Robust Topological Features for Deformation Invariant Image Matching

Edgar Lobaton
Ram Vasudevan
Ron Alterovitz
Ruzena Bajcsy

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2011-89
http://www.eecs.berkeley.edu/Pubs/TechRpts/2011/EECS-2011-89.html

August 5, 2011

Acknowledgement

This research is supported by the National Science Foundation under Awards CCR-0325274, CNS-0953823, CRA-0937060, ECCS-0931437, IIS-0703787, IIS-0724681, IIS-0840399, and IIS-0905344.

Technical Report: Robust Topological Features for Deformation Invariant Image Matching *

Edgar Lobaton[1], Ram Vasudevan[2], Ron Alterovitz[1], and Ruzena Bajcsy[2]

[1]Department of Computer Science
University of North Carolina, Chapel Hill
{lobaton,ron}@cs.unc.edu

[2]Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
{ramv,bajcsy}@eecs.berkeley.edu

Abstract

Local photometric descriptors are a crucial low level component of numerous computer vision algorithms. In practice, these descriptors are constructed to be invariant to a class of transformations. However, the development of a descriptor that is simultaneously robust to noise and invariant under general deformation has proven difficult. In this paper, we introduce the Topological-Attributed Relational Graph (T-ARG), a new local photometric descriptor constructed from homology that is provably invariant to locally bounded deformation. This new robust topological descriptor is backed by a formal mathematical framework. We apply T-ARG to a set of benchmark images to evaluate its performance. Results indicate that T-ARG significantly outperforms traditional descriptors for noisy, deforming images.

1. Introduction

Local photometric descriptors have found successful application in numerous areas such as object recognition [8], wide baseline matching [13], and image retrieval [10]. Traditionally these descriptors have been constructed in order to be invariant to a specific class of transformations while remaining robust to noise. In practice most have focused on the development of descriptors that are invariant under affine transformations as this is what occurs when a viewpoint changes relative to a rigid object with locally planar regions. Unfortunately, this class of transformations is unable to encapsulate the class of continuous deformations that describe how non-rigid objects transform, such as an animal moving its body or a cloth being folded. Observe that outside of occluding points, this model is able to describe the transformation between pairs of views of the same scene or the evolution of a deforming object seen from the same view. Under deformation, it is well known that the appropriate invariant is a topological one, i.e. the number of connected components or holes. However, such invariants have two principal shortcomings: first, they are not resilient to the presence of noise and second, they tend not to be distinct.

In this paper, we propose a novel framework for building a topological descriptor that is invariant under locally bounded deformations and addresses these two shortcomings. First, we make the topological invariants robust to noise by defining them in a local region over several inter-level sets of the intensity image. Second, we make the topological invariants distinct by describing the relational structure of nearby topological invariants. In so doing, we define a robust, distinct descriptor called the Topological- Attributed Relational Graph or *T-ARG*. Fig. 1 illustrates the performance of our descriptor on a typical pair of scenes.

*This research is supported by the National Science Foundation under Awards CCR-0325274, CNS-0953823, CRA-0937060, ECCS-0931437, IIS-0703787, IIS-0724681, IIS-0840399, and IIS-0905344.

Figure 1. Sample Results. The image on the right is constructed after deforming the left image by 30% and the red circles connected by red lines correspond to matches according to the algorithm presented in this paper.

1.1. Related Work

Due to space constraints, we refer the reader requiring an introduction to local descriptors to the comprehensive survey of the field by Mikolajczyk et al. [11]. They also provide a useful performance evaluation of several local descriptors including complex filters [12], gradient location and orientation histograms (GLOH) [11], shape context [1], scale invariant feature transform (SIFT) [9], spin image [4], and steerable filters [2]. In particular, SIFT and GLOH outperform all other descriptors. Our method can be loosely categorized with the so-called distribution based descriptors, like GLOH and SIFT, which use histograms to capture local image information. Importantly our method differs from these aforementioned approaches since it is invariant to all locally bounded deformations.

Others have attempted to develop local descriptors that are deformation invariant. The most promising such descriptor was developed by Ling et al. [6] who treat an intensity image as a surface embedded in 3D space, with intensity weighted relative to distance in the x, y plane. They show that as this weight increases, geodesic distances on the embedded surface are less affected by image deformation. They use geodesic sampling to construct a descriptor called the geodesic-intensity histogram (GIH). The method, though interesting, is not resilient to the presence of noise, since noise in either image does not scale well with respect to the weighting parameter. Our method on the other hand is provably robust to the presence of noise since it operates over several inter-level sets of the intensity image simultaneously. This paper is motivated by our recent work that defines a topological description of occlusions during deformation [7].

1.2. Overview

Our principal contributions are two-fold. First, in Section 3 we develop a notion of topological invariance under the assumption of locally bounded deformation. Second, in Section 4 we show how these topological invariants can be combined locally to define a descriptor that is distinct. The rest of the paper is organized as follows: Section 2 describes our imaging model and describes the type of properties we wish our descriptor to satisfy; Section 5 describes explicitly how to employ T-ARG to compare descriptors; Section 6 compares the performance of T-ARG to SIFT and GLOH using a precision versus recall metric; and Section 7 concludes the paper.

2. Image Model

In this section we introduce our imaging model and formalize the objectives of our work. Our analysis is done on grayscale images, but can be generalized to a multi-channel imaging modality in a straightforward manner. Suppose we are given two grayscale images $I_0, I_1 : \Omega \to \mathbb{R}$ defined over an image domain $\Omega \subset \mathbb{R}^2$ and related by:

$$I_1(x) = I_0 \circ f(x) + p(x), \tag{1}$$

where $p : \Omega \to \mathbb{R}$ is a scalar-valued function and $f : \Omega \to \Omega$ is a homeomorphism that satisfies:

$$\begin{aligned} |p(x)| &\leq K_p, \\ h_L(||x - x'||) &\leq ||f(x) - f(x')||, \text{ and} \\ ||f(x) - f(x')|| &\leq h_U(||x - x'||) \end{aligned} \tag{2}$$

for all $x, x' \in \Omega$, where $h_L, h_U : \mathbb{R}^+ \to \mathbb{R}^+$ are monotonic increasing functions with $h_L(\rho) \leq h_U(\rho)$ for all $\rho \in \mathbb{R}^+$, and $|| \cdot ||$ is the Euclidean norm. The functions p and f can be thought of as a bounded perturbation and a locally bounded deformation, respectively. The constant K_p is a perturbation bound, and the functions h_L and h_U bound the amount of local deformation. Observe that outside of points of occlusion, this model is able to describe the transformation between pairs of views of the same scene or the evolution of a deforming object. An example of such deformation functions are:

$$h_L(\rho) = (1 - K_d)\rho \quad \text{and} \quad h_U(\rho) = (1 + K_d)\rho \tag{3}$$

which corresponds to a **Lipschitz deformation model** with deformation constant K_d. This bounding function requires that deformations are bounded linearly with respect to the distance between points in an image. Throughout the rest of the paper, we assume that the images I_0 and I_1 satisfy the deformation model specified in Equations (1) and (2), where the perturbation bound K_p and the bounding deformation functions h_L, h_U are known. However, the actual perturbation $p(x)$ and deformation $f(x)$ functions are unknown.

Next, we describe explicitly the two problems we attempt to address using our descriptor.

Problem 1. *Given an arbitrary point $x_0 \in \Omega$ and a set of points $\bar{\Lambda}_1 \subset \Omega$, find the point $x_1 \in \bar{\Lambda}_1$ such that:*

$$||f^{-1}(x_0) - x_1|| = \min_{x \in \bar{\Lambda}_1} ||f^{-1}(x_0) - x||. \tag{4}$$

The solution to this problem has direct implications for wide baseline matching and image retrieval. When $\bar{\Lambda}_1$ is taken as uniform grid of points, we call the solution to Problem 1 a **grid matching point**. Denoting the area of the set E by $|E|$ and defining $B(x, r) = \{y \in \Omega \mid ||x - y|| \leq r\}$, our second problem is related to the first but attempts to explicitly identify neighborhoods rather than points with sufficient overlap:

Problem 2. *Given a threshold $\tau \in [0, 1]$, called the* **overlap threshold**, *neighborhoods $B(x, R_0) \subset \Omega$ and $B(x', R_1) \subset \Omega$ for some $x, x' \in \Omega$ and $R_0, R_1 > 0$, is*

$$\frac{|f^{-1}(B(x, R_0)) \cap B(x', R_1)|}{|f^{-1}(B(x, R_0)) \cup B(x', R_1)|} > \tau? \tag{5}$$

Note that the above quantity corresponds to the ratio between the area in the intersection and the union of the pair of neighborhood after mapping to the domain of I_1. This quantity is directly related to determining region correspondences which has important applications for object recognition and registration.

