Matched Manifold Detection for Group-Invariant Registration and Classification of Images

Ziv Yavo, Yuval Haitman, Joseph M. Francos, and Louis L. Scharf, *Life Fellow, IEEE*

***Abstract*—Consider the set of possible observations turned out by geometric and radiometric transformations of an object. This set is generally a manifold in the ambient space of observations. It has been shown [1] that in those cases where the geometric deformations are affine and the radiometric deformations are monotonic, the radiometry invariant universal manifold embedding (RIUME) provides a mapping from the orbit of deformed observations to a single low dimensional linear subspace of Euclidean space. This linear subspace is invariant to the geometric and radiometric transformations and hence is a representative of the orbit. It thus naturally serves as an invariant statistic for solving problems of joint transformation estimation and detection or classification. In the unsupervised detection problem, subspaces evaluated from two observations are tested for the similarity of the observed object and their relative transformation is estimated from the RIUME matrix representation. In the classification set-up the RIUME subspace extracted from an experimental observation is tested against a set of subspaces representing the different object manifolds, in search for the nearest class. We show how to extract a set of mutually orthogonal subspaces, where each subspace represents a different object manifold. In the presence of observation noise, the observations do not lie strictly on the manifold and the resulting RIUME subspaces are noisy. We derive a method for estimating the mean subspace representation of a manifold of deformed observations. To optimize the performance of the matched manifold detector in the presence of observation noise, an analytic solution for choosing the RIUME nonlinear operators is derived, achieving the effect of simultaneous denoising of the object manifolds. The invariant representation of the object is the basis of a matched manifold detection and tracking framework for objects that undergo complex geometric and radiometric deformations. The experimental results on natural scenes demonstrate the generality and applicability of the RIUME framework for classification, detection, and dense registration.**

***Index Terms*—Object detection, invariant classification, affine coordinate transformation, matched manifold detection, subspace averaging, dense registration.**

Manuscript received July 9, 2020; revised March 12, 2021 and June 4, 2021; accepted June 25, 2021. Date of publication July 8, 2021; date of current version August 3, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Elias Aboutanios. This work was supported by NSF-BSF Computing and by Communication Foundations (CCF) under Grants CCF-2016667, CCF-1712788 and BSF-2016667 . *(Corresponding author: Joseph M. Francos.)*

Ziv Yavo, Yuval Haitman, and Joseph M. Francos are with the Department of Electrical and Computer Engineering, Ben-Gurion University, Beer-Sheva 84105, Israel (e-mail: zivyavo@gmail.com; yuvalhaitman@gmail.com; francos@ee.bgu.ac.il).

Louis L. Scharf is with the Department of Mathematics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: scharf@colostate.edu).

Digital Object Identifier 10.1109/TSP.2021.3095723

I. Introduction

THERE are many problems in image and signal analysis where an object to be detected presents itself subject to *a-priori* unknown geometric and radiometric transformations. Hence an understanding of the set of all possible observations of that single object is essential. We shall refer to these observations as *images*, where image is to be taken as a general term for scalar or vector valued measurements recorded at points in n-dimensional space.

As a result of the action of geometric and radiometric deformations, a set of observations (images) of an object is generally a manifold in the image space. Thus, although the data may be sampled and presented in a high-dimensional space because of the high resolution of the camera sensing the scene, in fact the intrinsic complexity and dimensionality of the observed physical phenomenon are low. While there are many cases where no prior knowledge of the sources of the variability in the appearances of an object is available, there are many scenarios in which such information is available, and hence can be exploited for efficient detection and classification of objects from their deformed images.

Radiometry invariant universal manifold embedding (RIUME) [1], [2] is a methodology for constructing a covariant matrix representation of an image, and then using this representation to identify a linear subspace that is invariant to monotonic amplitude transformations and to affine coordinate transformations of the image. The covariant RIUME matrix representation obtained by this procedure may be inverted for the parameters of the geometric transformation. Practical application of the method requires a high-quality estimate of the invariant subspace for each of K objects, and each of these subspaces must be estimated from one or more versions of an object, imperfectly imaged in one or more of its representative poses. This suggests that subspaces must be averaged in a training stage. Moreover, for reliable detection and classification, the subspaces for the respective images should be as well separated as possible. In this paper we address these issues by extending the theory of invariant target detection and classification, manifold denoising and wide baseline dense registration, in important ways:

1) By applying the order-fitting rule reported in [18] and [20], we extend the results in [19] to construct invariant subspaces for use in RIUME in the presence of observation noise.
2) We clarify the way in which *level-set images*, computed at each quantization level in an image, serve as a basis for the invariant subspaces in RIUME.

1053-587X © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Joseph Francos. Downloaded on May 11,2022 at 07:45:53 UTC from IEEE Xplore. Restrictions apply.

3) We derive an optimal companding of the level-set images for orthogonalizing as many as K RIUME subspaces for K objects; here $K \leq Q/(n+1)$, with Q the number of quantization levels in the image, and n determined by the number of degrees of freedom defining the group action (for example, $n = 2$ for the case of affine transformations of two-dimensional images).
4) In the presence of observation noise, the observations do not lie strictly on the manifold and the resulting RIUME subspaces are noisy. The derived analytic solutions for designing the RIUME operators and for estimating the mean RIUME representation of each manifold is equivalent to *simultaneous* denoising of all the object manifolds.
5) Since almost any imaged 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, we employ this framework to optimize a linear algorithm for estimating the homography transformation relating two observations taken from different angles on a planar surface. This approach is then extended and optimized for obtaining dense registration of complex scenes, where the shape of the object is *a-priori* unknown and no closed-form model of the transformation exists, such as in wide baseline multi-view registration.

The structure of this paper is as follows: In Section II we provide the basic definitions and properties of the radiometry invariant universal manifold embedding. Then, in Section III we define the basic principles of the Matched Manifold Detector (MMD). In Section IV we derive a method for estimating the mean subspace representation of a manifold of noisy and deformed observations, and determining its dimension. Section V elaborates on design procedures for optimizing the RIUME operator. This procedure is demonstrated using a detailed example, and a detailed experimental analysis of its performance in Section VI. An algorithm for robust homography estimation using the optimized local MMD is derived and its performance tested in Section VII. In Section VIII we demonstrate the effectiveness of the optimized MMD for wide baseline registration of a complex scene, where the shape of the scene is *a-priori* unknown and there is no closed-form model of the transformation. In Section IX we provide our conclusions.

II. Problem Formulation

Consider an object $s \in \{s_1, \ldots, s_K\}$, and an abstract *orbit* $\overline{\alpha}s, \overline{\alpha} \in \overline{G}$ of equivalent objects turned out by the transformation group $\overline{G}$. A typical group $\overline{G}$ is the rotation group of 3-D rigid objects, SO(3). In the framework of this paper we consider the case where the action of $\overline{G}$ on $\{s_1, \ldots, s_K\}$ can be approximated by (or inferred from) the action of another group G on a recorded, segmented, image of s, denoted $X(\mathbf{u}; s)$, where $\mathbf{u} \in \mathbb{R}^n$ is the image coordinate system, and $X : \mathbb{R}^n \longrightarrow \mathbb{R}$. More specifically, we concentrate on the special case where the action of G and its relation to $\overline{G}$ are approximated by

$$X(\mathbf{u}, \overline{\alpha} \circ s) = \alpha \circ X(\mathbf{u}, s) = U(X(\mathcal{A}(\mathbf{u}), s)), \qquad (1)$$

such that $U \in \mathcal{U} : \mathbb{R} \longrightarrow \mathbb{R}$ is a monotone radiometric map, and $\mathcal{A} \in \mathrm{Aff}[n] : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is an n-dimensional affine transformation of coordinates, parameterized by $\mathcal{A} : \mathbf{u} \mapsto \mathbf{v} = \mathbf{A}\mathbf{u} + \mathbf{c}$ where $\mathbf{A} \in \mathrm{GL}[n]$, $\mathbf{c} \in \mathbb{R}^n$; $\mathrm{Aff}[n]$ denotes the n-dimensional affine group and $\mathrm{GL}[n]$ the general linear group. The sets $\mathcal{U}$ and $\mathrm{Aff}[n]$ are closed under composition:

$$\begin{aligned} \alpha_2 \circ \alpha_1 \circ X(\mathbf{u}, s) &= U_2(U_1(X(\mathcal{A}_2(\mathcal{A}_1(\mathbf{u})), s))) \\ &= \alpha \circ X(\mathbf{u}, s), \alpha \in G. \end{aligned} \qquad (2)$$

The image $X(\mathbf{u}, s_k)$ of object s_k will be denoted X_k and the set $\psi_{X_k} = \{\alpha \circ X_k, \alpha \in G\}$ will denote the orbit of images turned out by the group G. There exists one such orbit for each object s_k. Our aim is to nonlinearly map each observation $\alpha \circ X_k$, taken from the orbit ψ_{X_k}, to a matrix representation $\mathbf{T}(X_k)$. This matrix is to be linearly covariant with the parametrization of G; Its column space, which we denote by $\langle \mathbf{T}(X_k) \rangle$ is to be G-invariant. In other words, the orbit ψ_{X_k} is mapped into a linear subspace $\langle \mathbf{T}(X_k) \rangle$, such that the mapping is G-invariant.

It is understood that the map from the group $\overline{G}$ to the group G will not precisely model the imaging of 3-D objects, for example. Similarly, radiometric variations are not globally monotone. In these cases, the mapping should be considered a *local* approximation of the mapping from object to image. Therefore, being able to infer the actions of G on an observation X provides local, or approximating, information about the hidden action of $\overline{G}$ on s.

In the context of this paper, the term "manifold" is adopted from the machine learning and dimensionality reduction literature, [4]–[12], to refer to the orbit of $X(\mathbf{u}, s)$ under G, *i.e.*, to the set of all possible observations $\{\alpha \circ X(\mathbf{u}, s), \alpha \in G\}$, due to the action of the group G. We note that most of these manifold learning techniques consider the case where the data lies on the manifold. However, only very few methods have been suggested to estimate the underlying manifold structure from noisy data, [14], [15].

It has been shown [1] that in the case where the observations on an object are determined by an affine geometric transformation of coordinates, jointly with a monotonic radiometric transformation, the RIUME operator returns a basis $\mathbf{T}(X)$ that is covariant to the coordinate transformation, and a subspace $\langle \mathbf{T}(X) \rangle$ that is G-invariant. That is, the set of *all* possible observations on an object under group action G is mapped by the RIUME operator into a *single* linear subspace which is invariant to both the geometric and radiometric transformations.

A. Radiometry Invariant Universal Manifold Embedding

Begin with image X and its deformation $\alpha \circ X$:

$$\alpha \circ X(\mathbf{u}) = U(X(\mathcal{A}(\mathbf{u}))), \quad \mathbf{u} \in \mathbb{R}^n \qquad (3)$$

where U is invertible and $\mathcal{A}$ is affine. In [1], [2] two maps are defined: $R : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ and $T : L^2(\mathbb{R}^n) \to \mathcal{T}(M, n+1)$, where $\mathcal{T}(M, n+1)$ is the space of $M \times (n+1)$ real-valued matrices, and M is the dimension of the embedding Euclidean space.

