Geometry and Radiometry Invariant Matched Manifold Detection and Tracking

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

***Abstract*—We present a novel framework for detection, tracking and recognition of deformable objects undergoing geometric and radiometric transformations. 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. 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. 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 *distinct* linear subspace - invariant with respect to both the geometric and radiometric transformations. This 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.**

***Index Terms*—Matched manifold detection, Manifold Learning, Dimensionality Reduction, Principal angles.**

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.*, [6]. The common underlying idea unifying existing manifold learning approaches is that although the data is sampled and presented in a high-dimensional space, for example because of the high resolution of the camera sensing the scene, in fact the intrinsic complexity and dimensionality of the observed physical phenomenon is very low.

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 [7] linear methods for dimensionality reduction such as PCA and MDS generate faithful 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 [11], [12], and the references therein. The need 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 [9], [10] 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 [16] we presented an alternative 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 very 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) provides a set of nonlinear functionals 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 a high dimensional vector space. As such, there is *no loss* of information in the reduced dimension representation of the signals on the manifold.

The proposed 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. It 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 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.

In this paper we expand the framework of matched manifold detection based on the universal manifold embedding to the more general case where the observed object undergoes not only geometric transformations, but also radiometric transformations. More precisely, we consider the case where the observed object undergoes both an affine geometric transformation, and a monotonic radiometric transformation. However, in the presence of radiometric transformations, the geometry-induced low-dimensional manifold model, [16], for the set of possible observations becomes over simplified. Thus, in order to employ the geometry-induced manifold model, the observations must be projected onto the manifold by finding a transformation that makes the observation invariant to radiometric deformations while being covariant with the geometric transformation. This leads to the derivation of a *radiometry-geometry invariant matched manifold detector and tracker*. This detector is then further employed as the basic building block in tracking complex objects that undergo both geometric and radiometric deformations, where the observed surface is tessellated into a set of tiles such that *locally*, the deformation of each of the tiles is well approximated by an affine geometric transformation and a monotonic transformation of the measured amplitudes on the tile.

II. Radiometry Invariant Universal Manifold Embedding

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 such that

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

where U is invertible 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 specifically, let $\mathcal{O}$ be the space of observations (for example, images), let $\tilde{\mathcal{A}}$ be the set of possible geometric deformations, $\mathcal{U}$ the set of monotonic one-dimensional transformations, and let S be a set of objects. We assume that the observations are the result of the following procedure: We choose an object $s \in S$ 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 an object $s \in S$ such that it jointly undergoes an affine geometric deformation $\mathcal{A}$, and a monotonic radiometric transformation U, producing an observation (such as g or h, above). For a specific object $s \in S$ we will denote by ψ_s 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 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.

In this section we show, by construction, that under the above assumptions there exists a pair of maps $R : \mathcal{O} \to \mathcal{O}$ and $T : \mathcal{O} \to H$ such that H is a linear space, which we call the *reduced space*. This construction holds for every object $g \in S$. We call the map $T \circ R$, *radiometry invariant universal manifold embedding* as it universally maps each of the different manifolds, where each manifold corresponds to a single object, into a *distinct* linear subspace of H such that the overall map $T \circ R \circ \psi_s : \tilde{\mathcal{A}} \times \mathcal{U} \to H$ is linear. In other words, each manifold is mapped into a different linear subspace of H. The map R projects the entire set of possible observations that may result from monotonic amplitude transformations, for some fixed pose of the object, to unique point on the manifold which represents the orbit of geometry only deformations of the object. The map T then maps the result non-linearly such that the overall map $T \circ R$ maps any observation to a distinct linear subspace of H. Figure 1 schematically illustrates the concept of the radiometry invariant universal manifold embedding.

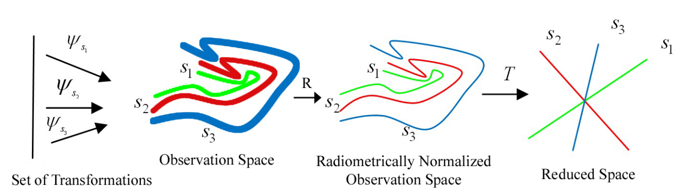

Fig. 1. Radiometry invariant universal manifold embedding.

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, $\mathcal{A} : \mathbf{x} \mapsto \mathbf{A}\mathbf{x} + \mathbf{c}$ where $\mathbf{A} \in \mathbb{R}^{n\times n}$ is non-singular, $\mathbf{c} \in \mathbb{R}^n$, and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, *i.e.*, $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$, $\mathbf{y} = [y_1, y_2, \ldots, y_n]^T$, such that $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{c}$, $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y} + \mathbf{b}$. $\mathcal{A}$ shall represent the geometric deformation. Let $U : \mathbb{R} \to \mathbb{R}$ be an invertible function, representing the non-linear radiometric deformation.