3. Set Filtrations

In this section, we introduce the results necessary to robustly characterize a neighborhood of a point in terms of topology. We begin by describing several results from algebraic topology, most importantly the homology group of a set. We then describe how our deformation model dictates the allowed transformations of the homologies for the pair of images. We conclude the section by strengthening the results on the transformation of homologies to local neighborhoods of the image.

3.1. Background

The objective of this section is to give a brief overview of algebraic topology. A comprehensive introduction to these ideas can be found in Chapter 2 of [3]. Algebraic topology explicitly characterizes the properties of spaces that are preserved under continuous deformation in terms of algebraic objects. Homology theory in particular transforms the study of topological invariants into the study of groups. If, for example, one wants to determine whether a pair of spaces are homeomorphic, one can transform the problem into determining whether a pair of groups are equivalent. In fact, by comparing the rank of the pair of groups, which is equal to the number of basis elements required to generate the group, one can effectively determine whether the pair of spaces are homeomorphic.

Naïvely comparing pairs of homology is generally insufficient to perform matching between pairs of images transforming under a homeomorphism for two reasons. First, though pairs of images maybe transforming under a homeomorphism the

effect of digitization (especially along edges) can ruin the applicability of homology. To address this deficiency, we define conditions on the homology over processed images. Second, homology is too coarse a construct. This is due to the fact that comparison between the homologies of different spaces is done via a counting argument and because homology is generally defined over entire spaces. We address homologies coarseness by localizing homology over intensity and space.

To understand these various extensions of homology, we must begin by describing homology more explicitly. The 0-**homology**, denoted $H_0(E)$, is a group whose rank is equal to the number of connected components in the space E. Whereas, the 1-**homology**, denoted $H_1(E)$, is a group whose rank is equal to the number of distinct cycles in the space E that cannot be shrunk via continuous deformation to a single point. Generalizing this notion let each of the k-**homology** groups of the space E be denoted $H_k(E)$. Suppose that one is given a map $\sigma : E_1 \to E_2$ between two spaces, we can in fact determine how topological properties transform under σ by considering the homomorphism (this generalizes the notion of a linear map to groups) induced by σ denoted $\sigma* : H_k(E_1) \to H_k(E_2)$. The case when $E_1 \subset E_2$ and σ is the **inclusion map** is called a **filtration** and is particularly important. To illustrate its utility consider the following result:

Lemma 1. *Given the filtration $E_1 \subset E_2 \subset E_3 \subset E_4$ and inclusion maps $\sigma_{i,j} : E_i \to E_j$ where $i < j$, then*

$$rank\left(\frac{H_k(E_1)}{ker\ \sigma_{1,4}*}\right) \leq rank\left(\frac{H_k(E_2)}{ker\ \sigma_{2,3}*}\right), \quad \forall k \geq 0, \tag{6}$$

where ker computes the kernel of its argument.

Proof. We split the proof into three steps:
Step I:

Let G_1, G_2 and G_3 be free abelian groups with finite rank, and let $\phi_{i,j} : G_i \to G_j$ for $i < j$ be homomorphisms such that

$$\phi_{1,3} = \phi_{2,3} \circ \phi_{1,2}. \tag{7}$$

By the first isomorphism theorem *c.f.*§1.3 in [5], we can define isomorphisms:

$$\pi_{1,3} : \frac{G_1}{ker\ \phi_{1,3}} \to \phi_{1,3}(G_1) \subset G_3 \tag{8}$$

and

$$\pi_{2,3} : \frac{G_2}{ker\ \phi_{2,3}} \to \phi_{2,3}(G_2) \subset G_3 \tag{9}$$

where $\phi_{1,3}(G_1) \subset \phi_{2,3}(G_2)$ by assumption. Then,

$$\alpha_{1,2} := \pi_{2,3}^{-1} \circ \pi_{1,3} : \frac{G_1}{ker\ \phi_{1,3}} \to \frac{G_2}{ker\ \phi_{2,3}} \tag{10}$$

is a 1-1 homomorphism, which implies that

$$rank\left(\frac{G_1}{ker\ \phi_{1,3}}\right) \leq rank\left(\frac{G_2}{ker\ \phi_{2,3}}\right) \leq rank\left(G_2\right). \tag{11}$$

Therefore,

$$rank\left(\frac{G_1}{ker\ \phi_{1,3}}\right) \leq rank\left(G_2\right). \tag{12}$$

Step II: Let G_1, G_2, G_3 and G_4 be free abelian groups with finite rank, and let $\phi_{i,j} : G_i \to G_j$ for $i < j$ be homomorphisms such that

$$\phi_{1,3} = \phi_{2,3} \circ \phi_{1,2} \tag{13}$$

and

$$\phi_{1,4} = \phi_{3,4} \circ \phi_{1,3}. \tag{14}$$

We can then define the following sequence of maps between groups

$$G_1 \overset{\beta}{\to} \frac{G_1}{ker\ \phi_{1,3}} \overset{\alpha_{1,2}}{\to} \frac{G_2}{ker\ \phi_{2,3}} \overset{\pi_{2,3}}{\to} G_3 \tag{15}$$

and

$$G_1 \xrightarrow{\gamma_1} \frac{G_2}{ker\ \phi_{2,3}} \xrightarrow{\gamma_2} G_4 \tag{16}$$

where $\beta : G_1 \to \frac{G_1}{ker \phi_{1,3}}$ is the natural surjective map, $\gamma_1 := \alpha_{1,2} \circ \beta$, $\gamma_2 := \phi_{3,4} \circ \pi_{2,3}$, and $\alpha_{1,2}$ and $\pi_{2,3}$ are defined as in the previous step. By construction we have that $\alpha_{1,2} := \pi_{2,3}^{-1} \circ \pi_{1,3}$, so

$$\gamma_2 \circ \gamma_1 = \phi_{3,4} \circ \pi_{1,3} \circ \beta = \phi_{3,4} \circ \phi_{1,3} = \phi_{1,4}. \tag{17}$$

Hence, by applying the results from Step I to the sequence in Equation 16, we have that

$$rank\left(\frac{G_1}{ker\ \phi_{1,4}}\right) \leq rank\left(\frac{G_2}{ker\ \phi_{2,3}}\right). \tag{18}$$

Step III:

From algebraic topology *c.f.*§2.1 in [3], we know that

$$\sigma_{i,j}* : H_k(E_i) \to H_k(E_j), \tag{19}$$

for $i < j$ are homomorphisms such that

$$\sigma_{1,3}* = \sigma_{2,3} * \circ \sigma_{1,2}* \tag{20}$$

and

$$\sigma_{1,4}* = \sigma_{3,4} * \circ \sigma_{1,3} * . \tag{21}$$

The sequence of homeomorphisms satisfy the assumptions made for the groups G_i in Step II. Hence, we obtain the desired result. □

This result gives a straightforward method to quantify the topological structure that must by carried from E_2 to E_3 by analyzing the structure carried from E_1 to E_4. Importantly, notice that neither of the mappings from E_1 to E_2 or from E_3 to E_4 are needed in this result. In the next few subsections, we describe how this result can be used to extend homology to address its aforementioned deficiencies.

3.2. Global Filtration

Let us begin by defining a set of pre-processed images.

