On the Set of Images Modulo Viewpoint and Contrast Changes

Ganesh Sundaramoorthi* Peter Petersen† Stefano Soatto‡

May 16, 2012

Abstract

We consider regions of an image that exhibit smooth variation, and pose the question of characterizing the "essence" of these regions that matters for visual recognition. Ideally, this would be a function of the image that does not depend on viewpoint and illumination (an "invariant"), and yet is "discriminative." In this manuscript, we show that such an invariant exists. That is, one can compute deterministic functions of the image that contain all the "information" present in the original image, except for the effects of viewpoint and illumination, when the underlying three-dimensional shape of the scene is unknown. We also show that such statistics are supported on a "thin" (one-dimensional) subset of the image domain, and thus the "information" in an image that is relevant for recognition is sparse. Yet, from this thin set one can reconstruct an image that is equivalent to the original up to a domain diffeomorphism and a contrast transformation.

Preamble

In this manuscript we characterize the quotient of positive-valued Morse functions of the real plane under diffeomorphic deformations of the domain and monotonic continuous transformations of the range. The motivation comes from the desire to characterize functions of a grayscale image of an unknown scene that are invariant to changes of viewpoint and illumination. Under the assumptions of Lambertian reflection, changes of ambient illumination away from cast shadows and vignetting effects can be characterized by monotonic continuous range transformation, also known as contrast functions. Under the same assumptions, changes of viewpoint away from occluding boundaries can be characterized by Epipolar deformations of the image domain, that are an infinite-dimensional subset of the group of plane diffeomorphisms implicitly defined by the Epipolar constraint [12]. This subset, however, is not a group, and its closure is the entire group of diffeomorphisms. Thus, in order to define a function that is invariant to viewpoint and contrast, in the absence of knowledge about the underlying three-dimensional shape of the scene, we characterize the quotient with respect to domain diffeomorphisms and range contrast transformations. It is well known that any function of an image that is invariant to viewpoint is also invariant to (three-dimensional, 3-D) shape, in the sense that images produced from scenes that are deformed versions of each other are lumped into the same equivalence class, so long as deformations do not produce self-occlusions. As already pointed out in [21], this does not mean that one cannot recognize scenes that have different 3-D shape; it means that one cannot do so by means of comparing invariants. Instead, shape has to be (implicitly or explicitly) inferred as part of the recognition process, or marginalized if priors on shape were available. In this manuscript we deal with viewpoint changes in the absence of singular perturbations due to occlusions or cast shadows. We also assume that the images have infinite resolution. Both assumptions are unrealistic in practice; however, the analysis of this simplified case is relevant to understanding the sparse nature of visual information (as

*Department of Computer Science, KAUST
†Department of Mathematics, UCLA
‡Department of Computer Science, UCLA

we will see, the Attributed Reeb Tree is supported on a zero-measure subset of the image domain). Lifting these assumptions is clearly important for engineering applications and is discussed at length in [20].

1 Introduction: Image Representations for Recognition

Visual recognition is difficult in part because of the large variability that images of a particular object exhibit depending on *extrinsic factors* such as vantage point, illumination, occlusions and other imaging artifacts. The problem is only exacerbated when one considers object categories subject to considerable *intrinsic variability.*

Attempts to "learn away" such variability and to tease out intrinsic and extrinsic factors result in explosive growth of the training requirement, so it is common to attempt to factor out as many of these "nuisances" as possible in a "pre-processing" phase. Ideally, one would want a representation that is *invariant to nuisance factors,* intrinsic or extrinsic[1] and that is *"sufficient"* for the task at hand. Typical nuisances in recognition are (a) viewpoint, (b) illumination, (c) visibility artifacts such as occlusions and cast shadows, (d) quantization, noise and other unmodeled phenomena.[2] The latter two are "non-invertible nuisances", in the sense that they cannot be "undone" in a pre-processing stage: For instance, whether a region of the scene occludes another cannot be determined from an image alone, but can be ascertained as part of the matching process [2]. Such non-invertible nuisances are beyond the scope of this paper, that focuses on the former two: *Can one devise image representations that are invariant to both viewpoint and illumination, away from visibility artifacts*[3] *such as occlusions and cast shadows?*

Viewpoint? Contrast? Both? . . .