Denote by $B_c(\mathbb{R}^n)$ the space of bounded, compactly supported, Lebesgue measurable functions from $\mathbb{R}^n$ to $\mathbb{R}$ and let $g \in B_c(\mathbb{R}^n)$. Throughout, we shall use $\circ$ to denote the composition of functions, and supp$\{f\}$ shall be used to denote the support of a function f, *i.e.*, the closure of the set where f does not vanish. With these notations, the first problem addressed in this paper is the following: *Given two observations g and h related by (1), find a* representation *for g (and h) which is invariant to both the radiometric transformation U and the affine transformation $\mathcal{A}$.*

Let λ denote the Lebesgue measure on $\mathbb{R}^n$. Following [15] we define the *sample distribution transformation* V on $B_c(\mathbb{R}^n)$:

$$[Vg](t) = \frac{\lambda\{\mathbf{x} \in \text{supp}\{g\} : g(\mathbf{x}) \leq t\}}{\lambda\{\text{supp}\{g\}\}}, \qquad g \in B_c(\mathbb{R}^n). \tag{2}$$

The sample distribution may be thought of as a continuous "cumulative histogram" of a function. Define next an auxiliary operator R on $B_c(\mathbb{R}^n)$ by

$$Rh = [Vh] \circ h. \tag{3}$$

Since $[Vh](t)$ may be viewed as the "cumulative histogram" of h, roughly speaking, the operator R maps each value t in the range of h, to its "accumulated relative frequency", $[Vh](t)$.

Applying R to the basic relation $h = U \circ g \circ \mathcal{A}$ given in (1), we have, [15]

$$Rh = [Rg] \circ \mathcal{A}. \tag{4}$$

Hence, letting $\mathcal{H}(\mathbf{x}) = [Rh](\mathbf{x})$ and $\mathcal{G}(\mathbf{x}) = [Rg](\mathbf{x})$, for all $\mathbf{x} \in \mathbb{R}^n$ the following relation holds

$$\mathcal{H}(\mathbf{x}) = [\mathcal{G} \circ \mathcal{A}](\mathbf{x}) = \mathcal{G}(\mathcal{A}(\mathbf{x})). \tag{5}$$

Thus, (5) 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 their corresponding transformations, $\mathcal{H}(\mathbf{x})$ and $\mathcal{G}(\mathbf{x})$ are two "new" observations related by an affine transformation $\mathcal{A}$. We next derive new representations for $\mathcal{H}(\mathbf{x})$ and $\mathcal{G}(\mathbf{x})$ that are invariant to the unknown affine transformation.

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 $\mathbf{D} = \begin{bmatrix}\mathbf{b} \ \mathbf{A}^{-1}\end{bmatrix}$. Let $L \in \mathbb{N}$ and let $w_l \;\; l = 1, \ldots, L$ 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],

$$\int_{\mathbb{R}^n} x_k w_\ell \circ \mathcal{H}(\mathbf{x}) d\mathbf{x} = \left|\mathbf{A}^{-1}\right| \int_{\mathbb{R}^n} (\mathbf{D}_k\tilde{\mathbf{y}}) w_\ell \circ \mathcal{G}(\tilde{\mathbf{y}}) d\tilde{\mathbf{y}} \tag{6}$$

Let f be some observation on a deformable object and let

$$\mathbf{T}(f) = \begin{bmatrix} \int_{\mathbb{R}^n} w_1 \circ f(\mathbf{y}) & \int_{\mathbb{R}^n} y_1 w_1 \circ f(\mathbf{y}) & \cdots & \int_{\mathbb{R}^n} y_n w_1 \circ f(\mathbf{y}) \\ \vdots & & \ddots & \vdots \\ \int_{\mathbb{R}^n} w_L \circ f(\mathbf{y}) & \int_{\mathbb{R}^n} y_1 w_L \circ f(\mathbf{y}) & \cdots & \int_{\mathbb{R}^n} y_n w_L \circ f(\mathbf{y}) \end{bmatrix} \tag{7}$$

Denote $\tilde{\mathbf{D}} = [\mathbf{e}_1 \ \mathbf{D}^T]$ where $\mathbf{e}_1 = [1, 0, \ldots, 0]^T$. Then, if $\mathcal{H}$ is an observation of $\mathcal{G}$ undergoing an affine deformation represented by the matrix $\mathbf{D}$, then from (6) we get:

$$\mathbf{T}(\mathcal{G}) \left|\mathbf{A}^{-1}\right| \tilde{\mathbf{D}} = \mathbf{T}(\mathcal{H}) \tag{8}$$

Since the deformations at hand are invertible, this implies that the column space of $\mathbf{T}(\mathcal{G})$ and the column space of $\mathbf{T}(\mathcal{H})$ are the same subspace. Thus, after applying the mapping $\mathbf{T}$ to the space of observations "normalized" by the operator R, the problems of detection and recognition of objects undergoing both an affine geometric transformation, and a monotonic radiometric transformation become a problem of classifying subspaces.

Measuring the distance between two subspaces of a larger subspace is a well explored problem. 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 L. One way of measuring distance between subspaces is by principal angles. An alternative is to use the fact 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, we get that the projection matrix onto the column space of $\mathbf{A}$ is identical to the projection matrix onto the column space of $\mathbf{B}$. This enables us to measure the distance between subspaces by measuring the Frobenius norm of the difference between the projection matrices onto the different object subspaces. This metric provides the sum of the sinusoid squared of the principle angles.