Definition 1. *Let the* **pre-processed images** *be defined as:*

$$\begin{array}{rcl} I_{i--}(x) & = & \inf_{y \in B(x, g_i(\rho))} I_i(y) \\ I_{i-}(x) & = & \inf_{y \in B(x,\rho)} I_i(y) \\ I_{i+}(x) & = & \sup_{y \in B(x,\rho)} I_i(y) \\ I_{i++}(x) & = & \sup_{y \in B(x, g_i(\rho))} I_i(y) \end{array} \tag{22}$$

for $i \in \{0, 1\}$, *where* $\rho \geq 0$ *and*

$$g_i(\rho) = \begin{cases} h_U(\rho) & \text{if } i = 0 \\ h_L^{-1}(\rho) & \text{if } i = 1 \end{cases} \tag{23}$$

We define the following **inter-level sets***:*

$$\begin{array}{rl} E_{i--} & = I_{i--}^{-1}[a + K_p, \infty) \cap I_{i++}^{-1}(-\infty, b - K_p] \\ E_{i-} & = I_{i-}^{-1}[a, \infty) \cap I_{i+}^{-1}(-\infty, b] \\ E_{i+} & = I_{i+}^{-1}[a, \infty) \cap I_{i-}^{-1}(-\infty, b] \\ E_{i++} & = I_{i++}^{-1}[a - K_p, \infty) \cap I_{i--}^{-1}(-\infty, b + K_p] \end{array} \tag{24}$$

for constants a *and* b *such that* $b - a > 2K_p$.

The inter-level sets are the objects upon which we perform homology computation and help us localize homology over each image's intensity space. Fig. 2 illustrates these pre-processed images. Their corresponding inter-level sets are drawn in Fig. 3. The inter-level sets satisfy certain properties:

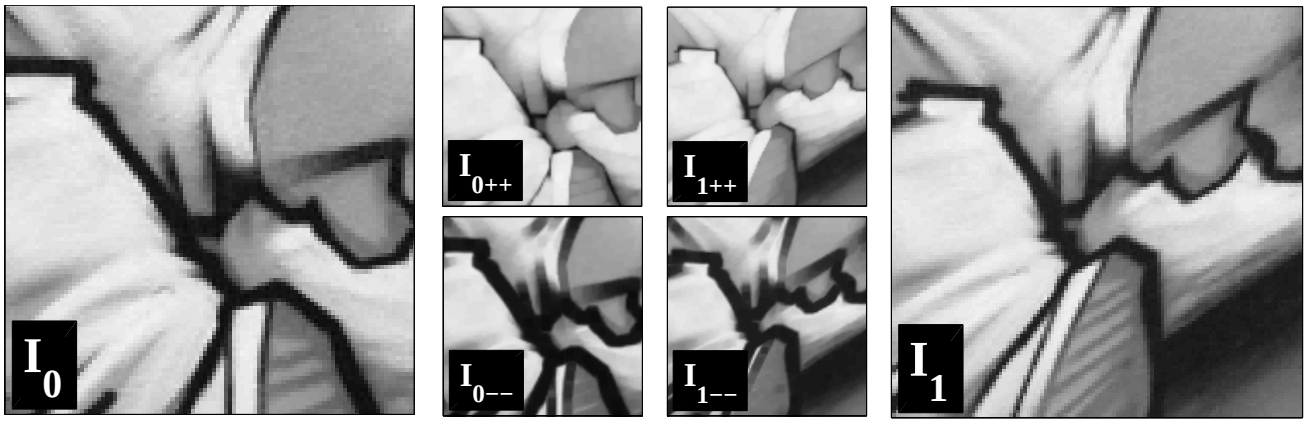

Figure 2. Pre-processed Images. Original images I_0 (left) and I_1 (right). Illustration of several pre-processed images (middle). These pre-processed images are defined to compensate for digitization effects along edges.

Lemma 2. *The inter-level sets for any a and b such that $b - a > 2K_p$ satisfy:*

$$\begin{array}{c} E_{0--} \subset f(E_{1-}) \subset f(E_{1+}) \subset E_{0++} \\ f(E_{1--}) \subset E_{0-} \subset E_{0+} \subset f(E_{1--}) \end{array}. \tag{25}$$

Proof. Note that given our deformation model we have that

$$I_{0--}(x) - K_p \leq I_{1-} \circ f^{-1}(x) \leq I_{1+} \circ f(x) \leq I_{0++}(x) + K_p, \tag{26}$$

which implies

$$I_{0--}^{-1}[a + K_p, \infty) \subset f\left(I_{1-}^{-1}[a, \infty)\right) \subset f\left(I_{1+}^{-1}[a, \infty)\right) \subset I_{0++}^{-1}[a - K_p, \infty) \tag{27}$$

and

$$I_{0++}^{-1}(-\infty, b - K_p] \subset f\left(I_{1+}^{-1}(-\infty, b]\right) \subset f\left(I_{1-}^{-1}(-\infty, b]\right) \subset I_{0--}^{-1}(-\infty, b + K_p]. \tag{28}$$

Hence,

$$E_{0--} \subset f(E_{1-}) \subset f(E_{1+}) \subset E_{0++}. \tag{29}$$

The other inclusions can be proven similarly. □

The result of the previous lemma ensures our choice of pre-processed images allows us to define a filtration between sets in images 0 and 1. As a result of Lemma 1, we obtain the following result:

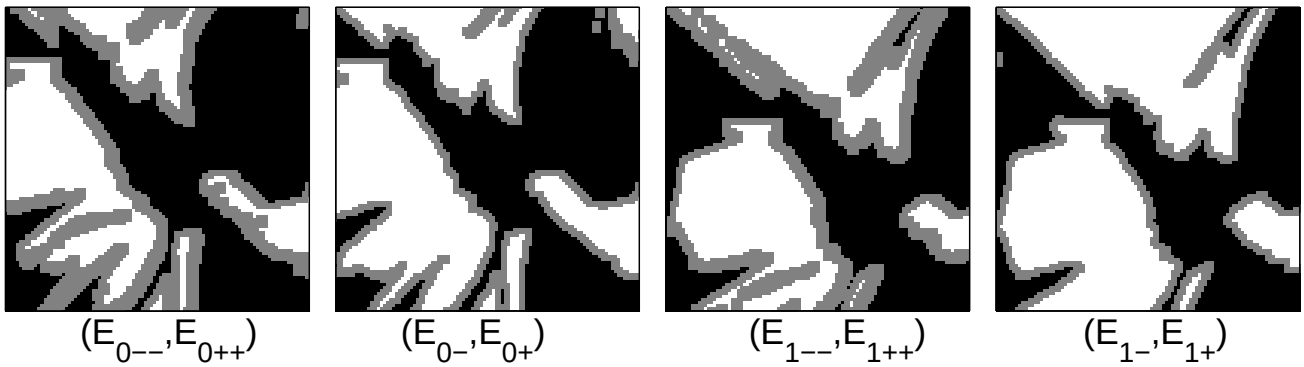

Figure 3. Inter-level sets for images 0 and 1 from Fig. 2. The left panel shows the set E_{0--} in white, and the set E_{0++} in gray. The other plots follow the same labeling convention. As stated by Theorem 1, $rank(\frac{H_0(E_{1--})}{ker\sigma_{1--,1++*}}) = 3$ (i.e. the number of connected components that persist from E_{1--} to E_{1++}) which is less than $rank(\frac{H_0(E_{0-})}{ker\sigma_{0-,0+*}}) = 5$.

Theorem 1. *Given images I_0 and I_1, and constants a, b such that $b - a > 2K_p$, then*

$$rank\left(\frac{H_k(E_{i--})}{ker\ \sigma_{i--,i++*}}\right) \leq rank\left(\frac{H_k(E_{j-})}{ker\ \sigma_{j-,j+*}}\right) \tag{30}$$

where $i, j \in \{0, 1\}$ such that $j \neq i$, and $\sigma_{i--,i++} : E_{i--} \to E_{i++}$ and $\sigma_{j-,j+} : E_{j-} \to E_{j+}$ are the inclusion maps.

Proof. By combining the set inclusion result from Lemma 2 and the rank results from Lemma 1 we have that

$$E_{0--} \subset f(E_{1-}) \subset f(E_{1+}) \subset E_{0++} \tag{31}$$

implies that

$$rank\left(\frac{H_k(E_{0--})}{ker\ \sigma_{0--,0++*}}\right) \leq rank\left(\frac{H_k(f(E_{1-}))}{ker\ \sigma_{1-,1+} * \circ f^{-1}*}\right). \tag{32}$$

By noting that $H_k(f(E_{1-})) = H_k(E_{1-})$ and $H_k(f(E_{1+})) = H_k(E_{1+})$ since f is a homeomorphism then we obtain the desired result for $i = 1$ and $j = 0$. The other result can be proven similarly. □

This theorem gives a computable condition in terms of the ranks of homologies that must be satisfied by the corresponding inter-level sets defined by image 0 and 1. To understand this result, observe that the rank of $\frac{H_0(E_{i-})}{ker\sigma_{i-,i+*}}$ is equal to the number of connected components in E_{i+} that have a non-empty intersection with E_{i-}. This can be understood as the number of components that persist from E_{i-} to E_{i+}. Hence, this theorem tells us that the number of connected components that persist from E_{1--} to E_{1++} is less than the number of components that persist from E_{0-} to E_{0+}. If this condition is violated for a pair of corresponding inter-level sets, then the pair of images that were used to construct these inter-level sets do not satisfy our image model. Observe that by defining the inter-level sets on pre-processed images, we avoid the problem of digitization effects along edges. An example illustrating an application of this theorem can be found in Fig. 3.

