{d_{1,1}, \ldots, d_{n,n+1}, d_{1,1}d_{1,1}, \ldots, d_{1,1}d_{1,n+1}, \ldots, d_{n,n+1}d_{n,n+1}\}$.

Theorem 2: Let $\{1, x_1, \ldots, x_n, x_1\ x_1, x_1\ x_2, \ldots x_n x_n\}$ be a set of basis functions. Let $g \in \tilde{S}_g$ and let h be some other function in $\tilde{S}_g$ such that $h = U \circ g \circ \mathcal{A}$, where $\mathcal{A} \in \tilde{\mathcal{A}}$, $U \in \mathcal{U}$. Let M be some positive integer. Then, there exists a linear space $H \subset \mathbb{R}^{\mathrm{M}}$ and a pair of maps $R : \mathcal{O} \to \mathcal{O}$ and $T^2 : O \to H$, such that the restriction of the overall map $L^{g,2} = T^2 \circ R \circ \psi_g : \tilde{\mathcal{A}} \times \mathcal{U} \to H$ to $\tilde{S}_g$, is such that the composed map $L^{g,2}$ is a linear map (up to a fixed scale factor) from Q^2 to H. As a linear operator on the parametrization Q^2, $T^{\mathcal{G},2}$ admits a matrix representation, $\mathbf{T}^{\mathcal{G},2}$ of the form (9).

The proof follows the same lines as the proof of Theorem 1, and hence is omitted.

Corollary 2: Let $g, h \in \tilde{S}_g$ and $\mathcal{A} \in \tilde{\mathcal{A}}$. Then rewriting (8) in a matrix form we have

$$\mathbf{T}^{\mathcal{G},2}\left|\mathbf{A}^{-1}\right|\tilde{\mathbf{D}}^2 = \mathbf{T}^{\mathcal{H},2} \qquad (10)$$

where the elements of $\tilde{\mathbf{D}}^2$ are obtained by rearranging the elements of Q^2. Assuming $\tilde{\mathbf{D}}^2$ is invertible, the column space of $\mathbf{T}^{\mathcal{G},2}$ and the column space of $\mathbf{T}^{\mathcal{H},2}$ are identical subspaces of H.

Note that the procedure yielding (8) and Theorem 2 can be extended to employ higher order moments by repeating the same methodology. This implies that by increasing the order of the moments from order one in (4), to order two in (8), and to higher orders, a more detailed characterization of the object manifold is obtained by projecting it onto a linear subspace spanned by the columns of $\mathbf{T}^{\mathcal{G},1}$, $\mathbf{T}^{\mathcal{G},2}$, to $\mathbf{T}^{\mathcal{G},K}$ for K-th order moments.

IV. THE DISTANCE BETWEEN MANIFOLDS

Measuring the distance between two subspaces of a larger subspace is a well explored problem, [23, p. 584], [24]–[26]. More formally, this problem is that of measuring the distance on the Grassmann manifold of subspaces of dimension $n+1$ in an ambient space of dimension M. 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 $range\,(\mathbf{A})$.

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

$$\begin{aligned}
\cos(\theta_k) &= \max_{\substack{x\in\mathbb{R}^p \\ y\in\mathbb{R}^q}} \frac{\left|x^T\mathbf{A}^T\mathbf{B}y\right|}{\|\mathbf{A}x\|_2\,\|\mathbf{B}y\|_2} \\
&= \frac{\left|x_k^T\mathbf{A}^T\mathbf{B}y_k\right|}{\|\mathbf{A}x_k\|_2\,\|\mathbf{B}y_k\|_2} \qquad (11)
\end{aligned}$$

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 $[0, \frac{\pi}{2}]$.

An alternative method for measuring the distance between subspaces is using the property that projection matrices have one-to-one correspondence to subspaces. That is, given two matrices $\mathbf{A}, \mathbf{B}$ whose column spaces are the same, the orthogonal projection theorem implies 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 Frobenius norm of the difference between the projection matrices onto the different object subspaces: Let $\mathbf{P}_A$ denote the projection matrix onto the column space of $\mathbf{A}$. $\mathbf{P}_B$ is similarly defined. Then, the relation between the Frobenius norm of the difference between the projection matrices onto the two subspaces, and the principal angles between the two subspaces, is given by

$$2^{-1/2}||\mathbf{P}_A - \mathbf{P}_B||_F = ||\sin\boldsymbol{\theta}||_2 \qquad (12)$$

Let $\mathbf{P}_f$ denote the projection matrix onto the $n+1$ dimensional column space of $\mathbf{T}^f$ defined in (5). Assume g and h are two observations on the same object, such that they are related by both an affine geometric transformation, and a monotonic radiometric transformation of their amplitudes, *i.e.*, $h(\mathbf{x}) = U(g(\mathcal{A}(\mathbf{x})))$. Using (3) we have that $\mathcal{H}(\mathbf{x}) = \mathcal{G}(\mathcal{A}(\mathbf{x}))$. Applying (7) we have proved the following theorem:

Theorem 3: Let $g \in \tilde{S}_g$ and let h be some other function in $\tilde{S}_g$ such that $h = U \circ g \circ \mathcal{A}$, where $\mathcal{A} \in \tilde{\mathcal{A}}$, $U \in \mathcal{U}$. Then $\mathbf{P}_{\mathcal{H}} = \mathbf{P}_{\mathcal{G}}$.