Let P_f denote the projection matrix onto the $n+1$ dimensional column space of $\mathbf{T}(f)$ defined in (7). Using (4) and (8) we have thus proven the following theorem.

Theorem 1. *Let g and h be two observations on the same object, such that they are related by both an affine geometric transformation, and a monotonic radiometric transformation, i.e., $h(\mathbf{x}) = U(g(\mathcal{A}(\mathbf{x})))$. Then $P_{\mathcal{H}} = P_{\mathcal{G}}$.*

Theorem 1 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 $P_{\mathcal{H}}$ and $P_{\mathcal{G}}$ will be identical, while projection matrices that result from observations on other objects will be different and hence will yield non-zero distances form $P_{\mathcal{G}}$.

III. 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. Therefore the following approach is adopted (in order to simplify the explanation, we concentrate on the case where the observations are 2-D images): Given two observations that contain an object to be tracked, the first step in constructing the radiometry-geometry invariant matched manifold detection framework, is to apply some point matching algorithm in 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 each image 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 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. In general, the proposed algorithm can work on top of any state-of-the-art point matching algorithm. However, 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). The probability for a successful tile match naturally increases with the precision of these point matches. Thus, the proposed matched manifold tracker achieves significantly faster results when employed on top of more accurate point matching algorithm that includes lower rates of false matches. In this work, we used the expansion based matching approach introduced in [17], where matches are extracted with a low rate of false matches, and thus provide successful matches of many object tiles. Another advantage of using the expansion approach from [17], is that it supplies vaster coverage of the scene relative to other state-of-the-art solutions. This allows for a more complete coverage of the scene by matching tiles.

In practice, the locations of the correspondence points are noisy and hence do not exactly match even in the case where the triangles should indeed match. Hence, the above evaluation of the distance between surfaces through measuring the distance between their corresponding projection matrices, is employed in order to refine the estimates of the locations of the points defining the triangle sides. Using the metric for the distance between subspaces, false matches between triangular surfaces are efficiently rejected.

Figure 2 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. Each of these areas is, in fact, a union of triangular surfaces that were identified by the matched manifold detector to be identical. The vertices of the triangles are the initial point matches, employed to initialize the procedure. 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).

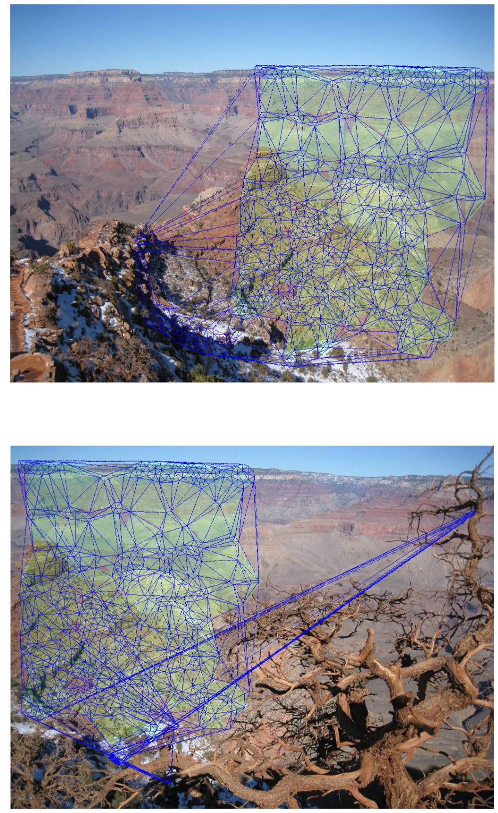

Fig. 2. Two images, taken from different view points but contain objects in common. The green shaded areas in both images were identified by the matched manifold detector as identical objects in both images.

IV. Conclusions

We have presented a novel framework for implementing geometry and radiometry invariant matched manifold detection and tracking. It is assumed that the observed surface undergoes both an affine geometric transformation, and a monotonic radiometric transformation, or that it can be tesselated into a union of such surfaces. By applying to each of the observations on a surface tile, 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 tile is mapped by the universal manifold embedding into a distinct linear subspace of some high dimensional Euclidean space. Thus, by tessellating the observed surface into a set of tiles it is shown that even in the general case where the observed surface deformation is highly complex, a sufficiently large number of triangular tiles can be detected on the surface such that by evaluating the Frobenius norm between the UME generated projection matrices, efficient detection and tracking of an object, under large deformations is achievable.

REFERENCES

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

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

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

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

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

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

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

[8] E. Vural and P. Frossard, "Learning Smooth Pattern Transformation Manifolds," *IEEE Trans. Image Process.,* pp. 1311-1325, 2013.

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

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

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

[12] 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.

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

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

[15] 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, pp. 940-946, May 2010.

[16] R. Sharon, J. M. Francos, R. Hagege, "Detection and Recognition of Deformable Objects Using Structured Dimensionality Reduction," *IEEE Int. Conf. Acoust., Speech, Signal Processing*, Brisbane, 2015.

[17] E. Farhan and R. Hagege, "Point matching via affine region expansion," *Proc. IEEE Int. Conf. Image Process.*, 2014.