The answer to the question above is trivially "yes" as any constant function of the image meets the requirement. More interesting is whether there exists an invariant which is non-trivial, and even more interesting is whether such an invariant is "sufficient," in the sense of containing all and only the "information" that the image contains for the purpose of the task. For the case of viewpoint, although earlier literature [4] suggested that general-case view-invariants do not exist,[4] it has been shown that it is always possible to construct non-trivial viewpoint invariant functions of images for Lambertian objects of any shape [21]. For instance, a (properly weighted) local histogram of the intensity values can be shown to be viewpoint invariant. For the case of illumination, it has been shown [6] that general-case (global) illumination invariants do not exist, even for Lambertian objects. However, there is a considerable body of literature [1, 15, 3][5] dealing with more restricted illumination models that induce monotonic continuous transformations of the image intensities, a.k.a. *contrast transformations.* For instance, [1] show that the geometry of the level curves (the iso-contours of the image), is contrast invariant, and therefore so is its dual, the gradient direction.[6]

But even if we restrict our attention to this more constrained illumination model, one can easily see that *what is invariant to viewpoint is not invariant to illumination, and vice-versa.* So it seems hopeless that we would be able to find *anything* that is invariant to both. However, we will show that under certain conditions (i) viewpoint-illumination invariants do exist; (ii) they are a "thin set" i.e., they are supported on a zero-measure subset of the image domain; finally, despite being thin, (iii) these invariants are sufficient, in the sense that they are equivalent to the original data for any task that requires invariance to viewpoint and contrast.

It is intuitive that discontinuities (edges) and other salient intensity profiles such as blobs and ridges are important, although exactly how important they are for a given recognition task has never been elucidated analytically.*But what about regions with smooth variation?* These would include shaded regions (Fig. 1)

[1]What constitutes a nuisance depends on the task at hand; for instance, sometimes viewpoint is a nuisance, other times it is not, as in discriminating "6" from "9".
[2]Note that we intend (a) and (b) to be absent of visibility artifacts, that are considered separately in (c).
[3]The case of visibility and quantization is addressed in [19].
[4]The results of [4] refer to functions of perspective projection of point ensembles, rather than images.
[5]Note that the image representation presented in these papers are not viewpoint invariant.
[6]This fact is exploited by many local representations used for recognition, such as SIFT [11].

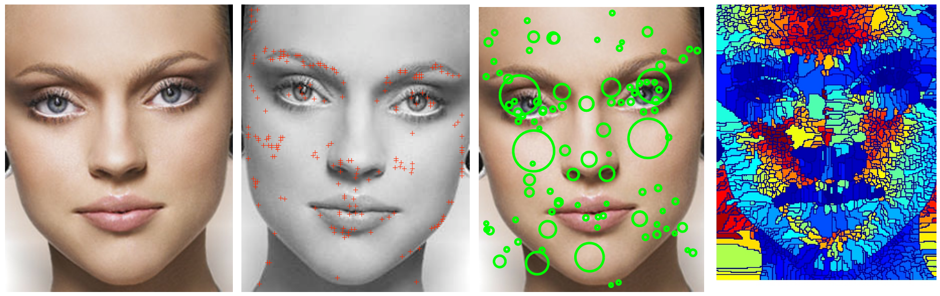

Figure 1: *Regions of an image that exhibit smooth texture gradient are not picked up by local feature detectors (Harris-affine, SIFT), and are over-segmented by most image segmentation algorithms. How do we "capture" the essence of these regions that matters for recognizing an object regardless of its viewpoint and illumination?*

as well as texture gradients at scales significantly larger than that of the local detectors employed for the structures just described. Feature selectors would not fire at these regions, and segmentation or super-pixel algorithms would over-segment them placing spurious boundaries that change under small perturbations. So, *how can one capture the "information" that smooth variations contain for the purpose of recognition?* We articulate our contribution in a series of steps:

1. We assume that some image statistic (intensity, for simplicity, but could be any other region statistic) is smooth, and model the image as a square-integrable function extended without loss of generality to the entire real plane or - for convenience - to the sphere $\mathbb{S}^2$.