Theorem 3 provides the basis for matched manifold detection in the presence of both radiometry and geometry transformations between observations. It is concluded that as long as two observations on the same object differ by an affine transformation of coordinates and some monotonic transformation of the pixel amplitudes, the corresponding projection matrices $\mathbf{P}_{\mathcal{H}}$ and $\mathbf{P}_{\mathcal{G}}$ will be identical, while projection matrices that result from observations on other objects will be different and hence will yield non-zero distances form $\mathbf{P}_{\mathcal{G}}$. The entire

Algorithm 1 Generation of RIUME Invariant Representation

- Given an image h, substitute the intensity of each non-zero pixel with the accumulated relative frequency of this intensity level, to generate a radiometry-invariant geometry covariant realization $\mathcal{H}$ as detailed in (2).
- Compute the matrix $\mathbf{T}^{\mathcal{H}}$ as defined in (6).
- Compute $\mathbf{P}_{\mathcal{H}} = \mathbf{T}^{\mathcal{H}}((\mathbf{T}^{\mathcal{H}})^T\mathbf{T}^{\mathcal{H}})^{-1}(\mathbf{T}^{\mathcal{H}})^T$, the projection matrix onto the column space of $\mathbf{T}^{\mathcal{H}}$.
- $\mathbf{P}_{\mathcal{H}}$ is the RIUME representation of the image h.

procedure for generating a RIUME representation of a function h is summarized in **Algorithm 1**.

V. PERFORMANCE EVALUATION EXAMPLES

In this section, we detail several synthetic experiments aimed at investigating the basic properties and parameters of choice in the UME and RIUME framework. The synthetic set-up and the knowledge of the experimental parameters enable a better understanding of the interplay between the different parameters in generating the UME/RIUME descriptor of the manifold. In the next section (Section VI) we demonstrate, on real-world scenarios, the performance of detection and tracking algorithms which are based on the properties of the RIUME. In all the presented experiments image intensities are normalized to the interval [0, 1].

A. Identification of Object Manifolds

In this example, we demonstrate that the UME can indeed provide a robust representation of the object manifold. It is shown that the UME maps different manifolds to corresponding different subspaces, and that these subspaces are distinguishable using standard methods of subspace classification, such as principal angles. In this example 250 object templates were used, and different observations on them were synthesized using random affine deformations. Sample images from the data set are depicted in Fig. 2. Next, the UME was applied in order to map each observation to an affine-invariant subspace. The chosen family of nonlinear operators $\{w_i\}_{i=1}^{M}$ is of the form

$$w_i(x) = \begin{cases} 1 & \frac{i-1}{M} < x \leq \frac{i}{M} \\ 0 & \text{else} \end{cases}. \tag{13}$$

For each observation, we computed the principal angles between its UME representation and the representations of both its corresponding template and non-matching templates. Figure 3 shows the distributions of the angles between an object's observations and its *matching* and *non-matching* templates. The principal angles between UME representations provide an equivalent metric to that of the Frobenius norm of the difference between the corresponding projection matrices, as given in (12).

One can readily observe that these distributions are very different and hence easily separable. It is therefore concluded that the principal angles between the UME representations of different object manifolds form a highly informative statistic for discriminating between the manifolds of different objects. This example shows that the UME indeed provides an efficient affine-invariant representation of the possible observations in the manifold. This result is achieved although the *uniform and universal* choice of the nonlinear operators according to (12) made here ($M = 10$ was used), implies that no data driven training procedure is applied in order to optimize the choice of the nonlinear operators to the empirical statistics of some training set.

Fig. 2. Sample images from the data set of 250 images used for the examples in this section.

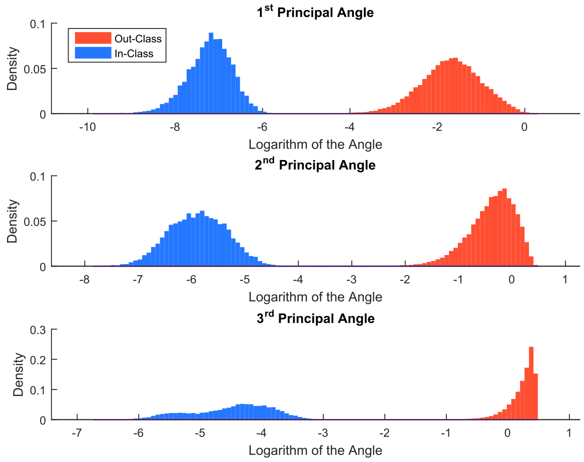

Fig. 3. Distributions of the principal angles between the UME representations of an object template and observations, where the observations are on different deformations of the *same* object (blue), in comparison with the case where the observations are on *different* objects (red).

It should be further emphasized that the in-class distributions contain distances from observations on multiple *different* objects and their respective templates. In fact, when building a classifier which tells apart only a *single* specific object from the others, the resulting distributions are even further apart. Thus, in optimal conditions, the UME is shown to be a reliable method for evaluating the distances between the manifolds of different objects.

B. The Choice of Nonlinear Operators $\{w_i\}_{i=1}^{M}$ and Its Effects on the Detector Performance

In this experiment we concentrate on the family of nonlinear operators $\{w_i\}_{i=1}^{M}$ defined in (13), where we compare the UME based classification for $M = 10$ and $M = 255$ in the presence of additive observation noise. More specifically, for each choice of M we estimate the distributions of the distances of the UME representations of differently deformed observations taken from the same object manifold in comparison with