3.3. Spatially Localized Filtrations

The previous subsection gave a topological method to robustly determine whether entire images satisfied our deformation model. In this subsection, we define a spatial localization of this result. In order to obtain a localized characterization of an image, we begin by letting $\Lambda_1 = \{x_{1,ij}\} \subset \Omega$ denote a uniform grid of points spaced s_1 units apart. Our objective in this subsection is to construct a local set of conditions similar to those described in Theorem 1 that must be satisfied by a point in Λ_1 that is a solution to Problem 1, i.e. a grid matching point.

First, we obtain neighborhoods around a point x_0 and its grid matching point $x_1 \in \Lambda_1$ that satisfy a sequence of inclusions.

Lemma 3. *Given a point $x_0 \in \Omega$, a corresponding grid matching point x_1, radius $r_0 > h_U(s_1/\sqrt{2})$, and radius $r_1 > h_L^{-1}(h_U(s_1/\sqrt{2}))$, then*

$$B(x_0, r_{0-}) \subset f(B(x_1, r_1)) \subset B(x_0, r_{0+}) \tag{33}$$

and

$$B(x_1, r_{1-}) \subset f^{-1}(B(x_0, r_0)) \subset B(x_1, r_{1+}) \tag{34}$$

where

$$\begin{aligned} r_{1-} &= h_U^{-1}(r_0 - h_U(s_1/\sqrt{2})) \\ r_{0-} &= h_L(r_1) - h_U(s_1/\sqrt{2}) \\ r_{0+} &= h_U(r_1) + h_U(s_1/\sqrt{2}) \\ r_{1+} &= h_L^{-1}(r_0 + h_U(s_1/\sqrt{2})) \end{aligned}. \tag{35}$$

Proof. We split this proof into four steps.

Step I:

We show that given r_1 and only the deformation bounds h_L and h_U, the largest radius r that we can choose such that

$$B(x_0, r) \subset f(B(x_1, r_1)) \tag{36}$$

is $r = r_{0-}$.

First, we note that

$$B(f(x_1), h_L(r_1)) \subset f(B(x_1, r_1)) \tag{37}$$

by our deformation model. In order to have

$$B(x_0, r) \subset B(f(x_1), h_L(r_1)), \tag{38}$$

we must have

$$||x_0 - f(x_1)|| + r \leq h_L(r_1). \tag{39}$$

Since $||f(x_0) - x_1|| \leq s_1/\sqrt{2}$ which implies $||x_0 - f(x_1)|| \leq h_U(s_1/\sqrt{2})$, then the largest r that we can choose is $r = r_{0-}$.

Step II:

Given r_1, we note that

$$f(B(x_1, r_1)) \subset B(f(x_1), h_U(r_1)). \tag{40}$$

In order to have

$$B(f(x_1), h_U(r_1)) \subset B(x_0, r) \tag{41}$$

we need

$$||f(x_1) - x_0|| + h_U(r_1) \leq r. \tag{42}$$

Since $||f(x_1) - x_0|| \leq h_U(s_1/\sqrt{2})$, then the smallest r that we can choose such that

$$f(B(x_1, r_1)) \subset B(x_0, r) \tag{43}$$

is given by $r = r_{0-}$.

Step III:

Given r_0, we note that

$$f(B(x_1, r)) \subset B(f(x_1), h_U(r)). \tag{44}$$

In order to guarantee that

$$B(f(x_1), h_U(r)) \subset B(x_0, r_0) \tag{45}$$

we must have

$$||f(x_1) - x_0|| + h_U(r) \leq r_0. \tag{46}$$

Since $||f(x_1) - x_0|| \leq h_U(s_1/\sqrt{2})$, then the largest r that we can choose such that

$$f(B(x_1, r)) \subset B(x_0, r_0) \tag{47}$$

is given by $r = r_{1-}$.

Step IV:

Given r_0, we note that

$$B(f(x_1), h_L(r)) \subset f(B(x_1, r)). \tag{48}$$

In order to guarantee that

$$B(x_0, r_0) \subset B(f(x_1), h_L(r)) \tag{49}$$

we must have

$$||f(x_1) - x_0|| + r_0 \leq h_L(r). \tag{50}$$

Since $||f(x_1) - x_0|| \leq h_U(s_1/\sqrt{2})$, then the smallest r that we can choose such that

$$B(x_0, r_0) \subset f(B(x_1, r)) \tag{51}$$

is given by $r = r_{1+}$. □

As before, the inclusions in the previous lemma define a filtration in terms of inter-level sets.

Definition 2. *Let the* (r_0, r_1)**-localized inter-level sets** *between a point* $x_0 \in \Omega$ *and its grid matching point* x_1 *be defined as*

$$\begin{aligned} E_{i--}(x_i) &= E_{i--} \cap B(x_i, r_{i-}) \\ E_{i-}(x_i) &= E_{i-} \cap B(x_i, r_i) \\ E_{i+}(x_i) &= E_{i+} \cap B(x_i, r_i) \\ E_{i++}(x_i) &= E_{i++} \cap B(x_i, r_{i+}) \end{aligned} \tag{52}$$

for $i \in \{0, 1\}$, $r_0 > h_U(s_1/\sqrt{2})$, $r_1 > h_L^{-1}(h_U(s_1/\sqrt{2}))$. *The radii* r_{i-} *and* r_{i+} *are given as in Lemma 3.*

Note that the (r_0, r_1)-localized inter-level sets satisfy inclusion relations similar to the ones stated in Lemma 2 by construction. Hence, we have the following localized counterpart to Theorem 1:

Theorem 2. *Given images I_0 and I_1, constants a, b such that $b - a > 2K_p$, and radii $r_0 > h_U(s_1/\sqrt{2})$ and $r_1 > h_L^{-1}(h_U(s_1/\sqrt{2}))$, then for any $x_0 \in \Omega$ and grid matching point $x_1 \in \Lambda_1$ we have that*

$$rank\left(\frac{H_k(E_{i--}(x_i))}{ker\ \sigma_{i--,i++*}}\right) \leq rank\left(\frac{H_k(E_{j-}(x_j))}{ker\ \sigma_{j-,j+*}}\right) \tag{53}$$

where $i, j \in \{0, 1\}$ such that $j \neq i$, and $\sigma_{i--,i++} : E_{i--}(x_i) \to E_{i++}(x_i)$ and $\sigma_{j-,j+} : E_{j-}(x_j) \to E_{j+}(x_j)$ are the natural inclusion maps.

Proof. By combining the results from Lemmas 2 and 3 we obtain the filtrations

$$E_{0--}(x_0) \subset f(E_{1-}(x_1)) \subset f(E_{1+}(x_1)) \subset E_{0++}(x_0) \tag{54}$$

and

$$f(E_{1--}(x_1)) \subset E_{0-}(x_0) \subset E_{0+}(x_0) \subset f(E_{1--}(x_1)). \tag{55}$$

By following the same steps as in the proof for Theorem 1 we obtain the desired results. □

This result gives a way to identify the existence of a local homeomorphism between neighborhoods around x_0 and x_1. That is, if these conditions are not satisfied, then x_1 cannot be a point that corresponds to x_0 under our image model.

4. Topological Attributed Relational Graph

At this point, we in fact have a robust localized topological descriptor that could be used to perform matching. In this section, we describe how to construct a graphical representation that integrates the localized topological characterization developed so far to render the constructed descriptor more distinct.