2. Again without loss of generality, we approximate the extended image with a Morse function.

3. We introduce the Attributed Reeb Tree (*ART*) [7], a deterministic construction that is uniquely determined from an image and is defined on a zero-measure subset of the image domain.

4. We show that computing viewpoint-invariant image statistics from the image itself (without knowledge of the underlying 3-D scene) is equivalent to computing image statistics that are invariant to the *entire* set of diffeomorphisms of the domain of the image.

5. We show that the *ART* is invariant to domain diffeomorphisms and contrast transformations, and therefore a viewpoint-illumination invariant.[8]

6. Finally, we show that the *ART* is a not just an invariant, but it is *the maximal invariant*, and it is sufficient in the sense that it is equivalent to the original image up to a domain diffeomorphism and contrast change.[9]

The complexity of the *ART*, measured for instance by its coding length with respect to a code, or by its entropy with respect to a prior, reflects the notion of "visual information" advocated by J. Gibson [8], and formalized in [19].

[7]A construction similar to ART has been proposed in [17] for filtering, segmentation, and information retrieval. However, the contribution of the current paper is not to merely to introduce this tree representation of the image, but to show that the ART is a viewpoint/contrast sufficient statistic (see 4-8 above), which is not considered in [17]. Also, we should mention that the ART of an image essentially characterizes the connected components of the level sets of an image treated as a function, and that a few of these connected components are considered in [13]. However, such "extremal regions" are not not viewpoint invariant; indeed, these regions co-vary with changes of viewpoint.

[8]Note that in [10], invariants to the constrained case of affine domain changes and contrast changes are considered. Our paper considers the *general* case of viewpoint changes.

[9]Note that this does not necessarily mean that a viewpoint-illumination invariant is a unique signature for an object. As [21] have pointed out, different objects that are diffeomorphically equivalent in 3-D (i.e. they have equivalent albedo profiles) yield identical viewpoint-invariant statistics. Discriminating objects that differ only by their shape can be done, but *not* by comparing viewpoint-invariant statistics, as shown in [21].

Clearly this is only a piece of the puzzle. It would be simplistic to argue that our key assumption of the image being a Morse function is made without loss of generality (Morse functions are dense in $\mathbb{C}^2$, which is dense in $\mathbb{L}^2$, and therefore they can approximate any discontinuous, square-integrable function to within an arbitrarily small error). Co-dimension one extrema (ridges, valleys) and discontinuities (edges) are qualitatively different than regions with smooth variation and should be treated as such, rather than generically approximated by "elongated blobs." This is beyond our scope in this paper, where we restrict our analysis to regions of images away from such structures. Our goal here is to show that viewpoint-illumination invariants exist under a precise set of conditions, and to provide a proof-of-concept construction. Yet it is interesting to notice that some of the most recent systems for face recognition [18] and shape coding [**?**] use a representation closely related to the *ART*. This paper's contribution is theoretical; however, we believe the theory will be useful in designing better visual recognition systems. For example, many of the ideas presented in this paper have guided the design of an end-to-end visual recognition system [9] (`http://www.youtube.com/watch?v=cMv-McHw660`).