The map $T \circ R$ is called a *radiometry invariant universal manifold embedding*, RIUME. It maps every observation X from the orbit ψ_X to a matrix $\mathbf{T}(X) \in \mathcal{T}(M, n+1)$ such that $\mathbf{T}(X)$

Authorized licensed use limited to: Joseph Francos. Downloaded on May 11,2022 at 07:45:53 UTC from IEEE Xplore. Restrictions apply.

is covariant with the geometric transformation and invariant to the radiometric transformation. It allows for simple estimation of the affine transformation between any two observations on the same object.

The map $\mathcal{Q}: \mathcal{T}(M, n+1) \to \text{Gr}(M, n+1)$, where $\text{Gr}(M, n+1)$ is the Grassmann manifold of $n+1$-dimensional linear subspaces of M-dimensional Euclidean space, maps $\mathbf{T}(X)$ to its column space $\langle \mathbf{T}(X) \rangle$. We conclude that the RIUME maps the orbit ψ_X into the G-invariant subspace $\langle \mathbf{T}(X) \rangle \in \text{Gr}(M, n+1)$. That is, $\mathcal{Q} \circ (T \circ R) : \psi_X \longrightarrow \langle \mathbf{T}(X) \rangle$.

The details are these. Consider the mapping R of $X(\mathbf{u}), \mathbf{u} \in \mathbb{R}^n$ to a new and "normalized" observation $\tilde{X}(\mathbf{u})$ (for breivity we omit the dependence on s from the notation), where

$$\tilde{X}(\mathbf{u}) = R(X(\mathbf{u})) = \frac{\lambda[\mathbf{x} : X(\mathbf{x}) \leq X(\mathbf{u})]}{\lambda[\text{supp}\{\text{X}\}]} \tag{4}$$

and λ is Lebesgue measure. It is shown in [3] that $R(U(X(\mathcal{A}(\mathbf{u})))) = R(X(\mathcal{A}(\mathbf{u})))$. That is, the histogram equalization R applied to $U(X(\mathcal{A}(\mathbf{u})))$ returns a histogram-equalized image $R(X(\mathcal{A}(\mathbf{u})))$. The effects of the group action U have been removed, leaving only the group action $\mathcal{A}$. To simplify notation we remove the tilde with the understanding that the radiometric equalization has been applied.

To characterize the affine transformation of coordinates, begin with $\mathbf{v} = [v_1, \ldots, v_n]^T$ and let $\tilde{\mathbf{v}} = [1, v_1, \ldots, v_n]^T$ denote the homogeneous coordinates representation of $\mathbf{v}$. Thus, $\mathbf{u} = \mathbf{D}\tilde{\mathbf{v}}$ where $\mathbf{D}$ is an $n \times (n+1)$ matrix given by $\mathbf{D} = [\mathbf{b}\ \mathbf{A}^{-1}]$, where $\mathbf{b} = -\mathbf{A}^{-1}\mathbf{c}$. Let $w_l \ \ l = 1, \ldots, M$ be a set of bounded, Lebesgue measurable functions $w_l : \mathbb{R} \to \mathbb{R}$. Let $\mathbf{D}_k$ denote the kth row of the matrix $\mathbf{D}$. Then, [13],

$$\int_{\mathbb{R}^n} u_k w_\ell \circ (\alpha \circ X(\mathbf{u})) d\mathbf{u} = \left|\mathbf{A}^{-1}\right| \int_{\mathbb{R}^n} (\mathbf{D}_k \tilde{\mathbf{v}}) w_\ell \circ X(\mathbf{v}) d\mathbf{v}. \tag{5}$$

Define the $M \times (n+1)$ matrix, Eqn. (6) shown at the bottom of this page,

We call $\mathbf{T}(X)$ the RIUME matrix representation of the image X. It amounts to a mapping of X to an $M \times (n+1)$ matrix of first-order moments in each of n coordinate directions, plus one zeroth-order moment. Each of these moments is computed for one of M *companded* versions of the image, denoted $w_m \circ X$. The typical moment in the $(m, i+1)$ element of $\mathbf{T}$ is an integral of $u_i w_m \circ X$.

Denote $\tilde{\mathbf{D}} = [\mathbf{e}_1\ \mathbf{D}^T]$ where $\mathbf{e}_1 = [1, 0, \ldots, 0]^T$. Then, if $\alpha \circ X$ is an observation of X undergoing an affine deformation represented by the matrix $\mathbf{D}$, then from (5) we have

$$\mathbf{T}(\alpha \circ X) = \mathbf{T}(X) \left|\mathbf{A}^{-1}\right| \tilde{\mathbf{D}}. \tag{7}$$

Since $\mathbf{T}(\alpha \circ X)$ and $\mathbf{T}(X)$ are related by an invertible transformation that is a re-expression of the affine transformation matrix $\mathbf{D}$ relating the observations, we say that the basis $\mathbf{T}(X)|\mathbf{A}^{-1}|\tilde{\mathbf{D}}$ is *covariant* with the affine transformation. Hence it provides a method for estimating the affine transformation that relates any two observations. The affine transformation that relates X and $\alpha \circ X$ is evaluated by solving this linear system, [13]. That is, noise-free, $(\mathbf{T}(X)^T\mathbf{T}(X))^{-1}\mathbf{T}(X)^T\mathbf{T}(\alpha \circ X) = |\mathbf{A}^{-1}|\tilde{\mathbf{D}}$. The scale $|\mathbf{A}^{-1}|$ is determined from the ratio of any of the elements in the left column $\mathbf{T}(\alpha \circ X)$ to the corresponding element in $\mathbf{T}(X)$, and then the elements of $\mathbf{D}$ may be determined. Furthermore, since $\mathbf{T}(\alpha \circ X)$ and $\mathbf{T}(X)$ are related by a right invertible linear transformation, the column space of $\mathbf{T}(X)$ and the column space of $\mathbf{T}(\alpha \circ X)$ are identical. Their bases are different, but their range spaces are identical. Hence, the subspace $\langle \mathbf{T}(X) \rangle$ is a G-invariant statistic that is constant on image orbits, and hence distinguishes the orbits of different objects under affine coordinate transformation of their images.

Computation of covariant (sometimes called equivariant) moments, and the use of these moments to construct invariant functions of these moments, usually as rational functions of the computed moments, are often employed for classification in computer vision, see, *e.g.*, [25],–[27] and the references therein. Typically several invariant functions of many covariant moments are computed, and these are used to represent an image and its versions under a transformation group. Most of this research is devoted to finding invariants to the action of geometric transformations, and especially by considering their contours, as these are assumed to be less sensitive to illumination variations. In [27] this framework is generalized to include color moments in order to provide invariants in the presence of linear intensity variations. However, it is known that the use of high order moments is problematic especially in the presence of illumination model mismatches and noise. Here we are instead using only a zeroth-order moment and n first-order moments, but computing these for many companded versions of the image, to achieve invariance to affine geometric transformations and monotonic illumination variations. We are exchanging a large set of moments of a single image for a small number of moments of a large number of companded versions of the image. In some cases these compandings are level-set slices of the image. In this case a small number of moments are computed for what might be called the Lebesgue supports of the image at level slices. These moments are organized into an equivariant matrix which may be used to define an invariant subspace. Thus the major difference between the approaches is that unlike the moment invariant methods that use high-order moments and their nonlinear invariant functions for classification, the RIUME representation uses low-order moments of many compandings of

$$\mathbf{T}(X) = \begin{bmatrix} \int_{\mathbb{R}^n} w_1 \circ X(\mathbf{u}) d\mathbf{u} & \int_{\mathbb{R}^n} u_1 w_1 \circ X(\mathbf{u}) d\mathbf{u} & \ldots & \int_{\mathbb{R}^n} u_n w_1 \circ X(\mathbf{u}) d\mathbf{u} \\ \vdots & \ddots & & \vdots \\ \int_{\mathbb{R}^n} w_M \circ X(\mathbf{u}) d\mathbf{u} & \int_{\mathbb{R}^n} u_1 w_M \circ X(\mathbf{u}) d\mathbf{u} & \ldots & \int_{\mathbb{R}^n} u_n w_M \circ X(\mathbf{u}) d\mathbf{u} \end{bmatrix} \tag{6}$$

Authorized licensed use limited to: Joseph Francos. Downloaded on May 11,2022 at 07:45:53 UTC from IEEE Xplore. Restrictions apply.

an image. These produce a subspace representation of an image orbit, a subspace that may be used for covariant estimation and invariant detection-and-classification.

III. The Detection-Classification Problem and the Distance Between Equivalence Classes

The RIUME uses the operator $\mathbf{T}$ to universally map a manifold, generated by the set all monotone radiometric and affine coordinate transformations of an imaged object, into a G-invariant linear subspace. That is, the RIUME operator maps the orbit $\{\alpha \circ X, \alpha \in G\}$, to a point $\langle \mathbf{T}(X) \rangle$ on the Grassmannian $\mathrm{Gr}(M, n+1)$.

In the RIUME framework the problem of detection-classification of radiometrically and geometrically deformed objects is formalized as follows: Given an observation Z, (for example in the form of an image), where both its geometric and radiometric deformations are unknown, the problem is to determine whether $Z = \alpha \circ X$ or $Z = \beta \circ Y$, for some $\alpha, \beta \in G$, and X, Y some reference images of known objects.

Since the detection-classification is to be G-invariant, we propose in this paper to compute $\mathbf{T}(Z)$ using (6) and measure the distance between the subspace $\langle \mathbf{T}(Z) \rangle$ and the subspaces $\langle \mathbf{T}(X) \rangle$ and $\langle \mathbf{T}(Y) \rangle$. That is, the observation Z is determined to belong to the orbit ψ_X if the distance from $\langle \mathbf{T}(Z) \rangle$ to $\langle \mathbf{T}(X) \rangle$ is smaller than its distance to $\langle \mathbf{T}(Y) \rangle$, and is small enough to be considered a detection. Then, the observation Z is determined to be $\alpha \circ X$ for some $\alpha \in G$.

Following [22], [23] we compute the distance between a pair of subspaces as the extrinsic distance, evaluated using the projection Frobenius-norm

$$d(\langle \mathbf{T}(Z) \rangle, \langle \mathbf{T}(X) \rangle) = 2^{-\frac{1}{2}} ||\mathbf{P}_X - \mathbf{P}_Z||_F = ||\sin \boldsymbol{\theta}||_2 \quad (8)$$

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

All of these arguments extend to the classification of the observation Z as an element of the orbit ψ_s, where $\mathrm{argmin}_k \, d(\langle \mathbf{T}(Z) \rangle), \langle \mathbf{T}(X_k) \rangle) = \langle \mathbf{T}(X_s) \rangle$.