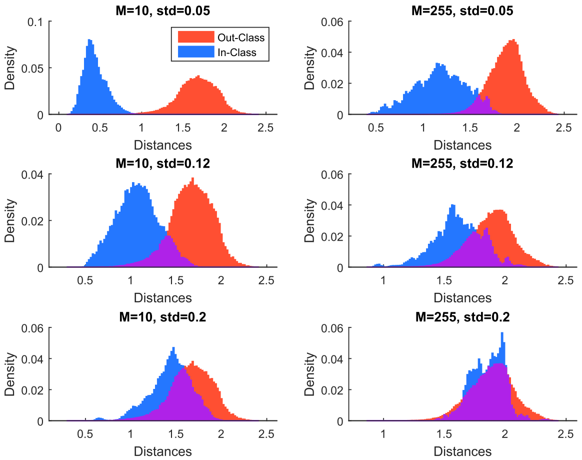

Fig. 4. Comparison of the distributions of the distances of the UME representations of differently deformed observations taken from the same object manifold (blue), in comparison with the distances of the UME representations of observations taken from different object manifolds (red). The upper two graphs show the distributions for the high SNR case, where relatively little noise is present, the middle and lower pairs show the middle and low SNR cases, respectively.

the distances of the UME representations of observations taken from different object manifolds. Figure 4 depicts the distance distributions for the high-, mid-, and low-SNR scenarios, respectively. To generate the noisy observations required for this example, pixel intensity values were normalized to the interval [0, 1] and Gaussian noise with a standard deviations of 0.05, 0.12 and 0.2 was added to generate the high-, mid-, and low-SNR cases, respectively. The distance between the manifolds of two objects is evaluated as the Frobenius norm of the difference between their corresponding UME projection matrices, as given in (12). The results indicate that the performance when $M = 10$ shows higher robustness in the presence of noise than in the case where $M = 255$. In addition, using a smaller number of operators results in a computationally more efficient procedure. We note that the *uniform and universal* choice of the nonlinear operators made in this paper implies that no data driven training procedure is applied in order to optimize the choice of the nonlinear operators to the empirical statistics of a training set. It is therefore guaranteed that the performance of the classification scheme is not biased towards any specific data set. On the other hand, this universal, data-independent choice of the nonlinear operators is clearly suboptimal in the case where the statics of the data can be learned from a training set.

C. Sampling Resolution

In this example we explore how limitations encountered in real-world scenarios affect the performance of the UME. Problems such as low resolution and incorrect segmentation may be unavoidable in many applications, and it is desirable to understand how robust the UME is to these issues. The next experiment demonstrates how the performance of the UME in identifying different manifolds behaves as a function of the sampling resolution of the observations. Note that in the previous sections images are described as functions from $\mathbb{R}^2$ to $\mathbb{R}$. In practice however, the intensity samples are available only on some finite grid. Approximating the integrals given in the description of the UME by summations causes an error in the computation of the matrix, which in turn causes a perturbation of the subspace. Yet, sampling is unavoidable, so we must analyze the performance of the UME with different resolutions to determine its operational range and limitations. In this experiment a data set of 250 objects was chosen and new observations on these objects were generated using random affine geometric deformations and at different resolutions of the observations. The aim here is to detect and identify a given object that has undergone an unknown affine deformation. The distance between the manifolds of two objects is evaluated as the Frobenius norm of the difference between their corresponding UME projection matrices, as given in (12). Figure 5 depicts the results of this experiment: an ROC graph is generated for each resolution, illustrating the detector's performance. For every point in the ROC the P_D coordinate is evaluated as the integral from $-\infty$ to the chosen threshold level of the in-class distribution, while the P_{FA} coordinate is evaluated as the integral from $-\infty$ to the same threshold level of the out-of-class distribution. It is evident that while at high sampling rates we observe the expected near perfect ROC, as the resolution deteriorates so does the performance.

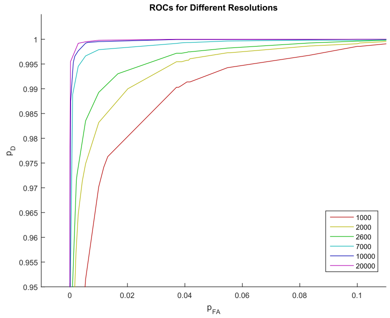

Fig. 5. ROC graphs based on UME representations constructed from images of different resolutions. The numbers in the legend provide the number of samples (pixels) per object.

D. Performance as a Function of the Number of Moments Used

In this set of examples, we wish to test how the number of moments taken when calculating the UME representation affects performance in the presence of noise. Theoretically, using more moments means extracting a more detailed representation of the manifold. Thus, from an information-theoretic perspective, the more moments taken, the easier it becomes to distinguish between manifolds. On the other hand, employing higher order moments in order to characterize each object manifold, results in a higher dimension UME subspace for each object, which leaves less room in the ambient space to separate different object classes. The experiments are performed in the presence of additive noise, the affine transformation parameters are selected at random, and the order of moments used to calculate the RIUME representation in (9) is modified from zero to four. The RIUME representation of the object in then employed to evaluate the corresponding projection matrix $\mathbf{P}_{\mathcal{G}}$.

The results of evaluating the distances between the projection matrices are depicted in Figure 6. From these results,

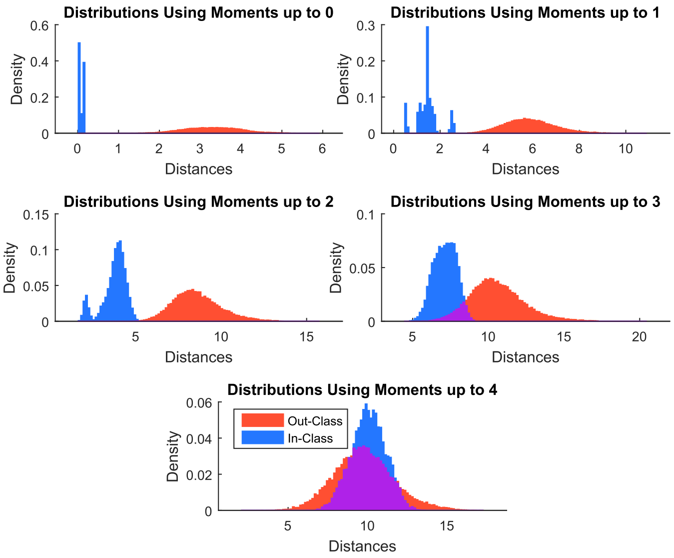

Fig. 6. In-class (blue) and out-of-class (red) distance distribution using moments of different orders.