2 Image Invariants

2.1 Invariance to Viewpoint and Illumination

Let $\mathcal{S}$ denote the set of closed, compact, smooth surfaces without boundary. The class $\mathcal{S}$ is an approximation of the set of bounding surfaces of objects in the real world. We denote by $\rho_S : S \to \mathbb{R}^+$, $\rho_S \in \mathcal{A}$ a function representing the albedo of $S \in \mathcal{S}$. Our model for the image formation process is the following. Let $\Omega \subset \mathbb{R}^2$ denote the imaging plane. Given a *viewpoint* $g \in SE(3)$ (an element of the special Euclidean group of rotations and translations in space) and an illumination (contrast) $h \in \mathcal{H}$ which is a monotonic continuous function $h : \mathbb{R}^+ \to \mathbb{R}^+$, we denote the process of image formation as a function $F : \mathcal{S} \times \mathcal{A} \times SE(3) \times \mathcal{H} \to \mathcal{I}$ where $\mathcal{I} = \{I : \Omega \to \mathbb{R}^+\}$ is the space of images:

$$I = F(S, \rho_S; g, h).$$

More specifically, the value I of an image at the pixel location $x \in \mathbb{R}^2$ is given by a contrast-transformed version of the albedo $h \circ \rho_S$ at a point in space with coordinates $X \in \mathbb{R}^3$, $I(x) = h \circ \rho_S(X)$, and the spatial coordinates X and the image coordinates x are related by a change of Euclidean coordinates to place the point in the camera reference frame, followed by an ideal central projection $\pi : \mathbb{R}^3 \to \mathbb{R}^2; X \mapsto (X_1/X_3,\ X_2/X_3)$, so we have that $x = \pi(g(X))$. We now define an invariant to viewpoint and contrast:

Definition 1. *Let $\mathcal{V}$ be a set. A functional $\mu : Range(F) \subset \mathcal{I} \to \mathcal{V}$ is* **invariant** *to the space $SE(3) \times \mathcal{H}$ (viewpoint and contrast) provided that for each $S \in \mathcal{S}$ and $\rho_S \in \mathcal{A}$ we have that*

$$\mu(F(S, \rho_S, g, h)) = \mu(F(S, \rho_S, g', h')), \text{ for all } g, g' \in SE(3), \text{ and } h, h' \in \mathcal{H}.$$

The set $\mathcal{V}$ is called the set of invariants.

Definition 2. *A* **non-trivial invariant** *$\mu : Range(F) \subset \mathcal{I} \to \mathcal{V}$ is an invariant such that there exists $S, S' \in \mathcal{S}$ and $\rho_S, \rho_{S'} \in \mathcal{S}$ so that $\mu(F(S, \rho_S, \cdot, \cdot)) \neq \mu(F(S', \rho_{S'}, \cdot, \cdot))$.*

Definition 3. *A* **maximal invariant** *μ is a (non-trivial) invariant such that $\mu(F(S, \rho_S, \cdot, \cdot)) \neq \mu(F(S', \rho_{S'}, \cdot, \cdot))$ if $F(S, \rho_S, g, h) \neq F(S', \rho'_S, g', h')$ for all $g, g' \in SE(3)$, $h, h' \in \mathcal{H}$ and $S, S' \in \mathcal{S}$.*

Remark 1. *A maximal invariant is such that two images that are formed from different scenes do not have the same invariant representation. Characterizing the maximal invariant is important because* any other invariant *is a function of it.* `[CHECK: IS IT THE OTHER WAY AROUND? IT IS A FUNCTION OF ALL OTHER INVARIANTS?]`

Remark 2. *It is important to note that μ is a* functional defined on the set of two-dimensional images*. Because there are infinitely many surfaces $S \in \mathcal{S}$ that can generate a given image $I \in Range(F)$, it is implicit in the definition above that μ also be invariant to* all possible surfaces *that generate the image I.*

Note that by the definition, the invariant is a property of the object $S \subset \mathbb{R}^3$. It is unrealistic to expect the existence of a non-trivial invariant for the entire group $SE(3)$ since for large parallax (the translational component of $g \in SE(3)$) in general parts of the surface S will be occluded. Therefore, in order to obtain non-trivial invariants, we must take into account occlusions in the definition, which needs a discussion of image generation and visibility, which we do next.