In practice we do not have noise-free observations on the set of objects $\{s_1, \ldots, s_K\}$. Therefore, the subspace $\langle \mathbf{T}(X_k) \rangle$ must be computed in a training step from noisy versions of $\alpha \circ X_k$, $\alpha \in G$. This suggests that for every $k = 1, \ldots, K$, experimental copies of the subspaces $\langle \mathbf{T}(X_k) \rangle$ must be averaged in order to arrive at a subspace that approximates the "noise-free" $\langle \mathbf{T}(X_k) \rangle$. This average subspace is the G-invariant statistic of the denoised manifold, obtained without explicitly first obtaining the denoised manifold as in [14], [15].

IV. The Subspace Mean

To compute the extrinsic distance between subspaces is to compute the Frobenius norm between the respective orthogonal projection matrices onto these subspaces. This suggests that the extrinsic "average" of several subspaces should be derived from an average of their projections. But this suggestion requires refinement, for the average of projections is no longer a projection. In [17], the problem of averaging affine transformed n-point configurations represented as subspaces of $\mathbb{R}^n$, is addressed. In [16] an analysis of different subspace means is presented. In [18], [19] the result to follow is derived from a slightly different perspective, and the order fitting rule of [18] is used to determine the dimension of the average subspace.

Let $\{\mathbf{P}_\ell, \ell = 1, 2, \ldots, L\}$ denote a set of $M \times M$ orthogonal projection matrices onto r-dimensional subspaces of M-dimensional Euclidean space. Each projection matrix projects a standard basis vector $\mathbf{e}_m \in \mathbb{R}^M$ onto a subspace, as $\mathbf{P}_\ell \mathbf{e}_m$. We seek an average projection $\overline{\mathbf{P}}$ which minimizes the averaged squared distance between all such projections, where the average is over all basis vectors and all projections. That is,

$$\begin{aligned} \overline{\mathbf{P}} &= \arg \min_{\mathbf{P} \in \mathcal{P}_r} \frac{1}{L} \sum_{m=1}^{M} \sum_{\ell=1}^{L} \|(\mathbf{P} - \mathbf{P}_\ell)\mathbf{e}_m\|^2 \\ &= \arg \min_{\mathbf{P} \in \mathcal{P}_r} \frac{1}{L} \sum_{m=1}^{M} \sum_{\ell=1}^{L} \mathbf{e}_m^T (\mathbf{P} - \mathbf{P}_\ell)^T (\mathbf{P} - \mathbf{P}_\ell)\mathbf{e}_m \end{aligned} \quad (9)$$

where $\mathcal{P}_r$ is the set of all rank r projections in the M-dimensional ambient space. Let $\mathbf{Q} = \frac{1}{L} \sum_{\ell=1}^{L} \mathbf{P}_\ell$. Then

$$\begin{aligned} \overline{\mathbf{P}} &= \arg \min_{\mathbf{P} \in \mathcal{P}_r} \sum_{m=1}^{M} \mathbf{e}_m^T (\mathbf{P} - \mathbf{Q}\mathbf{P} - \mathbf{P}\mathbf{Q} + \mathbf{Q})\mathbf{e}_m \\ &= \arg \min_{\mathbf{P} \in \mathcal{P}_r} \mathrm{tr}(\mathbf{P} - \mathbf{Q}\mathbf{P} - \mathbf{P}\mathbf{Q} + \mathbf{Q}) \\ &= \arg \min_{\mathbf{P} \in \mathcal{P}_r} \mathrm{tr}\left[(\mathbf{P} - \mathbf{Q})(\mathbf{P} - \mathbf{Q}) + (\mathbf{Q} - \mathbf{Q}^2)\right] \end{aligned} \quad (10)$$

The term $(\mathbf{Q} - \mathbf{Q}^2)$ is not affected by the choice of $\mathbf{P}$, so the equivalent problem is

$$\overline{\mathbf{P}} = \arg \min_{\mathbf{P} \in \mathcal{P}_r} \mathrm{tr}\left[(\mathbf{P} - \mathbf{Q})(\mathbf{P} - \mathbf{Q})\right] \quad (11)$$

Represent the orthogonal projection $\mathbf{P}$ as $\mathbf{P} = \mathbf{U}_r \mathbf{U}_r^T$, where $\mathbf{U}_r$ is a slice of an orthogonal matrix. Since $\mathbf{Q}$ is the average of projection (and hence symmetric) matrices, it is symmetric. Hence, its EVD is given by $\mathbf{Q} = \mathbf{V}\boldsymbol{\Gamma}\mathbf{V}^T$, $\boldsymbol{\Gamma} = diag(\gamma_i)$, $\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_M$, with $\mathbf{V}$ orthogonal. Then, the quadratic form to be minimized is [18]–[20]

$$\begin{aligned} V_r &\doteq \mathrm{tr}\left[(\mathbf{U}_r \mathbf{U}_r^T - \mathbf{V}\boldsymbol{\Gamma}\mathbf{V}^T)(\mathbf{U}_r \mathbf{U}_r^T - \mathbf{V}\boldsymbol{\Gamma}\mathbf{V}^T)\right] \\ &= \mathrm{tr}\left[\mathbf{U}_r \mathbf{U}_r^T\right] + \mathrm{tr}\left[\mathbf{V}\boldsymbol{\Gamma}^2\mathbf{V}^T\right] - 2\mathrm{tr}[\mathbf{U}_r^T \mathbf{V}\boldsymbol{\Gamma}\mathbf{V}^T \mathbf{U}_r] \\ &= r + \sum_{m=1}^{M} \gamma_m^2 - 2\mathrm{tr}[\mathbf{U}_r^T \mathbf{V}\boldsymbol{\Gamma}\mathbf{V}^T \mathbf{U}_r] \\ &\geq r + \sum_{m=1}^{M} \gamma_m^2 - 2\sum_{m=1}^{r} \gamma_m \end{aligned} \quad (12)$$

Authorized licensed use limited to: Joseph Francos. Downloaded on May 11,2022 at 07:45:53 UTC from IEEE Xplore. Restrictions apply.

with equality, for a given r, iff $\mathbf{U}_r = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r]$, where $\mathbf{v}_i, i = 1, .., r$ are the r eigenvectors of $\mathbf{Q}$ associated with its r largest eigenvalues. Hence, $\langle \mathbf{U}_r \rangle$ is the desired *G-invariant mean subspace*. The resulting mean-squared error is

$$V_r = \sum_{m=1}^{r} (\gamma_m - 1)^2 + \sum_{m=r+1}^{M} \gamma_m^2 \tag{13}$$

However, since $V_r \leq V_{r-1}$ as long as $1/2 \leq \gamma_r$, we conclude that V_r is minimized over rank r by choosing $r^* = \max m : \gamma_m \geq 1/2$ [18]. This completes the computation of the best rank r^* projection $\overline{\mathbf{P}}_{r^*}$ for averaging the set of projections $\{\mathbf{P}_\ell\}, \ell = 1, 2, \ldots, L$, with V_{r^*} the measure of average error.

The experimental procedure is now this: for each object s_k, record L noisy versions of the images $\alpha \circ X_k$; for each of these images extract its G-invariant subspace; average these subspaces according to the procedure above to estimate a G-invariant subspace for the orbit ψ_k.

V. The Optimal Set of Representatives for Multiple Orbits

We now show that the set of companding w-functions may be designed to orthogonalize the subspaces $\langle \mathbf{T}(X_k) \rangle, k = 1, \ldots, K$.

A. Representation by Level-Sets

Assume we are given an observation $X(\mathbf{u})$, $\mathbf{u} \in \mathbb{R}^n$, quantized at levels $\{q_i\}_{i=1}^{Q}$, so that it may be written as

$$X(\mathbf{u}) = \sum_{i=1}^{Q} q_i I_i^X(\mathbf{u}) \tag{14}$$

where $I_i^X(\mathbf{u})$ is the indicator function that equals 1 on the level-set of $\mathbf{u}$ where $q_{i-1} < X(\mathbf{u}) \leq q_i$, and zero elsewhere.

The w operators must be designed such that the result of their application is covariant with the geometric transformation, and hence they are not functions of the coordinates. The action of w_m on the image X is simply to map the levels q_i into levels $w_m(q_i)$, leaving the indicator images I_i^X unchanged. Then, each term in the matrix $\mathbf{T}(X)$ may be written as

$$\begin{aligned} T_{m,j} &= \int_{\mathbb{R}^n} w_m \circ X(\mathbf{u}) u_j d\mathbf{u} \\ &= \sum_{i=1}^{Q} w_m(q_i) \int_{\mathbb{R}^n} I_i^X(\mathbf{u}) u_j d\mathbf{u} \\ &= \sum_{i=1}^{Q} w_{m,i} F_{ij}^X, \end{aligned} \tag{15}$$

where $w_{m,i} = w_m(q_i)$. This makes the moments $F_{ij}^X = \int_{\mathbb{R}^n} I_i^X(\mathbf{u}) u_j d\mathbf{u}$, the image features of fundamental interest. Moreover, we can now write the moment matrix $\mathbf{T}(X)$ as

$$\begin{aligned} \mathbf{T}(X) &= \mathbf{W}\mathbf{F}^X;\ \mathbf{W} = \{w_{m,i}\} \in \mathbb{R}^{M \times Q}, \\ \mathbf{F}^X &= \{F_{ij}^X\} \in \mathbb{R}^{Q \times (n+1)}, \end{aligned} \tag{16}$$

where $\mathbf{F}^X$ may be called the *fundamental RIUME representation matrix* for image X. Since $M \leq Q$, the role of $\mathbf{W}$ is to transform the subspace $\langle \mathbf{F}^X \rangle \subset \mathrm{Gr}(Q, n+1)$ to a subspace $\langle \mathbf{W}\mathbf{F}^X \rangle$ in $\mathrm{Gr}(M, n+1)$. A single $\mathbf{W}$ has to serve for all the orbits $\psi_1, \ldots, \psi_K$.

We have thus reduced the problem of finding an optimal set of RIUME representations to a problem of finding $\mathbf{W}$. The special case where we choose $M = Q$, and set $\mathbf{W} = \mathbf{I}_M$ results from the choice of the w-functions to be indicator functions of the quantization levels.

B. The Optimal RIUME Operators: Grassmannian Dimensionality Reduction

We next wish to find the optimal set of w-functions that best separates the RIUME matrix representations of the different orbits. It is assumed that we have a set of K objects, such that $K \leq Q/(n+1)$ where for each object L observations are available. Applying the RIUME operator (6), using a set of functions chosen to be the indicator functions on the level-sets of the quantization levels, *i.e.*, $\mathbf{W} = \mathbf{I}_Q$, we find the fundamental RIUME representation for each of the observations, and obtain the set $\{\mathbf{F}^{k,j}\}_{k=1,j=1}^{k=K,j=L}$.