Suppose as in the previous section, we have a grid of points Λ_1 placed s_1 units apart, a set of radii and constants $\{(r_{0,\lambda}, a_\lambda, b_\lambda)\}_{\lambda\in\Gamma}$ that satisfy the conditions in Theorem 2 where Γ is some indexing set, and let

$$r_{1,\lambda} = (h_L^{-1}(r_{0,\lambda}) + h_U^{-1}(r_{0,\lambda}))/2. \tag{56}$$

Let us also assume a uniform grid of points $\Lambda_0 = \{x_{0,\gamma}\} \subset \Omega$ spaced $s_0 > 2h_U(s_1/\sqrt{2})$ units apart, and let $\{x_{1,\gamma}\} \subset \Lambda_1$, be the set of corresponding grid matching points. The following theorem gives a set of bounds between the distance of points in Λ_0 and grid matching points in Λ_1:

Theorem 3. *Given the pair of point $(x_{0,\alpha}, x_{0,\beta}) \in \Lambda_0 \times \Lambda_0$ and corresponding matching points $(x_{1,\alpha}, x_{1,\beta}) \in \Lambda_1 \times \Lambda_1$, then*

$$h_U^{-1}(||x_{0,\alpha} - x_{0,\beta}|| - 2h_U(s_1/\sqrt{2})) \leq ||x_{1,\alpha} - x_{1,\beta}|| \tag{57}$$

and

$$||x_{1,\alpha} - x_{1,\beta}|| \leq h_L^{-1}(||x_{0,\alpha} - x_{0,\beta}|| + 2h_U(s_1/\sqrt{2})). \tag{58}$$

Proof. By using our deformation bounds, we get the following inequalities

$$\begin{aligned} ||x_{0,\alpha} - x_{0,\beta}|| &\leq ||x_{0,\alpha} - f(x_{1,\alpha})|| + ||f(x_{1,\alpha}) - f(x_{1,\beta})|| + ||f(x_{1,\beta}) - x_{0,\beta}|| && (59)\\ &\leq 2h_U(s_1/\sqrt{2}) + h_U(||x_{1,\alpha} - x_{1,\beta}||) && (60)\end{aligned}$$

which gives the desired lower bound for $||x_{1,\alpha} - x_{1,\beta}||$.

We also have that

$$\begin{aligned} h_L(||x_{1,\alpha} - x_{1,\beta}||) &\leq ||f(x_{1,\alpha}) - f(x_{1,\beta})|| && (61)\\ &\leq ||f(x_{1,\alpha}) - x_{0,\alpha}|| + ||x_{0,\alpha} - x_{0,\beta}|| + ||x_{0,\beta} - f(x_{1,\beta})|| && (62)\\ &\leq 2h_U(s_1/\sqrt{2}) + ||x_{0,\alpha} - x_{0,\beta}|| && (63)\end{aligned}$$

which gives the desired upper bound. □

For each of the points $x_{0,\gamma}$ and corresponding grid matching points $x_{1,\gamma}$, the conditions specified by Theorem 2 must be satisfied by all tuples $(r_{0,\lambda}, r_{1,\lambda}, a_\lambda, b_\lambda)$. Additionally, for each pair $(x_{0,\alpha}, x_{0,\beta}) \in \Lambda_0 \times \Lambda_0$ and corresponding grid matching pair $(x_{1,\alpha}, x_{1,\beta}) \in \Lambda_1 \times \Lambda_1$ the distance bounds in Theorem 3 must be satisfied.

To check the simultaneous satisfaction of all of these conditions, we can recast our problem by constructing a **Topological-Attributed Relational Graph** or **T-ARG**, $\mathcal{G}_0$ as follows: let the nodes of this graph represent the points $x_{0,\gamma}$ and label these nodes with the rank conditions defined in Theorem 2 and let the edges of this graph represent the distance between points and be labeled using the distance bounds from Theorem 3. We can construct a similar T-ARG, $\mathcal{G}_1$, using the points $x_{1,\gamma}$. In fact, we can define a T-ARG $\hat{\mathcal{G}}_1$ using all the points in Λ_1, the results of the Theorems 2 and 3 give constraints to determine a subgraph isomorphism from $\mathcal{G}_0$ to $\mathcal{G}_1 \subset \hat{\mathcal{G}}_1$. More explicitly a correspondence between the graph $\mathcal{G}_0$ and a subgraph in $\hat{\mathcal{G}}_1$ is defined by the satisfaction of Theorems 2 and 3. Hence, the identification of $\mathcal{G}_1$ turns into the matching of Attributed Relational Graphs for which the attributes are topological rank conditions and distance bounds.

5. Implementation

In this section, we begin by explicitly describing how we can solve the problems defined in Section 2 using the T-ARG. We also describe how we implement our algorithm. The executables used to perform this implementation can be found online [1].

Solution 1. *Given an arbitrary set of points $\bar{\Lambda}_1$, we can first construct a grid of points Λ_1 placed s_1 units apart such that $\bar{\Lambda}_1 \subset \Lambda_1$. Then given a point $x \in \Omega$ and another point $x' \in \Lambda_1$, we identify x' as a possible grid matching point if there is a subgraph isomorphism from a graph $\mathcal{G}_0$ into a graph $\hat{\mathcal{G}}_1$. $\mathcal{G}_0$ is constructed using the points in $\Lambda_0 \cap B(x, R_0)$, and $\hat{\mathcal{G}}_1$ uses the points in $\Lambda_1 \cap B(x', R_1)$, where $R_1 = h_L^{-1}(R_0 + 2h_U(s_1/\sqrt{2}))$ and $R_0 > 0$. It is then straightforward to re-project the solution in Λ_1 to the solution in $\bar{\Lambda}_1$.*

Note that R_1 is defined using Theorem 3 and R_0 is set arbitrarily based on how large of a neighborhood around x we are interested in considering.

To solve Problem 2, we are interested in determining the overlap ratio between sets $f^{-1}(B(x, R_0))$ and $B(x', R_1)$. Note that if x_1 is a grid matching point to x, then:

$$B(x_1, R_{1-}) \subset f^{-1}(B(x, R_0)) \subset B(x_1, R_{1+}) \tag{64}$$

where R_{1-} and R_{1+} are defined as in Lemma 3. Using Solution 1, we obtain a set of possible grid matching points to x. An upper bound to the overlap ratio between $B(x, R_0)$ and $B(x', R_1)$ can be computed by considering the maximum of the bounds using all possible grid matches.

Solution 2. *Given neighborhoods $B(x, R_0)$ and $B(x', R_1)$ where $x, x' \in \Omega$, we identify a possible matching set with overlap greater than τ if*

$$\frac{|B(x_1, R_{1+}) \cap B(x', R_1)|}{|B(x_1, R_{1-}) \cup B(x', R_1)|} > \tau, \tag{65}$$

where x_1 is the possible grid matching point to x that is closest to x'.

In our implementation we construct our topological descriptors using the rank of the 0-homology group. Recall that the 0-homology corresponds to the connected components. For example, the rank of $\frac{H_k(E_{i-})}{ker\sigma_{i-,i+*}}$ is equal to the number of connected components in E_{i+} that have a non-empty intersection with E_{i-}. In our implementation we assume the Lipschitz deformation model defined in Equation 3. All the parameters required for our method are outlined in Tables 1 and 2. Note that the image model parameters are the only pieces of information required about the image. The algorithmic parameters (i.e. Table 2) on the other hand represent parameters in our algorithm and do not affect the validity of the approach.

6. Results

In this section, we describe our performance on a dataset constructed from standard benchmark images. We analyze the performance of our descriptor by computing its precision and recall as in [11]. To determine a ground truth, two regions are said to match if their overlap ratio is greater than a specified threshold τ. Precision and recall are defined as:

$$\text{Precision} = \frac{\#\text{ of correct matches by algorithm}}{\#\text{ of total detections by algorithm}} \tag{66}$$

[1] http://people.engr.ncsu.edu/ejlobato/Research/2011/FeatureMatching/

Parameter	Description
K_p	Perturbation bound. A value of 5 is used.
K_d	Deformation constant. A value of 0.1 is used unless otherwise specified.