2.2 Visibility

Given a viewpoint $g = (R, T) \in SE(3)$ ($R \in SO(3)$, $T \in \mathbb{R}^3$) and an object $S \in \mathcal{S}$, the pinhole is at the origin in $\mathbb{R}^3$, the imaging plane $\Omega' \subset \mathbb{R}^3$ (an embedding of $\Omega \subset \mathbb{R}^2$) is at T and its orientation is determined by R. A point $X \in Range(S)$ is **visible** from viewpoint g and the imaging plane Ω' if the line segment from the origin to the point X intersects Ω' and (the line segment) does not intersect any point in $Range(S)\backslash\{X\}$. A **camera projection** π from a viewpoint g is a map from the visible points of the object S to Ω given by the point of intersection described earlier. Now we may refine our definition of viewpoint/illumination invariance to take into account visibility.

Definition 4. *Let $\mathcal{V}$ be a set. A functional $\mu : Range(F) \subset \mathcal{I} \to \mathcal{V}$ is* **invariant** *to viewpoint/illumination provided that*

$$\mu(F(S, \rho_S, g, h)) = \mu(F(S, \rho_S, g', h')), \text{ for all } h, h' \in \mathcal{H},$$

and for all $S \in \mathcal{S}$, $g, g' \in SE(3)$ such that S is visible from g and g'.

Remark 3. *The definition of non-trivial and maximal invariant are the same as the definitions that do not account for visibility except that "for all $g, g' \in SE(3)$" is replaced by "for all $S \in \mathcal{S}$, $g, g' \in SE(3)$ such that S is visible from g and g'."*

2.3 Viewpoint Induced Image Transformations

Since a viewpoint/illumination invariant is a function defined on images, in order to describe such invariants, one must first describe the transformations between images induced by changes of viewpoint, which is the goal of the present section.

Let us first start by ignoring visibility, which we will address shortly. In an effort to characterize the smallest class of domain transformations induced by a change of viewpoint, we consider the subset of general diffeomorphisms $w : \mathbb{R}^2 \to \mathbb{R}^2; x \mapsto w(x) = [w_x(x),\ w_y(x)]^T$ specified by the assumption of Lambertian reflection and rigidity of the scene. From Lambertian reflection we that, if ρ_S is the diffuse albedo, then $I(x) = \rho(X)$, were $x = \pi(X)$, is related to another image of the same scene, $J(x)$, via $J(x') = \rho(X)$, where $x' = \pi(g(X)) \doteq w(x)$. Under the rigidity assumption $g = (R, T) \in SE(3)$, i.e. $T \in \mathbb{R}^3$ and $R \in SO(3)$ is a rotation matrix; more in general, for an uncalibrated camera [12], $g \in \mathbb{A}(3)$, the affine group in $\mathbb{R}^3$. Away from occlusions, we can represent the 3-D shape of the object as the graph of a function, for instance $X = \bar{x}Z(x)$ for a function $Z : \mathbb{R}^2 \to \mathbb{R}^+$, where the bar indicates the homogeneous coordinatization $\bar{x} = [x_1,\ x_2,\ 1]^T$. Therefore, we have

$$x' \doteq w(x) = \pi(R\bar{x}Z(x) + T),\ x \in \Omega \tag{1}$$

where $x \in \Omega \subset \mathbb{R}^2$ is the (subset of the) domain for which no (self-)occlusions occur. This limits the range of motions (R, T) depending on the shape $Z(\cdot)$, which is unknown.

It can be easily shown that the set of diffeomorphisms of the form (1) is given by

$$\boxed{\mathcal{W} \doteq \{w : \mathbb{R}^2 \to \mathbb{R}^2 \ |\langle \bar{w}(x), \widehat{T}R\bar{x}\rangle = 0, \text{ for some } (R, T) \in \mathbb{A}(3)\}.} \tag{2}$$

The derivation is in Appendix A. The 3×3 matrix $\widehat{T}R$ is a *fundamental matrix* [12] (it is an *essential matrix* when the cameras are calibrated and hence $(R, T) \in SE(3)$).