The procedure for finding the optimal set of w-functions is initialized by finding for every one of the K orbits its mean fundamental RIUME representation by evaluating $\bar{\mathbf{P}}_k$, the mean projection matrix, from the set $\{\mathbf{F}^{k,j}\}_{j=1}^{j=L}$; We next find among the available observations on orbit k some observation j such that its corresponding projection matrix $\mathbf{P}_{k,j}$ has the smallest distance to $\bar{\mathbf{P}}_k$. Thus observation j now becomes the representative of the kth orbit and $\mathbf{F}^k$ is the result of evaluating the fundamental RIUME representation of a "real" observation. The next step is to concatenate the K fundamental RIUME representative matrices into the *composite fundamental representation matrix* $\mathbf{F} \in \mathbb{R}^{Q \times K(n+1)}$, $K(n+1) \leq Q$

$$\mathbf{F} = \left[\mathbf{F}^1, \mathbf{F}^2, \ldots, \mathbf{F}^K\right]. \tag{17}$$

Since a single $\mathbf{W}$ has to serve for *all* the orbits, we define the companded representations

$$\mathbf{W}\mathbf{F} = \left[\mathbf{W}\mathbf{F}^1, \mathbf{W}\mathbf{F}^2, \ldots, \mathbf{W}\mathbf{F}^K\right]. \tag{18}$$

The optimal matrix $\mathbf{W} \in \mathbb{R}^{M \times Q}$ orthogonalizes the frames $\mathbf{W}\mathbf{F}^k$, $k = 1, 2, \ldots, K$ with respect to each other. So the problem is to choose $\mathbf{W} \in \mathbb{R}^{M \times Q}$ so that

$$(\mathbf{W}\mathbf{F})^T \mathbf{W}\mathbf{F} = \mathbf{I}_{K(n+1)}. \tag{19}$$

This makes the RIUME matrices $\mathbf{T}_k = \mathbf{W}\mathbf{F}_k$ orthogonal representations on the compact Stiefel manifold of $M \times (n+1)$ orthonormal matrices for all K image orbits under affine coordinate transformation; and the corresponding subspaces $\langle \mathbf{T}_k \rangle$ mutually orthogonal subspaces on $\mathrm{Gr}(M, n+1)$.

The solution for $\mathbf{W}$ is found from the SVD: $\mathbf{F} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \mathbf{\Sigma} = \mathrm{diag}[\sigma_1, \ldots, \sigma_{K(n+1)}]$. It is straightforward to show that in the case where M may be set to $K(n+1)$, the companding matrix

$$\mathbf{W} = \mathbf{\Sigma}^{-1}\mathbf{U}^T \in \mathcal{R}^{K(n+1) \times Q} \tag{20}$$

Authorized licensed use limited to: Joseph Francos. Downloaded on May 11,2022 at 07:45:53 UTC from IEEE Xplore. Restrictions apply.

is the solution for $\mathbf{W}_{opt}$, as indeed

$$\mathbf{F}^T\mathbf{W}^T\mathbf{W}\mathbf{F} = (\mathbf{V}\mathbf{\Sigma}\mathbf{U}^T)\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}^{-1}\mathbf{U}^T(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T) = \mathbf{V}\mathbf{V}^T = \mathbf{I}_{K(n+1)} \tag{21}$$

This solution is unique up to left multiplication by an orthogonal matrix. This amounts to rotating the entire space such that all distances between the subspaces remain the same.

There are several remarks to be made about this solution. First, in the noise-free case the derivation in (18)–(20) provides a methodology for jointly shaping the RIUME operators so that maximal separation is achieved between the representations of orbits of different objects, while minimizing the distance between observations that belong to the same orbit. This solution is achievable since the indicator functions on level-sets of the quantization levels span the entire space of w-functions when the observations are quantized. The solution preserves the dimension of the ambient space to be $Q = K(n+1)$. In the noise-free case, there is no need to have more companded images than $M = K(n+1)$. Second, it might be argued that there is no need to take the matrix $\mathbf{F}^T\mathbf{W}^T\mathbf{W}\mathbf{F}$ to identity, when taking it to a block-diagonal matrix of non-singular $(n+1)\times(n+1)$ blocks would be sufficient for orthogonality of the respective subspaces. The solution given here returns a block-diagonal matrix of $(n+1)\times(n+1)$identities for free. Thirdly, for $K(n+1) > Q$, in which case the matrix $\mathbf{F} \in \mathbb{R}^{Q\times K(n+1)}$ consists of more columns than rows, the problem is a problem of packing subspaces onto the Grassmannian in such a way that the intrinsic or extrinsic distances between subspaces are equalized, or perhaps the minimum distance is maximized. This problem is further complicated by the requirement that the base subspaces $\langle\mathbf{F}_k\rangle$ are fixed by the targets to be classified, and only the linear map $\mathbf{W} \in \mathbb{R}^{M\times Q}$ may be designed.

In the presence of noise, however, the suggested solution may be modified in order to control the noise contribution to classification errors. Hence, w-functions associated with subdominant modes of $\mathbf{F}$ that may be due to additive noise, may be excluded from the set of w-functions. We therefore search for the "knee" of the singular values in $\mathbf{\Sigma}$ and set to zero those singular values that are below a threshold. In that case M is reduced to r, the number of singular values above the threshold. Thus, in the presence of additive noise we evaluate $\mathbf{W}$ as

$$\mathbf{W}_r = \mathbf{\Sigma}_r^{-1}\mathbf{U}_r^T \tag{22}$$

where $\mathbf{\Sigma}_r$ is an $r\times r$ diagonal matrix containing the r dominant singular values, and the columns of $\mathbf{U}_r$ are composed of the r dominant left singular vectors of $\mathbf{F}$.

The subspaces $\langle\mathbf{W}_r\mathbf{F}_k\rangle$ are now taken to be estimates of signal subspaces. In this case, the orthogonality of the RIUME matrices $\mathbf{T}_k = \mathbf{W}_r\mathbf{F}_k$ and that of the corresponding subspaces $\langle\mathbf{T}_k\rangle$ on $\mathrm{GR}(r, n+1)$ is no longer guaranteed. This effect is similar to the one where an attempt to reduce an estimator "variance" results in increasing its "bias".

Remark: The derived procedure for designing the optimal w-functions in the presence of noise can be interpreted as a Grassmannian dimensionality reduction procedure aimed at improving the classification/registration/detection performance by mapping the problem from the original Grassmann manifold $\mathrm{GR}(M, n+1)$, where $M = Q$ to a lower dimensional ambient space Grassmann $\mathrm{GR}(r, n+1)$. So the problem stated in (18) is in fact to find the $\mathbf{W}$ and the reduced-dimension $r < M$ of the Grassmann ambient space where classification/registration/detection performance is optimized.

Finally, applying the designed w-functions (22) to the collection of available observations on each object and evaluating its mean G-invariant subspace by following the procedure derived in Section IV, we simultaneously obtain the G-invariant statistics of the denoised object manifolds for all orbits, without explicitly obtaining the denoised manifolds.

VI. Performance Evaluation on Synthetic Images

In this section, we conduct several synthetic experiments aimed at illustrating the procedure for designing the optimized w-functions. We demonstrate the performance gain of using the designed set of w-functions against the performance obtained by a naive choice of the w-functions as indicator functions. The synthetic set-up and the knowledge of the ground-truth and experimental parameters enable a detailed understanding of the effects of optimizing the RIUME operator.

A. Optimal Choice of the RIUME Operators

The experimental setup is this: Pick a noise-free image of an object. This image will serve as our noise-free ideal observation–a sample from the true manifold. This image is then deformed according to the assumed geometric deformation model, and zero-mean white Gaussian noise (WGN) is added to the observation (The noise std is 12 gray levels; amplitudes greater than 255 are clipped to 255 and negative amplitudes are clipped to zero). The top-left images in Fig. 2 and Fig. 3, as well as the images in Fig. 4 provide typical examples of images from the set. The number of different objects is $K = 30$. The training set is composed of 40 observations on each object, obtained by applying different affine geometric deformations to the original observation followed by adding zero-mean WGN as described above. The training set was employed to evaluate the fundamental RIUME representation of each observation, the mean fundamental RIUME representation of each object, and to design the optimal w-functions. In order to demonstrate the effect of quantizing the noisy images, in this experiment $Q = M = 100$, and hence the fundamental RIUME matrix representation is computed using 100 w-functions, chosen to be indicator functions on the level-sets, uniformly spread over the entire [0,255] range of amplitude levels. The optimal set of w-functions is then designed using the procedure of Section V. In this experiment the number of dominant singular values was found to be 10, and hence the number of optimal w-functions is 10. Therefore, using the set of optimal w-functions the dimension of the ambient space is $r = 10$. As a consequence, the extracted matrix $\mathbf{T}$ for any new image to be classified has dimension 10×3.

In Fig. 1 we depict the set of 6 optimal w-functions that correspond to the 6 most dominant singular values. The fundamental RIUME matrix representation is computed using 100 w-functions, chosen to be indicator functions uniformly spread

Authorized licensed use limited to: Joseph Francos. Downloaded on May 11,2022 at 07:45:53 UTC from IEEE Xplore. Restrictions apply.

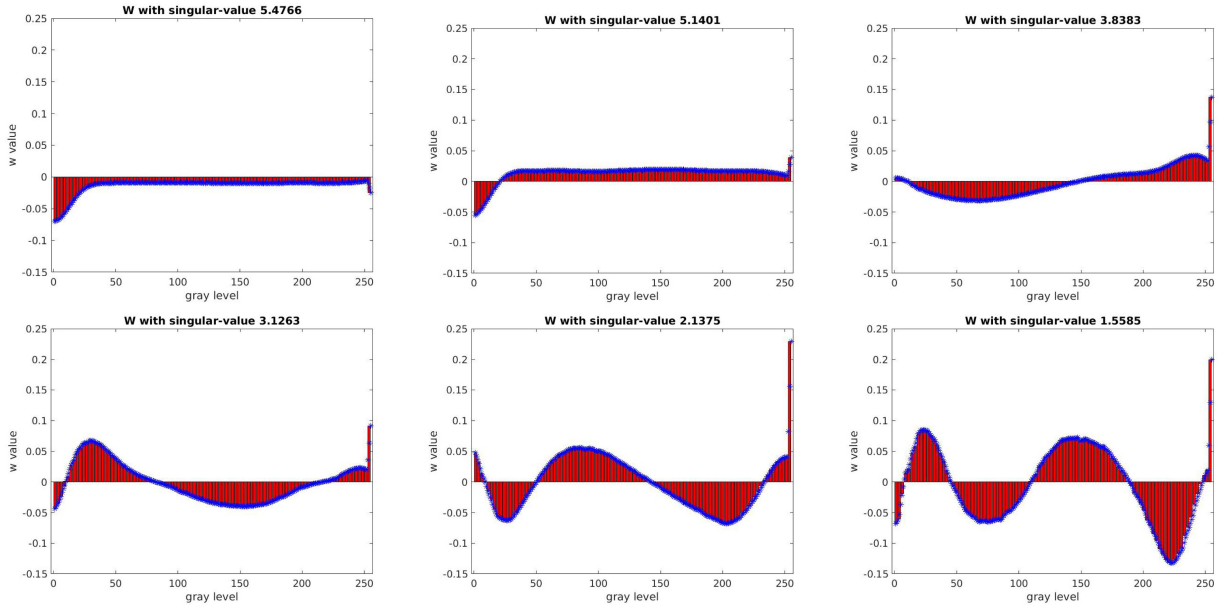

Fig. 1. w-functions that correspond to the six dominant singular values obtained using (22).