Table 1. Image Model Parameters

and

$$\text{Recall} = \frac{\#\text{ of correct matches by algorithm}}{\#\text{ of total true matches}}. \tag{67}$$

In contrast to other descriptors T-ARG does not employ a distance function to compare feature vectors. Rather, either a feature satisfies the conditions of Theorems 2 and 3 in which case it matches or it does not satisfy those conditions in which case it does not match. However, as described in Section 5, we construct bounds for the overlap given our choice of parameters. By thresholding our estimated bounds on the overlap using a parameter τ_{top}, we can get a precision/recall curve parameterized by the value of this parameter as other descriptors do with their appropriately defined distance function. The time required to construct a descriptor for a single region is approximately 20s on a 2.2 Dual-Core i7 CPU with 8 GB of memory. For our method, we set the parameters as described in Tables 1 and 2.

6.1. General Deformation Images

In order to analyze the performance of our approach in the presence of deformations we construct two synthetic datasets: *Graffiti* and *Boat*. The *Graffiti* images are of size 800×640 and the *Boat* images are 850×680. We consider a controlled perturbation using the function $f : \Omega \to \Omega$ given by:

$$f^{-1}\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} z_1 + 0.5\, c\, \cos(0.02 z_2) \\ z_2 + 0.5\, c\, \cos(0.02 z_1) \end{pmatrix}, \tag{68}$$

where the factor c specifies a maximum deformation factor (e.g. $c = 10$ indicates a maximum deformation of 10%). We also study the effect of noise by adding a random uniform perturbation to the images with magnitude equal to 10. Examples from our dataset are illustrated in Fig. 1 and 4. A subset of regions from the images to be matched are chosen to test. We select circular regions of radius 30 from both images centered around points from a uniform grid of points spaced 12 units from each other while removing the points that are 60 units from the boundary of the image. This gives around 2500 regions for each image in the datasets.

Fig. 5 shows the comparison of our approach to SIFT and GLOH for a ground truth threshold of $\tau = 0.5$. The first row corresponds to comparing the base images from the *Graffiti* and *Boat* datasets against an image with 30% deformation and no noise. The second row corresponds to the base image being matched to an image with 30% deformation and with uniform

Parameter	Description
ρ	Radius of morphological operator for pre-processing. Value set to 2.
s_0	Spacing for grid Λ_0. Value set to 6.
s_1	Spacing for grid Λ_1. Value set to 2.
$(r_{0,\lambda}, a_\lambda, b_\lambda)$	Radii and constant used to define the topological rank descriptors. We use all combination such that $r_{0,\lambda} \in \{2, 3, 4, 6, 9, 14, 19, 24, 30\}$, and $[a_\lambda, b_\lambda]$ corresponds to any of the intervals obtained from partitioning the range $[0, 255]$ into 4, 8 and 16 evenly spaced intervals.
R_0	Radius of neighborhood around x_0 used to construct graph $\mathcal{G}_0$ in Solution 1. A value of 30 is used.

Table 2. Algorithmic Parameters

noise in the interval $[-10, 10]$ added to each pixel. The final row corresponds to the base image being matched to an image constructed with 40% deformation and no noise. As expected from our method, we obtain a very high recall rate since our derivations attempt to avoid false-negatives. Observe that these results are obtained using an estimated deformation value of 10% (i.e. $K_d = 10$) which is far less than the actual maximal deformation for the images. Note that adding uniform random noise to an image (as shown in the middle row) and increasing the deformation (as shown in the bottom row) have little effect on the performance of our approach.

Fig. 6 illustrates similar results for a ground truth threshold of $\tau = 0.2$. In terms of the overlap ratio, this corresponds to treating pairs with less actual overlap as potential matches. In this case, the performance of our algorithm improves even though the performance of the SIFT and GLOH descriptors decreases. Finally, Fig. 7 illustrates the dependency of our approach on the choice of parameter K_d for the images with 30% deformation. We choose a ground truth threshold of $\tau = 0.5$ and an overlap threshold for our algorithm of $\tau_{top} = \tau/3$. The plot illustrates the precision and recall for our algorithm against SIFT and GLOH as we change the value of K_d from 0.05 to 0.20.

6.2. Homography Images

Next, we analyze the performance of our approach by constructing a new dataset, called the *Homography Graffiti Dataset*, by applying a homography to the left image of Fig. 1 corresponding to a change in viewing angle by $10°, 30°, 50°$, and $30°$ with uniform noise in the interval $[-10, 10]$ added to each pixel.

In order to give SIFT and GLOH an advantage we select 200 regions by using the Harris Affine Region Detector with a threshold of $80,000$ [11]. We determine the overlap estimate between two regions by first normalizing each detected region into a circular region of radius 30 and then apply our algorithm. Fig. 8 shows the comparison for a ground truth threshold of $\tau = 0.5$. Notice that regardless of choosing regions specifically preferred by SIFT and GLOH upon which to compare the performance of our matching approach on, our method still outperforms the traditional methods.

7. Conclusion

In this paper, we introduced T-ARG, a new local photometric descriptor that can effectively perform deformation invariant image matching. T-ARG is a robust topological descriptor backed by a formal mathematical framework. We applied T-ARG to a set of standard benchmark images with applied deformations and perturbations and demonstrated that T-ARG significantly outperforms traditional descriptors.

The utility of our approach is that it generalizes in a straightforward manner the comparison of multidimensional datasets undergoing bounded deformations. There is a deep relationship between the bounds on the rank of homology presented in Section 3 and the theory of Persistent Topology [14]. In the future we look forward to being able to employ the computational tools available to compute persistence to speed up our implementation and help build an even more powerful topological descriptor. Other potential extensions of this work include: (1) its generalization to account for larger lighting variations

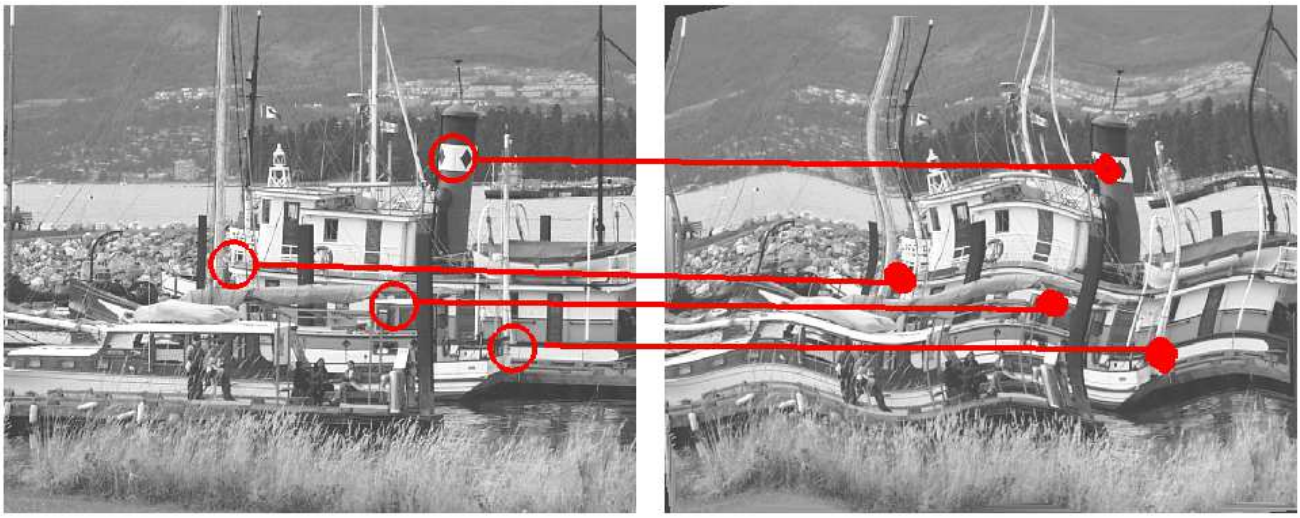

Figure 4. Point Matching Samples: The base image from the *Boat* dataset (left) with corresponding neighborhoods $B(x, R_0)$. Corresponding matches found on an image showing a 30% deformation, i.e. $c = 30$ in Equation 68 (right).