If $\mathcal{W}$ was a group under composition, then the maximal image invariant to viewpoint/contrast would be the orbit space, $\mathcal{S}/(\mathcal{H} \times \mathcal{W})$. Unfortunately, however, in general $\mathcal{W}$ is not a group.

Theorem 1 (Epipolar diffeomorphisms do not form a group)**.** *Let* $w_1 = w(x|R_1, T_1, Z_1) \in \mathcal{W}$ *and* $w_2 = w(x|R_2, T_2, Z_2) \in \mathcal{W}$*. Then* $w_3 = w_1 \circ w_2$ *is not, in general, an element of* $\mathcal{W}$*.*

The proof is in Appendix A

Remark 4. *Note that if both* $w_1, w_2 \in \mathcal{W}$ *are* known to come from the same scene, *then* $w_1 \circ w_2 \in \mathcal{W}$*. However, because* w_1 *and* w_2 *could be induced by different scenes, the composition is generally not an element of* $\mathcal{W}$*, and therefore an invariant has to quotient out the entire group closure of epipolar domain transformations.*

We now show that the group closure, i.e., the smallest group containing $\mathcal{W}$, under composition is the general set of diffeomorphisms, and this fact will be used in the next section. First, we introduce a restricted subset of $\mathcal{W}$ under which visibility conditions are satisfied:

$$\tilde{\mathcal{W}} \doteq \{w : \Omega \subset \mathbb{R}^2 \to \mathbb{R}^2; x \mapsto w(x|R, T, Z) \mid \exists\, Z'(\cdot) \mid R\bar{x}Z(x) + T = \bar{w}(x)Z'(w(x))\ \forall\, x \in \Omega\}. \tag{3}$$

The following theorem, proved in Appendix A, shows that the closure of $\tilde{\mathcal{W}}$ is the entire set of diffeomorphisms:

Theorem 2. *The group closure (i.e., the smallest group containing* $\tilde{\mathcal{W}}$*) is the entire set of (orientation preserving) diffeomorphisms of the plane.*

3 Maximal Viewpoint/Contrast Invariant

In this section, we are interested in giving a classification of the set of two-dimensional images under the equivalence of *domain diffeomorphism* and *contrast* changes, that is, we classify the set of images in which two images are equivalent if they are related by a domain diffeomorphism and/or contrast change. Note that if one wants to use viewpoint-invariant statistics, computed independently on the template and target images for image matching, and compare such invariants directly, then necessarily one has to quotient out all possible surfaces of the scene that could have generated the image (since we do not know the surface associated to the image from the image alone). By Theorem 2 in the previous section, this entails quotienting out the set of images by the entire group of diffeomorphisms. Thus, classifying the set of images under domain diffeomorphisms/contrast changes classifies the maximal viewpoint/illumination invariant, which we wish to seek. The price for doing this, as is well known (see [21]), is the loss of shape discrimination. The benefit is that, at decision time, one just compares statistics, as opposed to having to solve an optimization problem (to find the epipolar transformation that brings the target image into correspondence with the template), or to compute a complex integral (to marginalize all possible scenes according to their prior probability; it should be noted that the set of shapes is hard to even describe analytically and endow with a metric, let alone a suitable probability measure, and learning a prior on it).

3.1 Morse Functions As Image Approximations

For simplicity, we will represent an image by a function on the *plane*: $f : \mathbb{R}^2 \to \mathbb{R}^+$.

Definition 5 (Morse function)**.** *A Morse function* $f : \mathbb{R}^2 \to \mathbb{R}^+; x \mapsto f(x)$ *is a* C^2 *smooth function such that all critical points are non-degenerate. A critical point is a location* $x \in \mathbb{R}^2$ *where the gradient vanishes,* $\nabla f(x) = 0$*. A non-degenerate critical point is a critical point* x *where the Hessian is non-singular,* $\det(\nabla^2 f(x)) \neq 0$*.*

Remark 5. *Morse functions cannot have ridges, valleys and other critical structures of co-dimension one, although they can approximate them to an arbitrary degree. We will address the relevance of this restriction in Remark 14 in Section 3.4.*