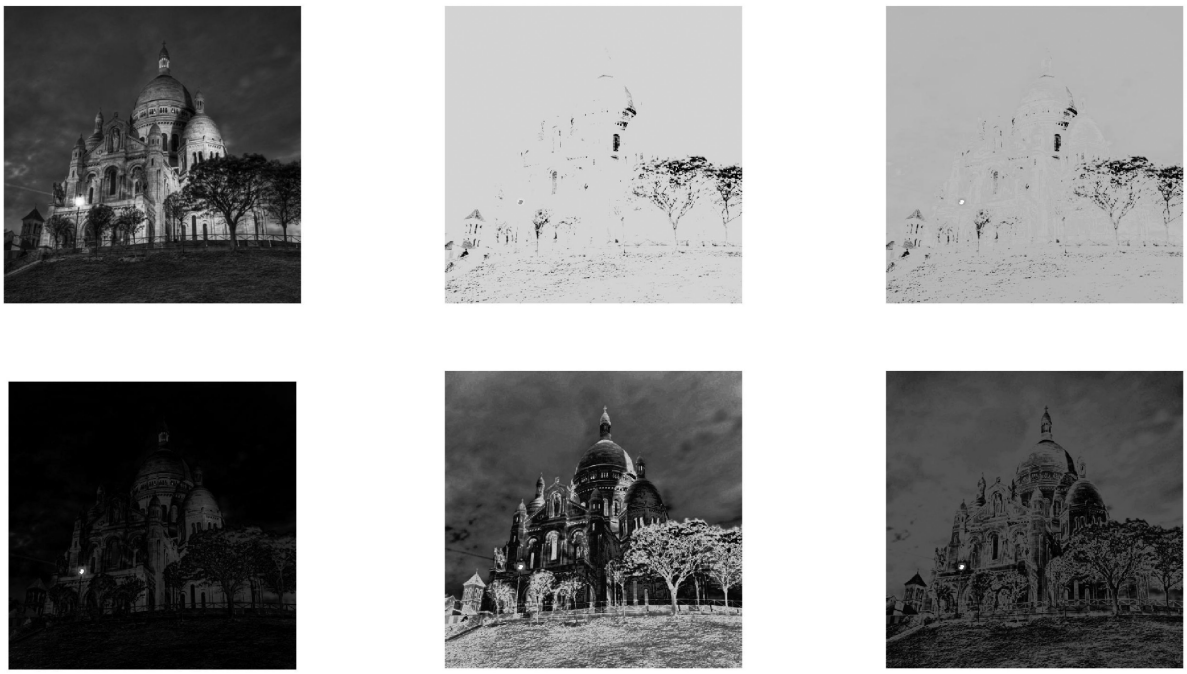

Fig. 2. Observed image and the results of applying the above 5 designed w-functions that correspond to the 5 most dominant singular values. Top row, from left to right: observed image X; $w_1 \circ X$; $w_2 \circ X$; Bottom row, from left to right: $w_3 \circ X$; $w_4 \circ X$; $w_5 \circ X$.

over the entire [0,255] range of amplitude levels and the plots in Fig. 1 depict the value $w(q_i)$ for each grey-level q_i.

In Fig. 2 we depict the results of applying the 5 optimal w-functions that correspond to the 5 most dominant singular values, to an example noise-free image. For display purposes the intensity levels of all images are scaled to the range 0–255. In Fig. 3, we depict the results of applying the same procedure to a deformed and noisy observation on the same image. It can be observed from the example images that each of the w-functions extracts different properties of the observation, while being covariant with the geometric transformation. We observe in these examples that while the optimal w-functions that correspond to the two most dominant singular values map the majority of the grey levels to nearly the same value, the next w-functions separate close grey level values in an oscillatory pattern, at increasing frequencies. This type of mapping implies that the designed w-functions aggregate pixels with substantially different gray levels to the same level-set.

Authorized licensed use limited to: Joseph Francos. Downloaded on May 11,2022 at 07:45:53 UTC from IEEE Xplore. Restrictions apply.

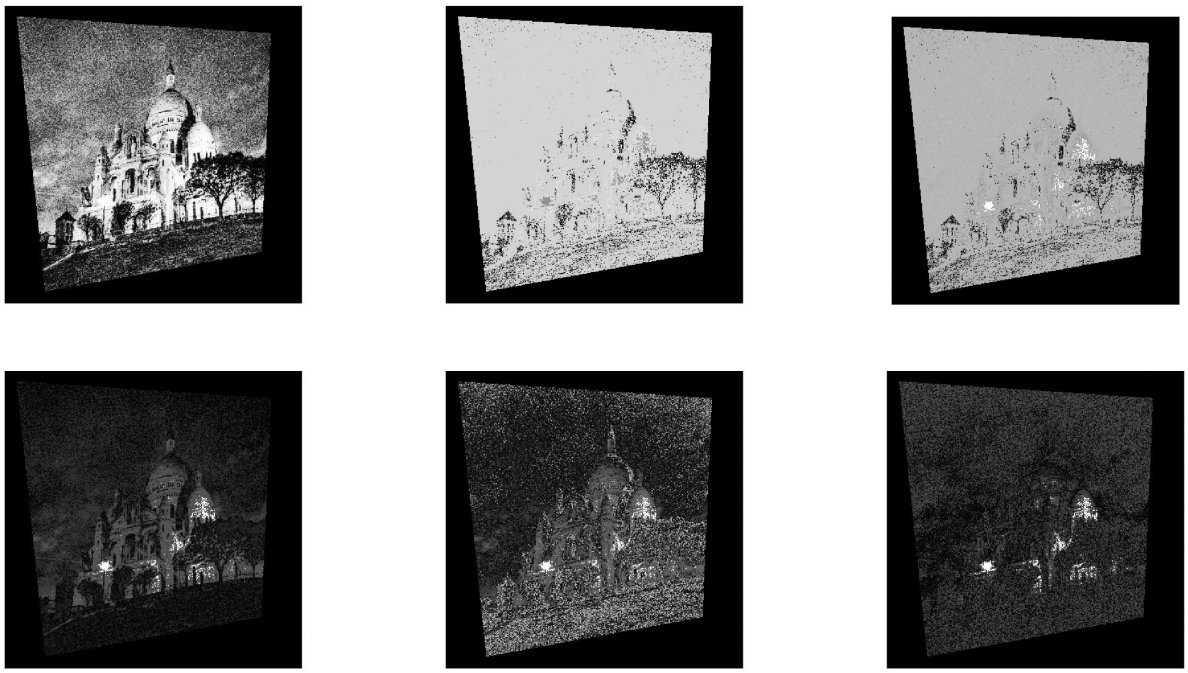

Fig. 3. Observed deformed and noisy image; and the results of applying the 5 designed w-functions that correspond to the 5 most dominant singular values. Top row, from left to right: observed image X; $w_1 \circ X$; $w_2 \circ X$; Bottom row, from left to right: $w_3 \circ X$; $w_4 \circ X$; $w_5 \circ X$.

Fig. 4. Example images from the test set employed to generate the results in Fig. 5 and Fig. 6.

B. Performance Evaluation of Applying the Optimal w-Functions

In this section we demonstrate in the context of a classifier design the performance gain of using the designed set of w-functions in comparison with the performance obtained by a naive choice of the w-functions as indicator functions.

A test set (different from the training set) composed of 40 deformed and noisy images was generated for each of the $K = 30$ objects. Example images from this test set are shown in Fig. 4. The geometric and radiometric deformations in the observations are assumed unknown. Throughout, distances between RIUME representations of different observations are evaluated using (8).

In the experiment to follow we evaluate the classifier performance by evaluating its ROC curve for 4 different classifier designs:

1) Use designed w-functions to compute the RIUME representation of an observation to be classified; Use the

Authorized licensed use limited to: Joseph Francos. Downloaded on May 11,2022 at 07:45:53 UTC from IEEE Xplore. Restrictions apply.

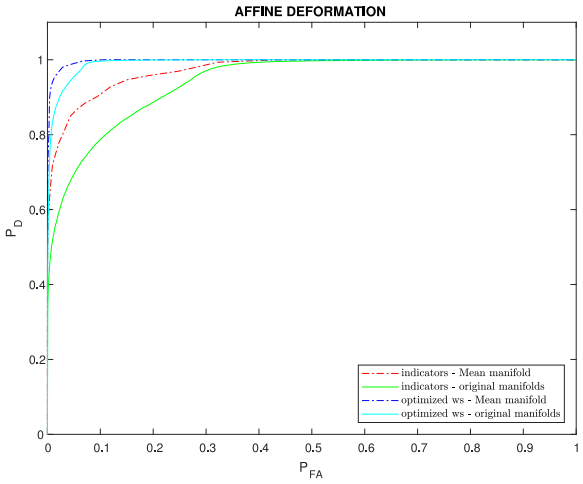

Fig. 5. The ROC curves of the four tested classification strategies on affine transformed observations: Blue - Classifier based on the minimum distance to the mean-RIUME representations of the different orbits, evaluated with optimized w-functions. Cyan - Classifier based on the minimum distance to the RIUME representations of all the observations in the training set, evaluated with optimized w-functions. Red - Classifier based on the minimum distance to the mean fundamental RIUME representations of the different orbits. Green - Classifier based on the minimum distance to the fundamental RIUME representations of all the observations in the training set.

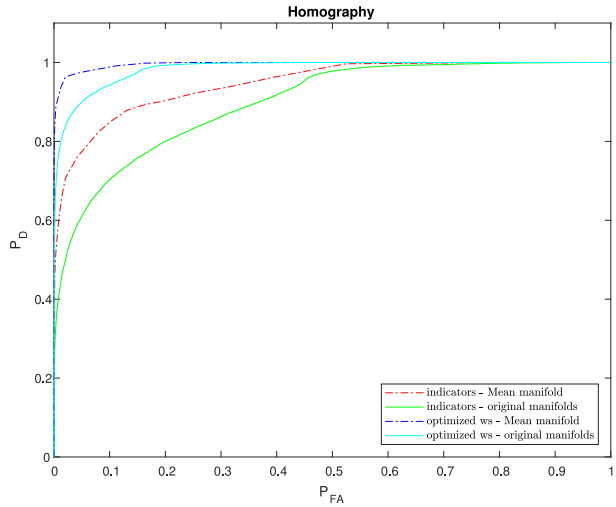

Fig. 6. The ROC curves of the four tested classification strategies on homography transformed observations: Blue - Classifier based on the minimum distance to the mean-RIUME representations of the different orbits, evaluated with optimized w-functions. Cyan - Classifier based on the minimum distance to the RIUME representations of all the observations in the training set, evaluated with optimized w-functions. Red - Classifier based on the minimum distance to the mean fundamental RIUME representations of the different orbits. Green - Classifier based on the minimum distance to the fundamental RIUME representations of all the observations in the training set.

mean-RIUME representation evaluated using the designed w-functions to represent each class; Decide the observation belongs to class j if the distance between its RIUME representation to the mean-RIUME representation of class j is the minimum (depicted in blue).