it is concluded that in the presence of noise the best representation of the object is achieved by the subspace created by the vector of the histogram, since the zero order moment using the choice of nonlinear operators defined in (13) is effectively the histogram of the image. However, we note that this experiment was performed in a scenario where it is a-priori known that illumination is constant (except for the noise effects). Since application of the operator R in order to achieve radiometric invariance in the practical case where illumination changes are present, reduces the discriminative power of the image histograms, it is concluded that the best detection and recognition performance is achieved by employing the zero and first order moments, as defined by $\mathbf{T}^{\mathcal{G},1}$ and Theorem 1.

E. On the Difference Between the RIUME and the UME

We bring several examples that illustrate the advantage of using the RIUME over the UME. First, the point is illustrated in a synthetic setting. Using a setup similar to that used in previous synthetic experiments in this paper, observations were generated on 250 different objects. This time, however, each observation was created using both a random affine geometric deformation and a monotonic radiometric deformation. RIUME and UME representations were computed from the observations and the undeformed templates. Then, we calculated the distance of each observation representation, given by the projection matrix $\mathbf{P}_{\mathcal{H}}$, from its corresponding template representation, given by the projection matrix $\mathbf{P}_{\mathcal{G}}$, (the in-class) and the distance of each observation representation from unrelated templates (the out-class). Figure 7 depicts the distributions of the in- and out- classes using the RIUME and the UME representations. It can be seen that in the presence of illumination changes, applying the radiometric normalization operator that is part of the RIUME is crucial to the success of the identification task. Note that following the application of the operator R to each of the observations, the distributions of the intensities of the resulting images become more similar.

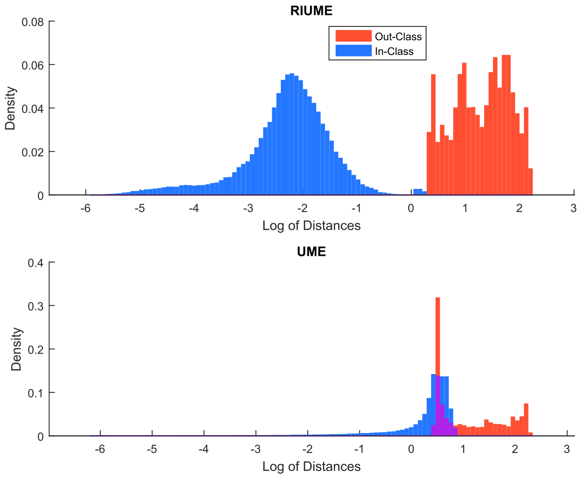

Fig. 7. In-class and out-of-class distributions - Performance comparison of the UME and RIUME.

Fig. 8. The two images used to demonstrate the performance of the UME *vs.* RIUME. The patches used in the example are marked with a dashed red line.

We now proceed to show the importance of using the RIUME in real-world scenarios. The next example involves two images taken by a camera (as opposed to the previously presented synthetically deformed examples). Figure 8 depicts two images of the same object taken from different angles and at different times such that the amplitude levels of identical features in the scene are significantly different. In this experiment we have extracted two corresponding patches out of the images (the triangular patches are marked with a dashed line in both images). Using the previously defined notations, we shall denote them g and h. Since locally, the geometric homography transformation between the two patches can be approximated by an affine transformation, a natural choice based on the geometric argument would be to evaluate the UME representations of both patches (*i.e.*, $\mathbf{T}^g$ and $\mathbf{T}^h$) and to verify that indeed the Frobenius norm of the difference between the projection matrices on the column spaces of the UME representations of both patches is close to zero. However, it turns out that the Frobenius distance between the two projection matrices is approximately 1.87, which is well within the range of distances from the UME representations of unrelated objects. On the other hand, the distance between the RIUME representations of the two triangular patches $\mathbf{T}^{\mathcal{G}}$ and $\mathbf{T}^{\mathcal{H}}$ yields a distance of 0.46, which is small enough to allow the identification of both patches as being observations on the same object in the world.

F. The RIUME in Comparison to SIFT as Descriptors

In the following we provide a comparison between the RIUME based projection matrix (denoted by $P_{\mathcal{H}}$ in the notation of **Algorithm 1**) as the descriptor of a patch h and the SIFT descriptor [31], chosen as a typical representative of the large family of local key-point descriptors (that includes methods such as SURF, GLOH, HOG). The goal of this comparison is to analyze using a comparative study the discriminative power of each of the descriptor types, when applied in similar conditions.