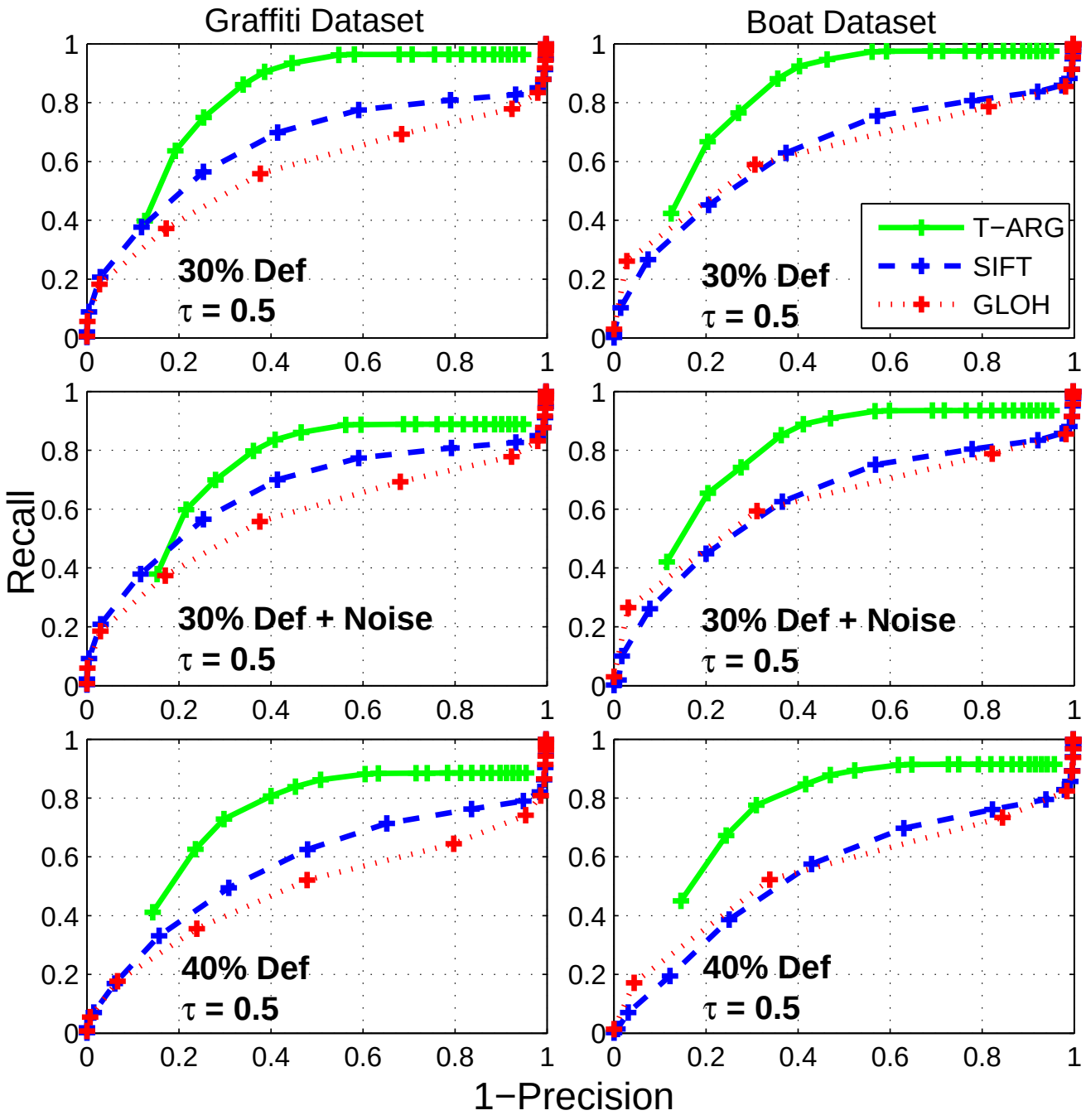

Figure 5. Precision and Recall ($\tau = 0.5$): Results for the *Graffiti* (left) and *Boat* (right) datasets. Results for image with 30% deformation (top), 30% deformation and uniform random noise in the interval $[-10, 10]$ (middle), and 40% deformation (bottom).

present in natural images, and (2) the development of a region detector to determine proper choice of scale for regions of interest.

References

[1] S. Belongie, J. Malik, and J. Puzicha. Shape Matching and Object Recognition Using Shape Contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522, 2002. 2

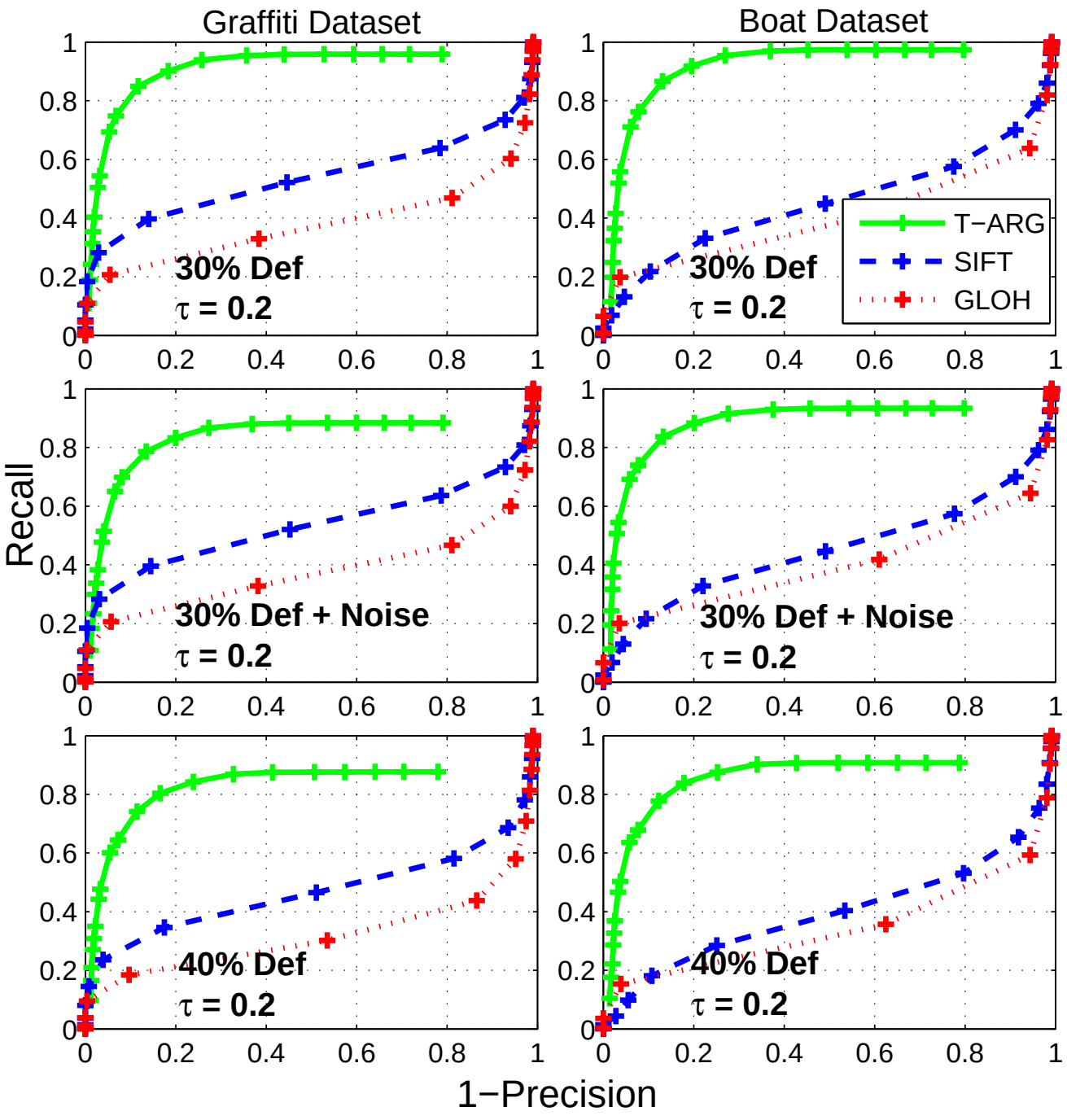

Figure 6. Precision and Recall ($\tau = 0.2$): Results for the *Graffiti* (left) and *Boat* (right) datasets. Results for image with 30% deformation (top), 30% deformation and uniform random noise in the interval $[-10, 10]$ (middle), and 40% deformation (bottom).