2) Use designed w-functions to compute the RIUME representation of an observation to be classified; Use designed w-functions to compute the RIUME representations of the observations in the training set. Decide the observation belongs to object j if the nearest neighbor to its RIUME representation is associated with class j (depicted in cyan).
3) Set the w-functions to be the indicator functions on the level-sets and compute the fundamental RIUME representation of an observation to be classified; Use the mean-fundamental RIUME representation to represent each class; Decide the observation belongs to class j if the distance between its fundamental RIUME representation to the mean fundamental RIUME representation of object j is the minimum (depicted in red).
4) Set the w-functions to be the indicator functions on the level-sets and compute the fundamental RIUME representation of an observation to be classified; Compute the fundamental RIUME representations of all observations in the training set. Decide the observation belongs to object j if the nearest neighbor to its RIUME representation is associated with class j (depicted in green).

In evaluating the ROC, P_D is estimated by counting the number of correct decisions for class j out of the total number of appearances of class j, averaged over classes. P_{FA} is evaluated by counting the number of incorrect decisions for class j out of the total number of experiments in which object j did not appear, averaged over classes.

The results of the experiment are summarized by the ROC curves depicted in Fig. 5. We conclude that designing the set of w-functions improves the performance of the classifier when the classifier is based on the entire set of RIUME representations of noisy observations in the training set, and when the classifier is based on the mean-RIUME representation derived from the set of noisy observations. Importantly, the use of the mean RIUME subspace requires just one subspace comparison, and not many comparisons to many noisy RIUME subspaces for each orbit.

Moreover, we find that using the designed set of w-functions *jointly* with the mean-RIUME representations of the object manifolds provides the best performance among all the tested classification strategies.

To further test the robustness of the proposed optimization procedure of the w-functions to model-mismatches, we repeated the test described above but this time the geometric deformation applied to each observation is a homography, which is clearly off-the-model. From the experimental results summarized in Fig. 6, it is concluded that in this case as well, using the designed set of w-functions *jointly* with the mean-RIUME representations of the object manifolds provides the best performance among all the tested classification strategies.

VII. Robust Homography Estimation Using Local Matched Manifold Detection

In general, the observed scene is not a single plane undergoing an affine transformation, and the radiometric variations across observations are not necessarily monotonic. Nevertheless, almost any scene can be well approximated by its tessellation into tiles, such that two observations on the same tile are related

Authorized licensed use limited to: Joseph Francos. Downloaded on May 11,2022 at 07:45:53 UTC from IEEE Xplore. Restrictions apply.

by simultaneous affine transformation of coordinates and a monotonic mapping of the intensities. This amounts to locally approximating a projective camera by an affine camera, [28].

In this and the next section we demonstrate, on sequences obtained by moving a camera in a 3-D scene, the performance of registration and transformation estimation algorithms that employ the RIUME framework: It is shown that tessellating the image into tiles, and modelling local transformations due to camera movement as if each tile undergoes simultaneous affine transformation of coordinates and monotonic mapping of the intensities, the RIUME-based MMD provides dense registration for an *a-priori* unknown and un-modeled scene structure. We also demonstrate the performance gain of using the designed set of w-functions in comparison with the performance of the fundamental RIUME operator.

Thus, in the case where the goal is to estimate the homography between two observations on a planar surface, the first step of the proposed estimation procedure is to apply a point matching algorithm, *e.g.*, SIFT [24], in order to find tentatively corresponding scene points in the two images. Given the two sets of tentatively corresponding points, Delaunay triangulation is applied to one of the images in order to tessellate it into a set of disjoint tiles. Each of these tiles is assumed to be a planar surface, such that if a set of three points defining a triangle in one image indeed matches a set of three points in the other image, then the resulting triangular surfaces will be related by simultaneous affine transformation of coordinates and monotonic mapping of the intensities. As we demonstrate in Section VIII the approach of approximating joint continuous coordinate and intensity transformations, by a set of piecewise affine geometric and monotonic radiometric transformations can be applied to much more complex scenarios than the homography discussed here.

Let $\mathbf{H}$ denote the homography matrix relating the coordinates of two images I_1 and I_2 of the same planar surface (using homogeneous coordinates notation). Using a point-matching algorithm we obtain N hypothesized corresponding pairs of points between I_1 and I_2: $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$, $i = 1, \ldots, N$. We next choose a subset $\{\mathbf{y}_k\}_{k=1}^K$, $K \leq N$ of the original set of correspondences such that the points $\{\mathbf{y}_k\}_{k=1}^K$ are spread as uniformly as possible across the image. Applying the Delaunay triangulation to the set of points $\{\mathbf{y}_k\}_{k=1}^K$, a tessellation of I_1 is obtained. The tessellation of I_1 is then mapped to a hypothesized tessellation of I_2 based on the set of correspondences to $\{\mathbf{y}_k\}_{k=1}^K$. As a result, each triangle in I_1 is associated with an hypothesized matching triangle in I_2. The hypotheses on the similarity of pairs of triangular tiles are then tested using the matched manifold detector by evaluating (8) for each hypothesized corresponding pair of tiles. In order to account for the unknown scale of the objects, this stage is repeated, each time with a different average spacing between the selected points and hence with different tile dimensions.

We next provide empirical performance evaluation of the proposed robust homography estimator. The method is tested on three well known data-sets: Graffiti, Wall, and Light, [29], including a total of 18 images in 3 different scenarios, for which ground truth measurement of the true homography is available.

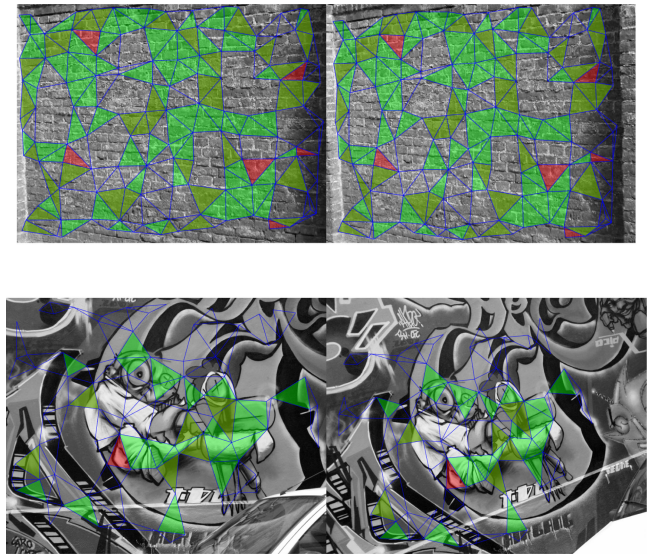

Fig. 7. Registration of homography related images. In each row, the blue lines on the left represent the Delaunay triangulation and the blue lines on the right represent the mapping of the triangulation according to the hypothesized point correspondences. Areas highlighted in green were found by the RIUME MMD to represent the same object using the optimized set of w-functions. Areas highlighted in red were found to represent the same object using the fundamental RIUME representations; The yellow triangles were identified as identical objects for both choices of w-functions.

Fig. 7 provides an example of the results obtained from a single stage of applying the RIUME MMD in two different cases and for different magnitudes of the geometric deformations. In each case, the two images, although taken from different view points and with different illumination, contain objects in common. The green shaded areas were identified as identical objects in both images using the optimized set of w-functions; The red shaded areas in both images were identified as identical objects using the fundamental RIUME representations; The yellow triangles were identified as identical objects for both choices of w-functions. 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 shaded). It is concluded that the piece-wise registration of the observations yields denser coverage when the optimized set of w-functions is employed. Since the scale of the scenes in both images is a-priori unknown, 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 correct scale. Hence, 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 foregoing procedure. The results of this dense matching procedure when the optimized RIUME operator is employed, vs. the results obtained using the fundamental RIUME operator are illustrated in Fig. 8 and Fig. 9. It is concluded that the optimized RIUME operator achieves a considerably denser and larger coverage of the overlapping areas in the images, and low

Authorized licensed use limited to: Joseph Francos. Downloaded on May 11,2022 at 07:45:53 UTC from IEEE Xplore. Restrictions apply.

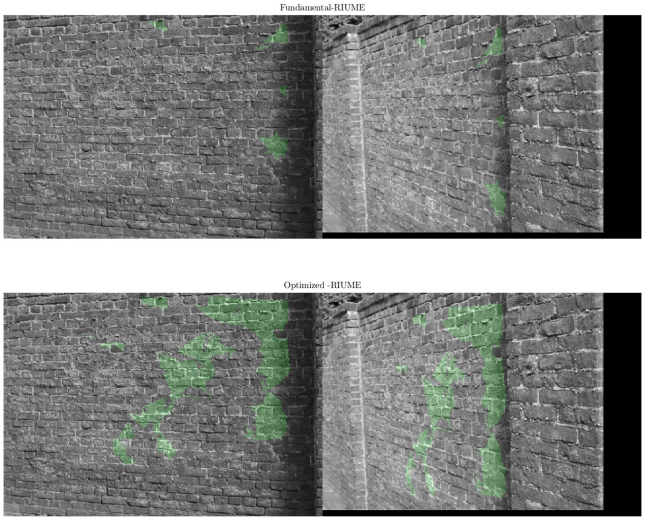

Fig. 8. Dense matching of related images using a locally affine geometric model and locally monotonic radiometric model. Top: using the fundamental RIUME operator and Bottom: using the optimized RIUME operator. The green shaded areas in both images were identified by the matched manifold detector as identical objects.