In this experiment, a data set of 250 objects was chosen and new observations on these objects were generated using random affine geometric deformations and monotonic radiometric deformations. Then, the SIFT feature detector was applied and the SIFT descriptor was attached to each detected key-point in both the template and the observation. Since the deformation is known, we were able to consider only pairs of features which were correctly matched and for these, we calculated the distances between their descriptors. Since SIFT features are extracted from a fixed size neighborhood at any given scale, we chose only pairs of features for which SIFT's scale selection was correct, in order to neutralize the deterioration in the SIFT performance related to the fact that SIFT uses fixed size neighborhoods at all scales. We compared these distances to distances between non-matching SIFT descriptors in order to produce the ROC graphs in Figure 10. To further illustrate the advantage of using the RIUME over SIFT, we controlled the amount of skew in the geometric deformations (See Fig 9 for typical examples). The results demonstrate that the SIFT performance deteriorates as the skew becomes more significant. The ROC curves of SIFT are compared to the corresponding ROC curve generated by the RIUME on 10 × 10 patches (in the original observation, before transformation) undergoing the same type of deformations. The RIUME performance is not affected by introducing skew, so only the results under maximal skew conditions are displayed. It is shown that in all the scenarios considered, the RIUME provides better performance than the SIFT. When patches are allowed to be bigger than 10 × 10, as is the typical case in the scenarios in which the RIUME should be employed, an even better classification on synthetic examples is achieved.

Finally, it is emphasize that the RIUME and SIFT are entirely different methods of extracting descriptors, which as explained in detail in the next section are aimed at different detail levels of characterizing the patch, and hence for different purposes. While the RIUME advantage is that it is designed to extract information from larger patches than SIFT does, and has better invariance to a larger class of deformations, it requires its input patches to be properly segmented. On the other hand, SIFT is designed to detect and characterize local key-point. As explained in the next section, the RIUME based matched manifold detection scheme provides a method for efficiently combining the advantages of the local, key point based methods, and the global registration methods.

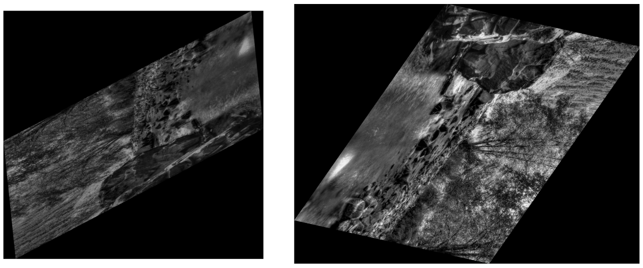

Fig. 9. Examples of the deformations used to produce comparison with SIFT. The left image represents the hard skew case, and the right image represents the medium skew case. The no skew case is simply a rotation and is not shown here.

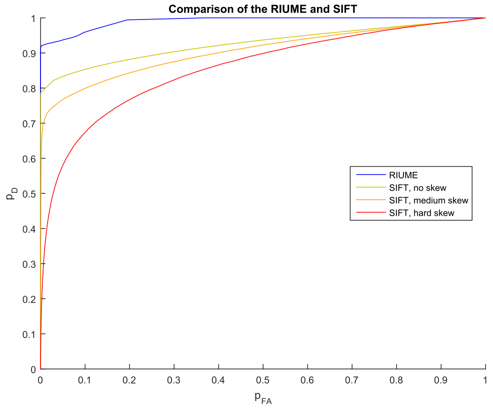

Fig. 10. Comparison with SIFT. ROC curves generated using the SIFT descriptor are displayed in yellow, orange and red. The corresponding ROC curve for RIUME is in blue.

VI. LOCAL MATCHED MANIFOLD DETECTION AND TRACKING

In general, the observed surface is not a single plane undergoing an affine transformation, and the radiometric variations across observations are not necessarily monotonic. Nevertheless, almost any surface can be well approximated by its tessellation into tiles, such that two observations on the same tile are related by simultaneous affine transformation of coordinates and a monotonic mapping of the intensities. In this scenario the modeling assumptions of the above derivation hold and hence the proposed detection scheme becomes highly practical as we explain in detail in the sequel.

A. Matched Manifold Detection

1) The Principles of Operation: Given two observations that contain an object to be detected, the first step in constructing the radiometry-geometry invariant matched manifold detection framework, is to apply some point matching algorithm in

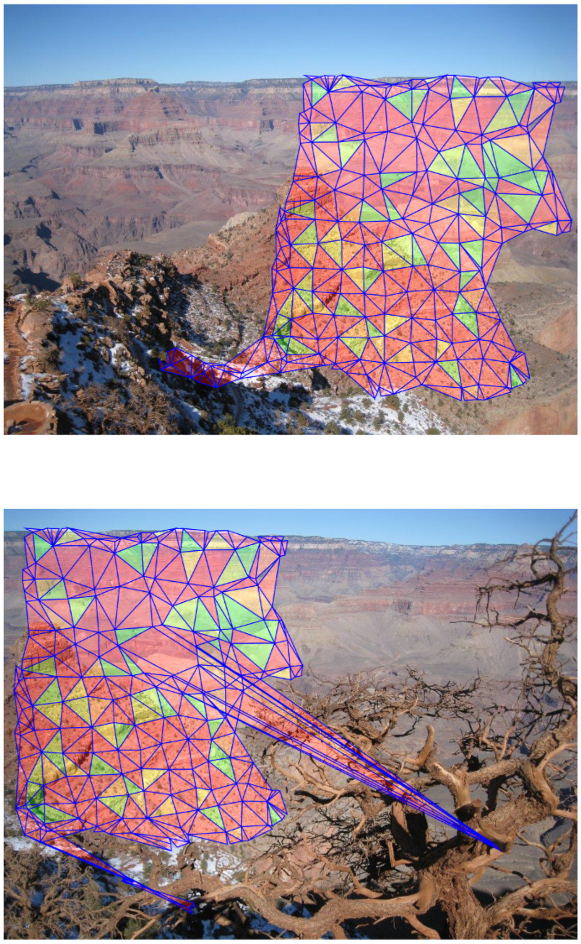

Fig. 11. Example of the operation of the matched manifold detector. A Delaunay triangulation is calculated on tentatively matched points in the upper image. This triangulation is projected onto the lower image. The irregular triangles seen in the lower image are due to both matching errors from the point-matching initializer, and severe geometric deformations. The distance between the RIUME representations of the tentatively matching triangular tiles in both images is color coded on the triangles themselves, from green indicating small distance to red indicating large distance.

order to find tentatively corresponding scene points in the two images. Given the two sets of tentatively corresponding points, the Delaunay triangulation is applied to tessellate one of the images into a set of disjoint tiles. Each of these tiles is assumed to be a planar surface, such that if the set of three points defining a triangular tile in one image indeed matches a set of three points on the other image, then the resulting triangular surfaces will be related by simultaneous affine transformation of coordinates and a monotonic mapping of the intensities.