To further simplify matters in our classification of images, we assume that the functions we consider fall in the following class

Definition 6 ($\mathcal{F}$). *A function $f : \mathbb{R}^2 \to \mathbb{R}^+$ is in class $\mathcal{F}$ ($f \in \mathcal{F}$) iff*

1. *f is Morse*
2. *the critical values of f (corresponding to critical points of f) are distinct*
3. *each* level set *(i.e. $L_a(f) = \{x \in \mathbb{R}^2 : f(x) = a\}$ for $a \in \mathbb{R}^+$) of f is* compact,
4. $\lim_{|x|\to+\infty} f(x) > f(y)\, \forall y \in \mathbb{R}^2$ *or* $\lim_{|x|\to+\infty} f(x) < f(y)\, \forall y \in \mathbb{R}^2$,
5. *there exists an $a \in \mathbb{R}^+$ so that $L_a(f)$ is a simple closed contour that encloses all* critical points *of f*

Remark 6. *If $f \in \mathcal{F}$, then we may identify f with a Morse function $\tilde{f} : \mathbb{S}^2 \to \mathbb{R}^+$ defined on the sphere, $\mathbb{S}^2$ via the inverse stereographic projection from the south pole, p. We then extend $\tilde{f}$ to the south pole, $-p$, by defining $\tilde{f}(-p) = \lim_{|x|\to+\infty} f(x)$, which will be either the global minimum or maximum of $\tilde{f}$. From now on in this article, we make this identification and any $f \in \mathcal{F}$ will be represented as a Morse function on $\mathbb{S}^2$ such that its global minimum or maximum is at the south pole.*

Conditions 1 and 2 make the class $\mathcal{F}$ stable *under small perturbations (e.g. noise in images); we will make this notion of stability more precise in Remark 13 in Section 3.4.*

Remark 7. *Images (e.g. the continuum version of digital images) are usually defined on a compact rectangular domain (e.g. $[0,1]\times[0,1]$). We may extend such a Morse function, $g : [0,1]\times[0,1] \to \mathbb{R}^+$ (with minimal distortion), to one that satisfies Condition 3-5 as follows. Let $c \subset [0,1]\times[0,1]$ denote a smooth simple closed curve that is arbitrarily close (say wrt a geometric L^∞ distance) to the boundary $\partial([0,1]\times[0,1])$. Define $b : \mathbb{R} \to \mathbb{R}$ as*

$$b_\epsilon(x) = \begin{cases} \exp\left(-\frac{\epsilon^2}{x^2}\right) & x > 0 \\ x\exp\left(-\frac{1}{x^2}\right) & x < 0. \end{cases}$$

Then the extended function $f : \mathbb{R}^2 \to \mathbb{R}^+$ is

$$f(x) = \begin{cases} g(x)b_\epsilon(dist_c(x)) & x \text{ is inside } c \\ b_\epsilon(-dist_c(x)) & x \text{ is outside } c \end{cases}$$

where $dist_c(x)$ is the distance from x to the curve c.

Now consider the set of surfaces that are the graph of a function in $\mathcal{F}$,

$$\mathcal{S} \doteq \{\{(x, f(x))|x \in \mathbb{S}^2\} \mid f \in \mathcal{F}\}. \tag{4}$$

The set of monotonic continuous functions, also called *contrast functions* in [5], is indicated by

$$\mathcal{H} \doteq \{h \in C^2(\mathbb{R}^+;\mathbb{R}^+) \mid 0 < \frac{dh}{dt} < \infty,\ t \in \mathbb{R}^+\}. \tag{5}$$