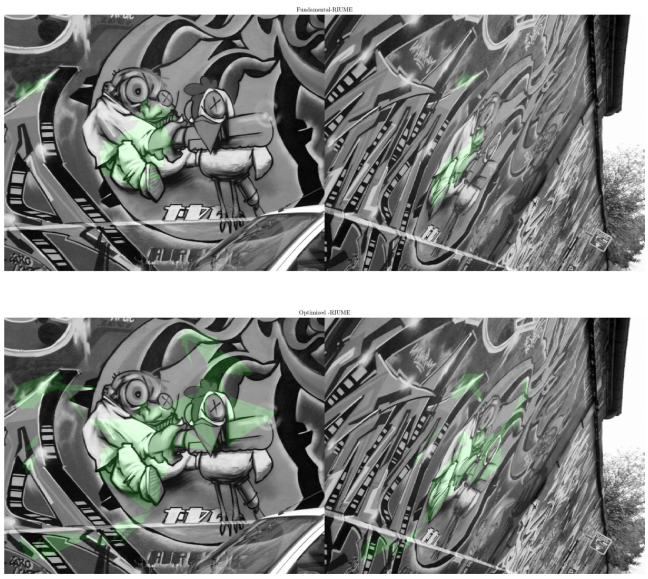

Fig. 9. Dense matching of related images using a locally affine geometric model and locally monotonic radiometric model. Top: using the fundamental RIUME operator and Bottom: using the optimized RIUME operator. The green shaded areas in both images were identified by the matched manifold detector as identical objects.

false coverage rates. Hence, it provides a larger coverage of correctly matched points.

The points that were found to match based on the locally affine model (see for example the green-shaded points in Fig. 8 and Fig. 9) are employed in the next stage to obtain $\hat{\mathbf{H}}$, an estimate of the homography $\mathbf{H}$, using the DLT linear estimation algorithm, [28]. The steps of the MMD-DLT robust estimator are summarized in Algorithm 1.

Algorithm 1: Robust MMD-DLT Homography Estimation.

1: Find hypothesized point correspondences between I_1 and I_2.
2: Uniformly dilute correspondences.
3: Tessellate the image according to the diluted point correspondences.
4: Match triangles using the RIUME based matched manifold detector (MMD).
5: Repeat steps 2-4 with varying spacing between the chosen matches.
6: Obtain $\hat{\mathbf{H}}$ from the DLT algorithm using the points in triangles that were found to match.

Next, we numerically evaluate the performance of the MMD-DLT homography estimator, implemented using the fundamental RIUME operator, and its implementation based on the optimized RIUME operator, relative to ground truth measurements of the homography transformations. The performance is also compared to the performance of the standard homography-RANSAC estimator, [28]. We further compare the performance of these methods to the performance of two state of the art DNN-based Deep-Homography estimators [30] and [31]. We note that the deep-homography estimation networks were designed and optimized for the specific task of homography estimation. The MMD-DLT homography estimator is an implementation using the DLT algorithm and the RIUME-MMD with no special training or adaptation of the RIUME-MMD to the case of homography estimation.

For the two variants of the MMD-DLT homography estimator we employ SIFT to obtain tentative point matches, and then follow Steps 1 to 6 of Algorithm 1. In the standard homography-RANSAC implementation, the same tentative point matches obtained by the SIFT algorithm are employed. The quantitative analysis presented in Table I was evaluated on two datasets for which ground truth information is available: The first is "RE" [31], for which manually labeled corresponding points in homography related images are provided. The second dataset is "Graffiti" [29], where measurements of the ground truth homography relating any pair of observations is provided. In order to better assess the performance of the different methods, and since [31] is designed for small baseline homographies, we defined based on the ground truth homographies, two different subsets of Graffiti: One is made of pairs of images with a small baseline, and represents the scenario of "weak homography". The other subset is made of pairs of images with a wide baseline, and represents the scenario of "strong homography". The error metric in Table I is the average ℓ_2 norm of the reprojection error (in pixels) between a point as projected by the estimated homography and its labeled ground-truth corresponding point (also used by [31]). For each dataset the results of the best performing method appear in bold.

We conclude from the results in Table I that for the "RE" dataset, on which the two Deep Homography methods were trained and optimized, the performance of all tested methods is similar, with a slight advantage to the Deep Homography

Authorized licensed use limited to: Joseph Francos. Downloaded on May 11,2022 at 07:45:53 UTC from IEEE Xplore. Restrictions apply.

TABLE I
AVERAGED ℓ_2 REPROJECTION ERROR EVALUATED ON "RE" AND "GRAFFITI" DATASETS

Method	RE	Graffiti-Small Baseline	Graffiti-Wide Baseline	Graffiti - All
SIFT + RANSAC	2.15	1.14	3.48	2.49
Unsupervised Deep Homography [30]	1.88	51.81	170.01	115.7
Content-Aware Unsupervised Deep Homography [31]	**1.81**	106.23	211.19	164.78
MMD-DLT (Fundamental RIUME)	2.11	1.09	1.57	1.5
MMD-DLT (Optimized RIUME)	2.09	**0.99**	**1.47**	**1.41**

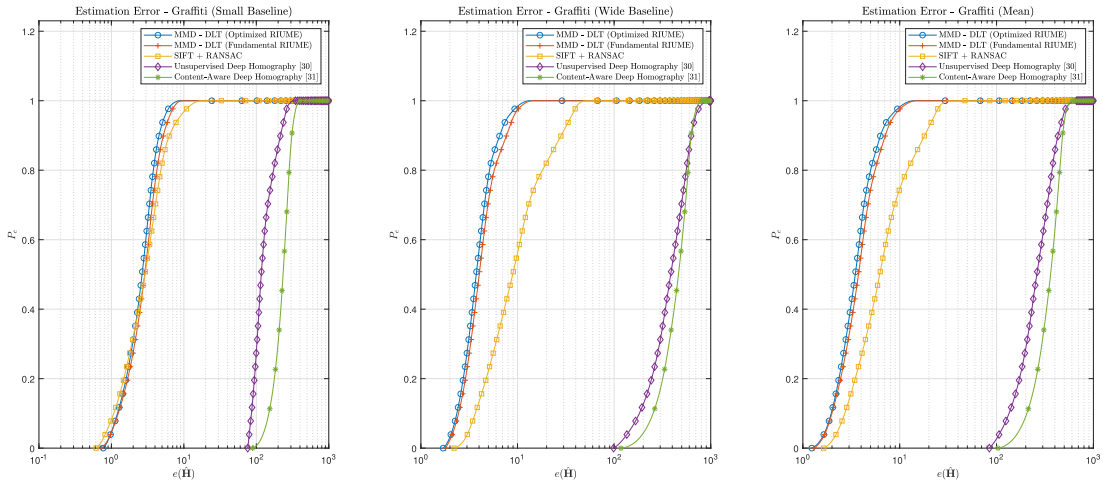

Fig. 10. Experimental cumulative probability distribution of the error in estimating the homography by the evaluated estimators, for the small baseline subset, wide baseline subset, and the entire dataset. e is defined in (23).

methods. However, on the Graffiti dataset the two versions of MMD-DLT, and the standard homography-RANSAC estimator outperform the Deep Homography methods, even for the case of small baseline (weak homography). The experimental results further indicate that the MMD implemented using the optimized RIUME operator outperforms the other estimators with a significant advantage in the more difficult scenarios of wide baseline homography.

In order to provide a more detailed evaluation of the performance of the MMD-DLT estimator, in comparison to that of the alternative estimators, Fig. 10 provides the experimental cumulative distribution function of the error between the location of each image point based on the true homography relative to its estimated location, evaluated on the Graffiti dataset. More specifically, the employed point-wise error metric (in pixels) is defined as

$$e(\hat{\mathbf{H}}) = \left\|(\mathbf{H} - \hat{\mathbf{H}})\mathbf{x}\right\|_2 + \left\|(\mathbf{H}^{-1} - \hat{\mathbf{H}}^{-1})\mathbf{x}'\right\|_2 \qquad (23)$$

where $\mathbf{H}$ is the ground truth homography and $\hat{\mathbf{H}}$ is its estimate.

As shown in Fig. 10, the probability mass of the MMD-DLT estimation errors is concentrated at low errors. It is concluded that both versions of the MMD-DLT homography estimator outperform the homography-RANSAC estimator and the Deep-Homography solutions. We further conclude using Fig. 10 and Table I that the use of the optimized RIUME in the MMD-DLT algorithm provides performance gain relative to the performance of the MMD-DLT algorithm implemented using the fundamental RIUME operator. This gain, is more significant in the more challenging scenarios where the baseline between observations is large, and deformations are larger. The gain is achieved due to the denser and larger coverage of correctly matched points as demonstrated in Fig. 8 and Fig. 9.

Finally, note that the optimized w-functions were obtained using the derivation in Section V-B aimed at finding the optimal set of w-functions that best separates the RIUME representations of the different objects considered. They were trained using the experimental set-up described in Section VI-A, that contains none of the images analyzed in this section. Thus, the experimental results of this section, on homography estimation and dense registration, demonstrate the generality and applicability of the design procedure of the optimized w-functions, for both classification and registration.

VIII. MATCHED MANIFOLD DETECTION FOR DENSE WIDE BASELINE REGISTRATION

Next, we demonstrate the effectiveness of the optimized Matched Manifold Detector for wide baseline registration of a complex scene, where the shapes of the objects are *a-priori* unknown and no closed-form model of the transformation is known. The method employs the tessellation-based MMD-and-registration. Each pair of frames is tesselated into tiles, and

Authorized licensed use limited to: Joseph Francos. Downloaded on May 11,2022 at 07:45:53 UTC from IEEE Xplore. Restrictions apply.