Fig. 11 depicts an example illustrating a single stage of the process: The tentatively corresponding points from the point-matching algorithm are dilated in order to control the size of tessellation tiles, and a Delaunay triangulation is calculated on the reduced set of points to obtain the tessellation. Next, the distance between the RIUME representation of each tentatively corresponding pair of triangles is calculated. This procedure is repeated with various choices of reduced sets of tentatively corresponding points, in varying density, in order to increase the probability of choosing correctly matched tiles at the "right" scale. The optimal scale is one that yields triangular tiles, large enough to contain enough information and yet small enough so that the assumption that each of these tiles represents a planar surface, related to its corresponding tile by a simultaneous affine transformation of coordinates and a monotonic mapping of the intensities, holds. In Fig. 11, the distance between each pair of tentatively matching triangular tiles in both images is color coded on the triangles themselves, from green indicating small distance to red indicating large distance. Note that triangular tiles that result from false initial point matches, yield projection matrices that cannot be matched with projection matrices of tiles in the other image. Hence, they are excluded from the set of matching tiles (and hence are not green shaded). Following the completion of the described test, the affine deformations computed from each pair of matching triangles are used to provide a mapping and dense matching for all the pixels in the area covered by the matched triangles.

In general, the proposed algorithm can work on top of any state-of-the-art point matching algorithm. Since any given tile pair (source and target) in our tessellation is created by vertices that constitute point matches, it is successfully matched only if this entire set of matches is correct (*e.g.*, in case of triangulation, all 3 matches have to be correct).

Thus the RIUME is employed in order to provide a different type of information than existing point matching algorithms on the one hand, or global registration algorithms, on the other hand. Point matching algorithms aim at finding key points in the observed image and characterize them through the properties of small regions around them. These inherently local approaches use relatively small amounts of information (small patches) in generating the descriptor of a key point. As a consequence they result in non distinctive descriptors, which in turn lead to high rates of false matches that need to be eliminated before any further processing can take place. Such verification procedures require knowledge of the global geometric transformation model. A prominent example is the usage of the RANSAC algorithm [27] to eliminate faulty matches. However, such global geometric transformation model is often unknown, or difficult to estimate. On the other end, global registration algorithms may be applied only when the family of expected geometric deformations is a-priori known, and the radiometric deformations between the two observations are minimal. The RIUME based matched manifold detection scheme provides a method for efficiently combining the advantages of the local, key point based methods, and the global methods. This is achieved by employing the observation that almost any surface can be well approximated by its tessellation into tiles, such that two observations on the same tile are related by simultaneous affine transformation of coordinates and a monotonic mapping of the intensities. Thus, the combinatorial complexity of RANSAC-like methods is eliminated by using the RIUME for directly testing large patches defined by key points for their similarity. Since each such triangle is much larger than the small neighborhoods used to extract the key point descriptors, the matching procedure of

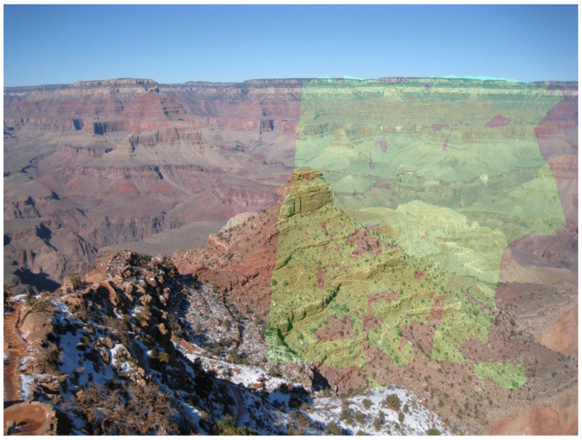

Fig. 12. Two images, taken from different view points but contain objects in common. The green shaded areas in both images were identified by the matched manifold detector as identical objects in both images.

triangular tiles is much more accurate than any procedure for matching local key point descriptors.

2) Matched Manifold Detection in Natural Scenes: The probability for a successful tile matching naturally increases with the precision of the point matches. Thus, the proposed matched manifold detector achieves significantly faster results when employed on top of more accurate point matching algorithm that produces lower rates of false matches. In this work, we used the expansion based matching approach introduced in [32], where matches are extracted with a low rate of false matches, and thus provide successful matches of many object tiles. Using the metric (12) for the distance between subspaces, false matches between triangular surfaces are efficiently rejected.

Figure 12 provides an example of the results obtained by applying the radiometry-geometry invariant matched manifold detector. The two images, although taken from different view points and at different times, contain objects in common. The green shaded areas in both images were identified as identical objects in both images. A decision that a pixel belongs to identical objects in both images is made only if it is identified to belong to matching triangles in both images, for at least two different tessellations obtained by the procedure described in the previous subsection.

Fig. 13. Sistine Chapel Image A.

Fig. 14. Sistine Chapel Image B.