[2] W. Freeman and E. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906, 2002. 2

[3] A. Hatcher. *Algebraic Topology*. Cambridge University, 2002. 3, 5

[4] A. Johnson and M. Hebert. Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5):433–449, 2002. 2

[5] S. Lang. *Algebra*, volume 211. Springer, 2002. 4

[6] H. Ling and D. Jacobs. Deformation Invariant Image Matching. In *International Conference on Computer Vision 2005*, pages 1466–1473, 2005. 2

[7] E. Lobaton, R. Vasudevan, R. Bajcsy, and R. Alterovitz. Local Occlusion Detection under Deformations Using Topological Invariants. In *European Conference on Computer Vision 2010*, pages 101–114, 2010. 2

[8] D. Lowe. Object Recognition from Local Scale-Invariant Features. In *International Conference on Computer Vision 1999*, pages 1150–1157, 1999. 1

[9] D. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. 2

[10] K. Mikolajczyk and C. Schmid. Indexing Based on Scale Invariant Interest Points. In *International Conference on Computer Vision 2001*, pages 525–531, 2002. 1

[11] K. Mikolajczyk and C. Schmid. A Performance Evaluation of Local Descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005. 2, 10, 12

[12] F. Schaffalitzky and A. Zisserman. Multi-View Matching for Unordered Image Sets. *European Conference on Computer Vision 2002*, pages 414–431, 2002. 2

[13] T. Tuytelaars and L. Van Gool. Matching Widely Separated Views Based on Affine Invariant Regions. *International Journal of Computer Vision*, 59(1):61–85, 2004. 1

[14] A. Zomorodian and G. Carlsson. Computing Persistent Homology. *Discrete and Computational Geometry*, 33(2):249–274, 2005. 12

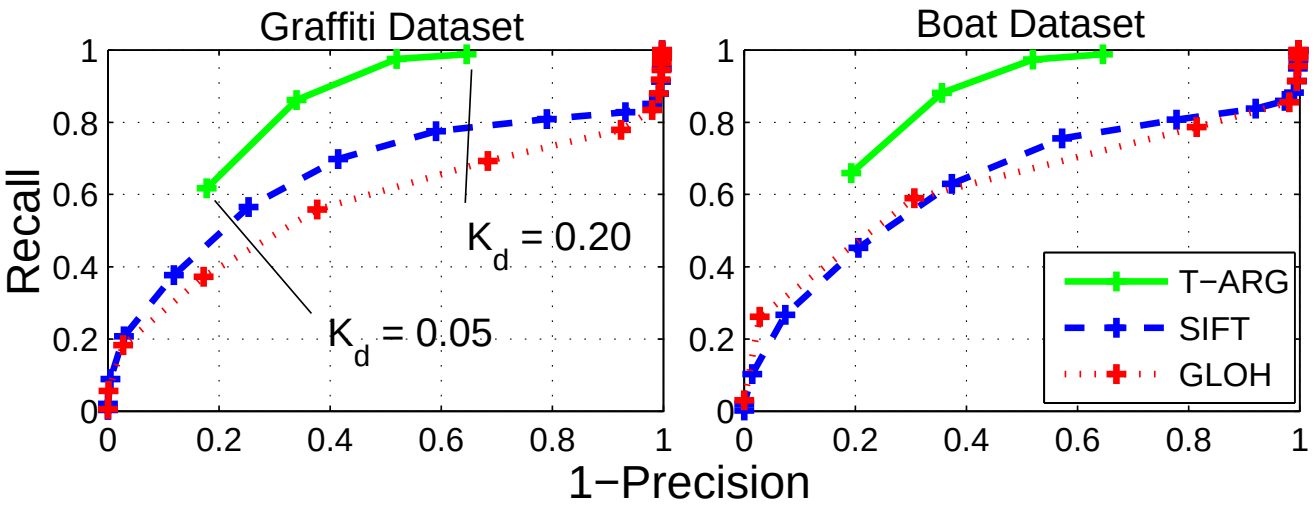

Figure 7. Dependency on deformation bound K_d. Results for *Graffiti* (left) and *Boat* (right) datasets as we change K_d from 0.05 to 0.20. The ground truth overlap threshold is set to $\tau = 0.5$. The overlap threshold for our algorithm is set to $\tau_{top} = \tau/3$.

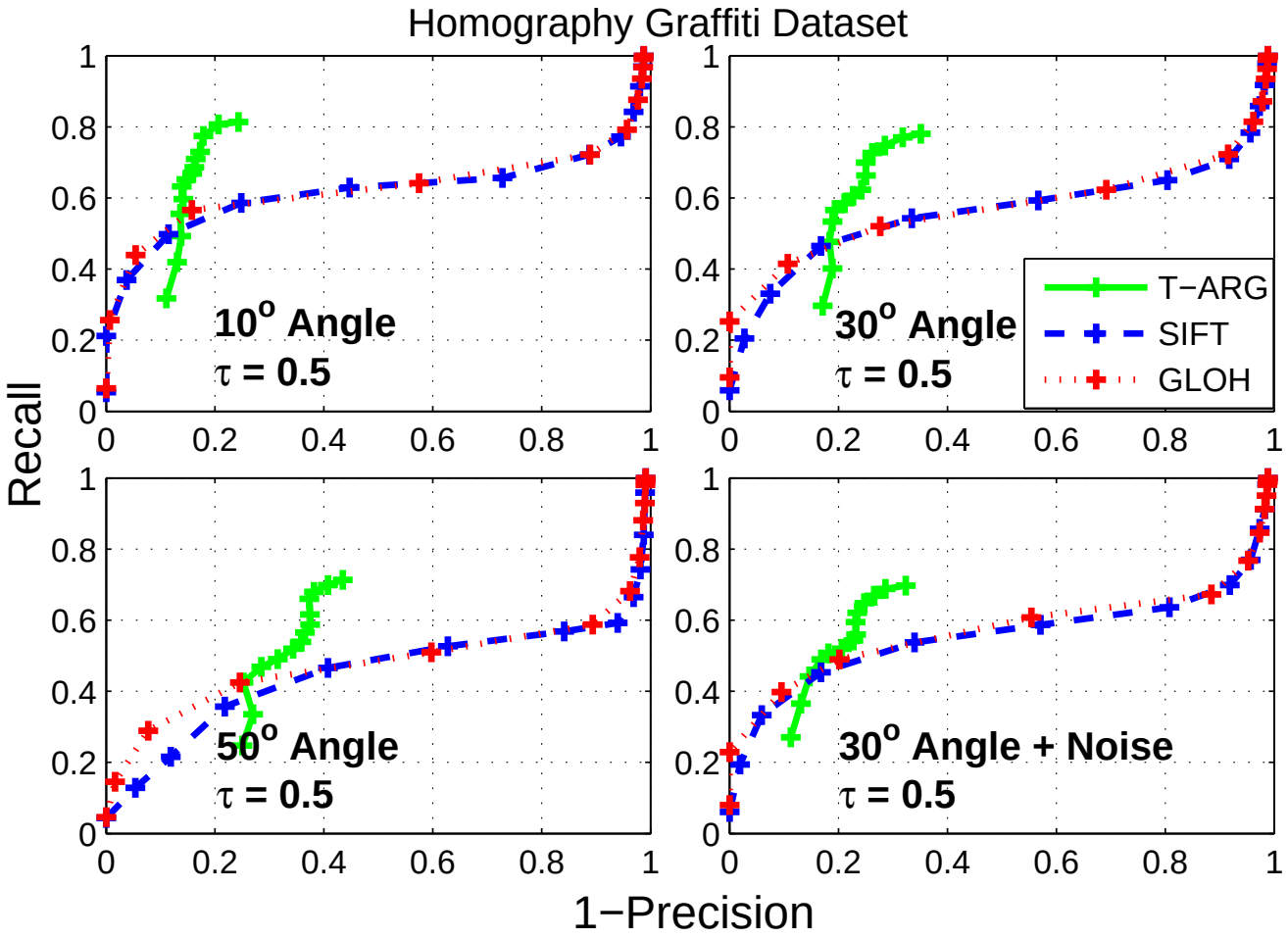

Figure 8. Precision and Recall ($\tau = 0.5$): Results for the *Homography Graffiti Dataset*. Results for image with 10° change in viewing angle (top left), 30° change in viewing angle (top right), 50° change in viewing angle (bottom left), and 30° change in viewing angle and uniform random noise in the interval $[-10, 10]$ (bottom right).