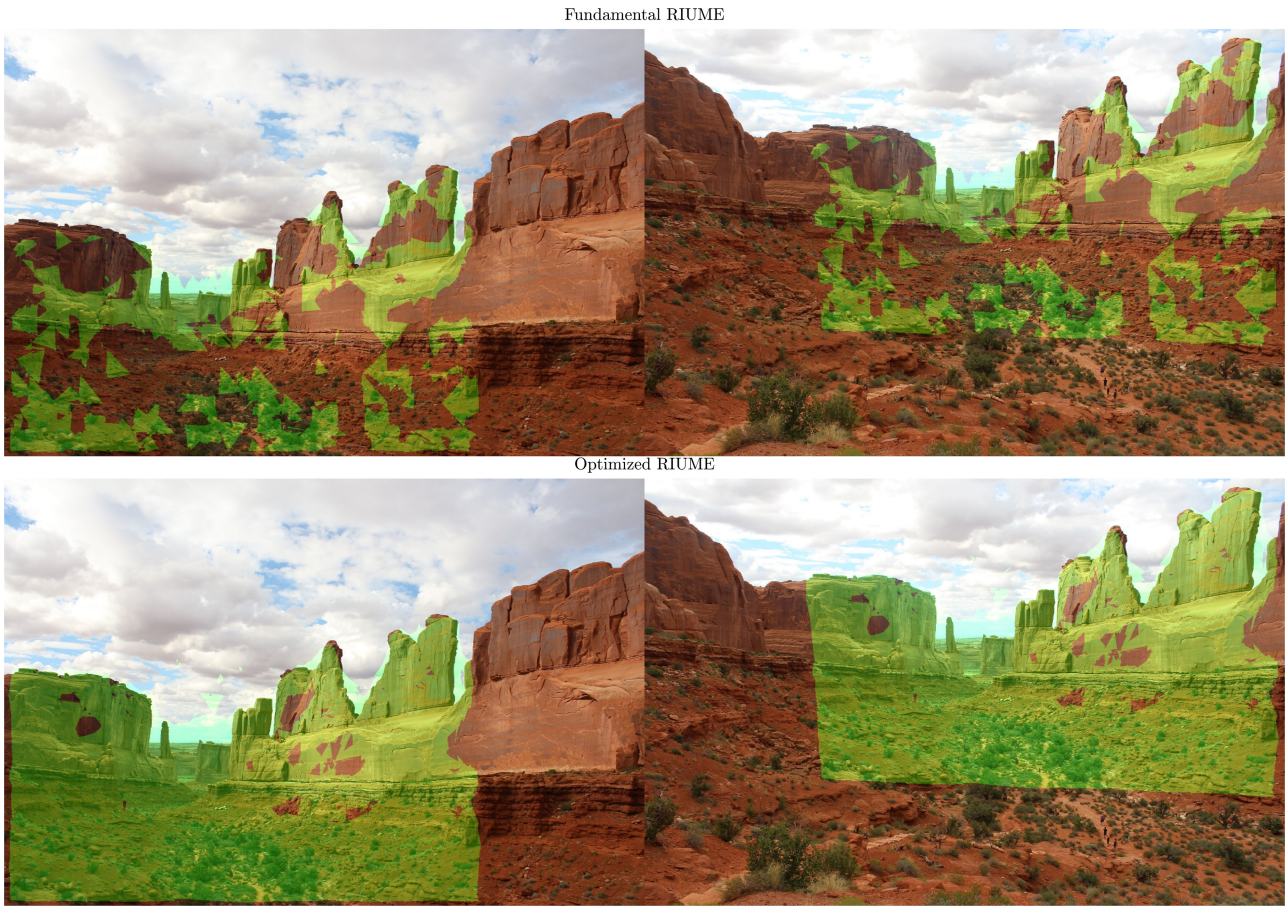

Fig. 11. Dense registration of wide base-line related images: The observed scene is tessellated into a set of tiles such that the deformation of each one is well approximated by an affine geometric transformation and a monotonic transformation of the measured intensities. Top: using the fundamental RIUME operator; Bottom: using the optimized RIUME operator. Green shaded areas in each of the images in a pair, were identified as identical objects.

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 one frame to the other. Points 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 Fig. 11 we provide the results of the dense registration of two images taken from different angles and distances from the scene and at different illumination conditions. We evaluate the results obtained by applying the fundamental RIUME operator, in comparison to the dense registration results of the optimized RIUME operator, for the same decision threshold in (8) when deciding whether the distance between the RIUME projection matrices of two triangular patches is indeed close enough to zero. Green shaded areas in both images were identified as identical objects in both images. It is concluded that the optimized RIUME operator provides a higher rate of correct detections, and hence a considerably larger and denser coverage of the overlapping areas in the images, at low false alarm rates. These results have many applications in structure-from-motion modelling problems.

IX. Summary

We have considered the problem of matched manifold detection and classification of noisy images, each undergoing geometric and radiometric deformations. For the case where the geometric deformation is affine and the radiometric deformation is monotonic, the RIUME maps the orbit of possible observations on each object to a *distinct linear subspace*. In the presence of observation noise, the observations do not lie strictly on the orbit and the resulting RIUME subspaces are fluctuations around the noise-free subspace. We have described a method for averaging these noisy subspaces in order to estimate the mean subspace representation for the orbit of each image under affine coordinate transformation. To optimize the performance of the matched manifold detector in the presence of observation noise, an analytic solution for choosing the RIUME operators is derived, such that the G-invariant representation of the denoised manifold is obtained without explicitly first obtaining the denoised manifold. It is shown that for object detection and classification, the effects of noise are reduced and the separability between observations originating from different orbits is improved by using the designed set of companding w-functions jointly with the mean subspace representation of each orbit.

Authorized licensed use limited to: Joseph Francos. Downloaded on May 11,2022 at 07:45:53 UTC from IEEE Xplore. Restrictions apply.

For registration applications, 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 characterizing them through the properties of small regions around them. These 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 and its variants to eliminate faulty matches. However, such a global geometric transformation model is often unknown. 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 small. The RIUME based matched manifold detection scheme provides a method for efficiently combining the advantages of the local, key-point methods, and the global methods. It thus enables dense matching and registration of complex scenes where the shapes of the objects are *a-priori* unknown and no closed-form model of the transformation is known.

REFERENCES

[1] R. Sharon, J. M. Francos, and R. Hagege, "Geometry and radiometry invariant matched manifold detection," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4363–4377, Sep. 2017.

[2] 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.

[3] S. Z. Kovalsky, R. Hagege, G. Cohen, and J. M. Francos, "Estimation of joint geometric and radiometric image deformations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 940–946, May 2010.

[4] "Special Issue on Dimensionality Reduction Methods," *Signal Process. Mag*., vol. 28, no. 2, Mar. 2011.

[5] P. Dollar, V. Rabaud, and S. Belongie, "Learning to traverse image manifolds," *Proc. Conf. Neural Inf. Proc. Syst.*, 2006.

[6] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.

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

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

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

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

[11] J. Zhao, Y. Peng, and Y. Yan, "Steel surface defect classification based on discriminant manifold regularized local descriptor," *IEEE Access*, vol. 6, pp. 71719–71731, 2018.

[12] R. Vidal, "Subspace clustering," *Signal Process. Mag.*, vol. 28, no. 2, pp. 52–67, Mar. 2011.

[13] 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.

[14] M. Hein and M. Maier, "Manifold denoising," in *Proc. Conf. Neural Inf. Proc. Syst.*, 2007, pp. 561–568.

[15] S. Deutsch, A. Ortega, and G. Medioni, "Manifold denoising based on spectral graph wavelets," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2016, pp. 4673–4677.

[16] T. Marrinan, J. Beveridge, B. Draper, M. Kirby, and C. Peterson, "Finding the subspace mean or median to fit your need," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 1082–1089.

[17] E. Begelfor and M. Werman, "Affine invariance revisited," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2006, pp. 2087–2094.

[18] I. Santamaria, L. L. Scharf, C. Peterson, M. Kirby, and J. M. Francos, "An order fitting rule for optimal subspace averaging," in *Proc. IEEE Statist. Signal Process. Workshop*, 2016, pp. 1–4.

[19] Z. Yavo, J. M. Francos, I. Santamaria, and L. L. Scharf, "Estimating the mean manifold of a deformable object from noisy observations," in *Proc. IEEE Image, Video, Multidimensional Signal Process. Workshop*, 2016, pp. 1–5.

[20] V. Garg, I. Santamaria, D. Ramírez, and L. L. Scharf, "Subspace averaging and order determination for source enumeration," *IEEE Trans. Sig. Process.*, vol. 67, no. 11, pp. 3028–3041, Jun. 2019.

[21] Z. Yavo and J. M. Francos, "Geometry and radiometry invariant matched manifold detection and robust homography estimation," in *Proc. IEEE Statist. Signal Process. Workshop*, 2018, pp. 169–173.

[22] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed., John Hopkins Univ. Press. Baltimore, 1989.

[23] A. Edelman, T. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, 1998.

[24] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, 1999, pp. 1150–1157.

[25] R. Mukundan and K. R. Ramakrishnan, "Moment functions in image analysis," World Scientific, Singapore, 1998.

[26] J. Flusser, B. Zitova, and T. Suk, *Moments and Moment Invariants in Pattern Recognition*, Hoboken, NJ, USA: Wiley, 2009.

[27] F. Mindru, T. Tuytelaars, L. Van Gool, and T. Moons, "Moment invariants for recognition under changing viewpoint and illumination," *Comput. Vis. Image Understanding*, vol. 94, no. 1–3, pp. 3–27, 2004.

[28] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge, U.K.: Cambridge Univ. Press, 2003.

[29] K. Mikolajczyk *et al.*, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, no. 1/2, pp. 43–72, 2005.

[30] T. Nguyen, S. W. Chen, S. S. Shivakumar, C. J. Taylor, and V. Kumar, "Unsupervised deep homography: A. fast and robust homography estimation model," *IEEE Robot. Automat. Lett.*, vol. 3, no. 3, pp. 2346–2353, Jul. 2018.

[31] J. Zhang *et al.*, "Content-aware unsupervised deep homography estimation," *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 653–669.

Ziv Yavo received the B.Sc. and M.Sc. degrees in electrical and computer engineering from Ben-Gurion University, Beer Sheva, Israel, in 2015 and 2019, respectively. He is currently an Algorithm Engineer with Mobileye, an Intel company. His research focuses on geometric problems in computer vision.

Yuval Haitman received the B.Sc. degree in electrical and computer engineering from Ben-Gurion University, Beer Sheva, Israel, in 2020. He is currently working toward the M.Sc. degree at the Electrical and Computer Engineering Department, Ben-Gurion University. His research interests include image and point cloud registration, manifold learning, and geometric problems in computer vision.

Authorized licensed use limited to: Joseph Francos. Downloaded on May 11,2022 at 07:45:53 UTC from IEEE Xplore. Restrictions apply.

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, Israel, in 1982 and 1991, respectively. 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. 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, MA, USA, with Electrical and Computer Engineering Department, University of California, Davis, CA, USA, with Electrical Engineering and Computer Science Department, University of Illinois, Chicago, IL, USA, INRIA Sophia-Antipolis, Biot, France, and with Electrical Engineering Department, University of California, Santa Cruz, CA, USA. His current research interests include parametric modeling and estimation of multidimensional signals, image registration, estimation of object deformations form images, manifold learning, parametric modeling, and estimation of 2D random fields, random fields theory.

Louis L. Scharf (Life Fellow, IEEE) is currently a Research Professor of mathematics and Emeritus Professor of electrical and computer engineering with Colorado State University, Fort Collins, CO, USA. He holds a courtesy appointment in statistics. He has coauthored the books, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-Wesley, 1991, and *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*, Cambridge University Press, 2010. His co-authored book, *A First Course in Electrical and Computer Engineering*, Addison-Wesley, Reading, MA, 1990, was re-published by Connexions in 2008. His research interests include statistical signal processing and machine learning as it applies to space-time adaptive processing for radar, sonar, and communication, modal analysis for electric power monitoring; spectrum analysis for nonstationary times series modeling and hyperspectral imaging, and image processing for group-invariant classification and registration. He has made original contributions to matched and adaptive subspace detection, group-invariant signal processing, spectrum analysis, and reduced-rank signal processing. He was the recipient of various awards for his contributions to statistical signal processing, including the Technical Achievement and Society Awards from the IEEE Signal Processing Society, the Donald W. Tufts Award for Underwater Acoustic Signal Processing, the Diamond Award from the University of Washington, the 2016 IEEE Jack S. Kilby Medal for Signal Processing, and the Education Award from the IEEE Signal Processing Society in 2021.

Authorized licensed use limited to: Joseph Francos. Downloaded on May 11,2022 at 07:45:53 UTC from IEEE Xplore. Restrictions apply.