3) The RIUME in Comparison with the UME in Matching Natural Scenes: We next elaborate on an example that demonstrates the difference in the effectiveness of employing the RIUME and the UME as detection building blocks of the matching-and-registration framework, built on top of the matched manifold detector as described in the previous subsection. Two images of the Sistine Chapel (Fig. 13 and Fig. 14), taken at different times, by different cameras and from different positions were used as inputs to the matching-and-registration process. The matching-and-registration process was performed twice, once with the RIUME based projection matrix (denoted by $P_{\mathcal{H}}$ in the notation of **Algorithm 1**) as the descriptor of the manifold of each triangle and once with the UME based projection matrix, P_h, as the descriptor of the manifold of each triangle. Figures 15 and 16 depict the pixels successfully mapped to their corresponding counterparts using each method. It can be observed that application of the RIUME based projection matrix as the descriptor of the manifold of each triangle yields much better performance, successfully mapping more pixels from one image to the other. This improved performance is highly important in tasks that are to follow that of matching-and-registration of the individual triangles of the tessellation result. These include tasks like detection, tracking (demonstrated below), structure from motion, and recognition of complex objects. As we have seen, applying the RIUME is a key factor in the ability to identify images undergoing illumination changes. In conclusion, applying the composed operator $T \circ R$ allows for successful object identification in the presence of significant variability in the illumination conditions.

Fig. 15. Registration with RIUME, mapped pixels are colored.

Fig. 16. Registration with UME, mapped pixels are colored.

B. Matched Manifold Tracking

Next, we demonstrate an application in tracking a complex object. Following a pre-segmented object of interest along a video sequence is a task of great importance in many applications, including video summarization, security-related products, and automated video photography using robots. The proposed matched manifold detector framework, well fits as a building block of a tracking scheme after complex objects, as it provides stable information on the pose of the object of interest in each frame of the video sequence. While the proposed matched manifold tracker can be incorporated in a more complete and robust tracking mechanism that includes for example a Kalman filter or a particle filter, we employ in this example the most basic version, that utilizes only the matched manifold detector with no additional feedback or correction mechanisms, to produce the presented results.

Fig. 17. Tracking application: Marking of the object of interest.

Fig. 18. Tracking application: Estimated object location after 2s.

Fig. 19. Tracking application: Estimated object location after 4s.

Fig. 20. Tracking application: Estimated object location after 6s.

The matched manifold tracker employs the tessellation based matched manifold detection-and-registration method proposed earlier in this section, to each consecutive pair of frames, thus advancing the prediction of the pose of the object of interest along the video sequence. More specifically, each pair of consecutive frames is tesselated into tiles, and the matching of the tiles is tested using the matched manifold detector. Once a pair of tiles has been verified to match, the affine transformation between the tiles is evaluated and hence the transformation of their interior points is known from each frame to the next one. Points in the object of interest are considered to be reliably tracked only if they were identified to belong to matching triangles in both images, for at least two different tessellations of the pair of images. In this way, each of the points of the object of interest is tracked from frame to frame. Figures 17-20 depict frames along the progress of the video sequence, with a temporal separation of 2 second between each frame (the video sequence itself is sampled at 10 frames per second). The green mask represents the successfully tracked pixels in the object of interest. The reader can observe that the object is tracked with high precision throughout the video despite the large changes in the viewing angle and illumination along the sequence. It should be noted that the algorithm uses no assumptions whatsoever on the camera, the deformations caused by the lens, or any side information on the object other than the initial marking in the first frame of the object to be tracked.

VII. CONCLUSIONS

We presented a novel approach for solving the problem of manifold learning, estimation, and detection for the case where the manifold is comprised of the set of all possible observations resulting from the simultaneous action of affine geometric deformations and monotonic radiometric transformations, on some object. Applying to each of the observations the radiometry invariant universal manifold embedding operator, the set of *all* possible observations on that object is mapped into a *single* linear subspace - invariant with respect to both the geometric and radiometric transformations. The derivation enables the representation of the object classification and detection problems in a linear subspace matching framework. Since the embedding of the space of observations in the linear space is independent of the specific observed object, it is universal. The derived RIUME provides an exact description of the manifold despite using as low as a single observation, and hence the need for using large amounts of observations in order to learn the manifold or a corresponding dictionary, is eliminated. Moreover, the proposed RIUME does not involve any discretization of the model, nor local approximations of the manifold, as the parametrization of the manifold remains in the continuum.

The invariant representation of the object by the RIUME is the basis of a matched manifold detection and tracking framework of objects that undergo complex geometric and radiometric deformations: The observed surface is tessellated into a set of tiles such that the deformation of each one is well approximated by an affine geometric transformation and a monotonic transformation of the measured intensities. Since each tile is mapped by the radiometry invariant UME to a distinct linear subspace, the detection and tracking problems are solved by evaluating distances between linear subspaces. Classification in this context becomes a problem of determining which labeled subspace in a Grassmannian is closest to a subspace in the same Grassmannian, where the latter has been generated by the RIUME from an unlabeled observation.

REFERENCES

[1] P. Dollár, V. Rabaud, and S. Belongie, "Learning to traverse image manifolds," in *Proc. NIPS*, Jan. 2006, p. 361.

[2] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[3] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[4] C. Walder and B. Schölkopf, "Diffeomorphic dimensionality reduction," in *Proc. NIPS*, 2009, pp. 1713–1720.

[5] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimension reduction via local tangent space alignment," *SIAM J. Sci. Comput.*, vol. 26, no. 1, pp. 313–338, 2002.

[6] K. Q. Weinberger and L. K. Saul, "Unsupervised learning of image manifolds by semidefinite programming," *Int. J. Comput. Vis.*, vol. 70, no. 1, pp. 77–90, Oct. 2006.

[7] "Dimensionality reduction via subspace and manifold learning," *IEEE Signal Process. Mag.*, Mar. 2011.

[8] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[9] R. Hagege and J. M. Francos, "Linear estimation of time-warped signals," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4423–4439, Jul. 2013.

[10] R. Hagege and J. M. Francos, "Parametric estimation of affine transformations: An exact linear solution," *J. Math. Imag. Vis.*, vol. 37, no. 1, pp. 1–16, Jan. 2010.

[11] S. Z. Kovalsky, R. Hagege, G. Cohen, and J. M. Francos, "Decoupled linear estimation of affine geometric deformations and nonlinear intensity transformations of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 940–946, May 2010.

[12] I. Tosic and P. Frossard, "Dictionary learning," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 27–38, Mar. 2011.

[13] Y. Ma, A. Yang, H. Derksen, and R. Fossum, "Estimation of subspace arrangements with applications in modeling and segmenting mixed data," *SIAM Rev.*, vol. 50, no. 3, pp. 413–458, 2008.

[14] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?" *Vis. Res.*, vol. 37, no. 23, pp. 3311–3325, 1997.

[15] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[16] E. Vural and P. Frossard, "Learning smooth pattern transformation manifolds," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1311–1325, Apr. 2013.

[17] L. Carin *et al.*, "Learning low-dimensional signal models," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 39–51, Mar. 2011.

[18] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," *Found. Comput. Math.*, vol. 9, no. 1, pp. 51–77, 2009.

[19] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[20] R. Vidal, "Subspace clustering," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 52–68, Mar. 2011.

[21] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, Aug. 1994.

[22] F. Mindru, T. Tuytelaars, L. Van Gool, and T. Moons, "Moment invariants for recognition under changing viewpoint and illumination," *Comput. Vis. Image Understand.*, vol. 94, nos. 1–3, pp. 3–27, 2004.

[23] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1989.

[24] P.-A. Absil, A. Edelman, and P. Koev, "On the largest principal angle between random subspaces," *Linear Algebra Appl.*, vol. 414, no. 1, pp. 288–294, 2006.

[25] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: Packings in Grassmannian spaces," *Experim. Math.*, vol. 5, no. 2, pp. 139–159, 1996.

[26] A. Barg and D. Yu Nogin, "Bounds on packings of spheres in the Grassmann manifold," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2450–2454, Sep. 2002.

[27] M. A. Fischler and R. C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[28] R. Sharon, J. M. Francos, and R. R. Hagege, “Detection and recognition of deformable objects using structured dimensionality reduction,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Brisbane, QLD, Australia, Apr. 2015, pp. 3442–3446.

[29] R. R. Hagege and J. M. Francos, “Universal manifold embedding for geometrically deformed functions,” *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3676–3684, Jun. 2016.

[30] R. Sharon, E. Farhan, R. R. Hagege, and J. M. Francos, “Geometry and radiometry invariant matched manifold detection and tracking,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Shanghai, China, Mar. 2016, pp. 3951–3955.

[31] D. G. Lowe, “Object recognition from local scale-invariant features,” in *Proc. IEEE 7th Int. Conf. Comput. Vis. (ICCV)*, Sep. 1999, pp. 1150–1157.

[32] E. Farhan and R. Hagege, “Geometric expansion for local feature analysis and matching,” *SIAM J. Imag. Sci.*, vol. 8, no. 4, pp. 2771–2813, 2015.

Ran Sharon received the B.Sc. and M.Sc. degrees in electrical and computer engineering from Ben-Gurion University, Beer Sheva, Israel, in 2014 and 2017, respectively. His research interests are image registration, structure from motion and other geometric topics in computer vision.

Joseph M. Francos received the B.Sc. degree in computer engineering and the D.Sc. degree in electrical engineering from the Technion-Israel Institute of Technology, Haifa, in 1982 and 1991, respectively.

From 1982 to 1987, he was with the Signal Corps Research Laboratories, Israeli Defense Forces. From 1991 to 1992, he was with the Department of Electrical Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA, as a Visiting Assistant Professor. Since 1993, he has been with Signal Processing Technology, Palo Alto, CA, USA. In 1993, he joined the Department of Electrical and Computer Engineering, Ben-Gurion University, Beer-Sheva, Israel, where he is currently a Professor. He heads the Mathematical Imaging Group, and the Signal Processing track. He also held visiting positions with the Massachusetts Institute of Technology Media Laboratory, Cambridge, with the Electrical and Computer Engineering Department, University of California at Davis, Davis, CA, USA, with the Electrical Engineering and Computer Science Department, INRIA Sophia-Antipolis, France, with the University of Illinois at Chicago, Chicago, IL, USA, and with the Electrical Engineering Department, University of California at Santa Cruz, Santa Cruz, CA, USA. His current research interests include parametric modeling and estimation of multi-dimensional signals, image registration, estimation of object deformations form images, manifold learning, parametric modeling and estimation of 2D random fields, and random fields theory.

Rami R. Hagege received the B.Sc. (*summa cum laude*) degree, the M.Sc. (*summa cum laude*) degree, and the Ph.D. degree from Ben-Gurion University, Beer Sheva, Israel, in 2002, 2004, and 2009, respectively, all in electrical and computer engineering. From 2009 to 2010, he was with the Massachusetts Institute of Technology Laboratory of Information and Decision Systems. From 2010 to 2016, he was with the Department of Electrical and Computer Engineering, Ben-Gurion University. He has been leading a Computer Vision Startup since 2016.