Contrast functions form a group under function composition, and therefore each surface in $\mathcal{S}$ that is the graph of a function f forms an orbit (equivalence class) of surfaces that are different from the original one, but related via a contrast change. We indicate this equivalence class by $[f]_\mathcal{H} = \{h \circ f \mid h \in \mathcal{H}\}$. The *topographic map* of a surface is the set of connected components of its level curves, $\mathcal{S}' \doteq \{\{x \mid f(x) = \lambda\}, \lambda \in \mathbb{R}^+\}$; it follows from Proposition 1 and Theorem 1 on page 11 of [5] that the orbit space of surfaces $\mathcal{S}$ modulo $\mathcal{H}$ is given by their topographic map,

$$\mathcal{S}' = \mathcal{S}/\mathcal{H}. \tag{6}$$

In other words, the topographic map is a sufficient statistic of the surface that is invariant to contrast changes. Or, all surfaces that are equivalent up to a contrast change have the same topographic map. Or, *given a topographic map, one can uniquely reconstruct a surface up to a contrast change* [5].

Remark 8. *In the context of image analysis, where the domain of the image is a rectangle (for instance a continuous approximation of the discrete lattice* $D = [0, 640] \times [0, 480] \subset \mathbb{Z}^2$*) and* $f(x)$ *is the intensity value recorded at the pixel in position* $x \in D$*, usually between* 0 *and* 255*, contrast changes in the image are often considered as a first-order approximation of illumination changes in the scene away from visibility artifacts such as cast shadows. Therefore, the topographic map, or dually the gradient direction* $\frac{\nabla f}{\|\nabla f\|}$*, is equivalent to the original image up to contrast changes, and represents a sufficient statistic that is invariant to* h*.*

Now consider the set of domain diffeomorphisms of functions in $\mathcal{F}$:

$$\mathcal{W} \doteq \{w \in C^2(\mathbb{R}^2; \mathbb{R}^2) : \text{ a diffeomorphism}\} \cong \{w \in C^2(\mathbb{S}^2; \mathbb{S}^2) : \text{a diffeomorphism s.t. } w(\sigma) = \sigma, \sigma \text{ is the south pole }\} \quad (7)$$

which is a group under composition, and therefore each surface determined by f generates an orbit $[f]_{\mathcal{W}} = \{f \circ w \mid w \in \mathcal{W}\}$. If we consider the product group of contrast functions and domain diffeomorphisms we have the orbits $[f] = \{h \circ f \circ w \mid h \in \mathcal{H}, w \in \mathcal{W}\}$. **The goal of this manuscript is to characterize** these equivalence classes. In other words, we want to characterize the orbit space

$$\mathcal{S}'' \doteq \mathcal{S}'/\mathcal{W} = \mathcal{S}/\{\mathcal{H} \times \mathcal{W}\} \quad (8)$$

of surfaces that are equivalent up to domain diffeomorphisms and contrast functions.

Remark 9. *In the above it is important to note that the orbit space above is defined algebraically, and that the group* $\mathcal{H} \times \mathcal{W}$ *acts on the set* $\mathcal{S}$*. Therefore, the quotient we seek above is just a set, and we do not seek to characterize the topology of the resulting quotient.*

Remark 10. *As one can check easily, it turns out that the orbit space* $\mathcal{S}/\{\mathcal{H} \times \mathcal{W}\}$ *is the set of* maximal *viewpoint/illumination invariants according to our definition of illumination change (a contrast change). See Definition 3 to recall the definition of maximal invariant.*

Remark 11. *The quotient above – if it is found to be non-trivial – is a sufficient statistic of the image that is invariant to viewpoint and illumination.*

3.2 Reeb Graphs: Towards Viewpoint/Contrast Invariants

We now introduce Reeb graphs [16], and their basic properties. Reeb graphs, as will be apparent in the next sections, will be the basis for the construction of viewpoint/contrast invariants of images.

Definition 7 (Reeb Graph of a Function)**.** *Let* $f : \mathbb{S}^2 \to \mathbb{R}$ *be a continuous function. We define*

$$Reeb(f) = \{[(x, f(x))] : x \in \mathbb{S}^2\}$$

where

$$(y, f(y)) \in [(x, f(x))] \text{ iff } f(x) = f(y) \text{ and there is a continuous path from } x \text{ to } y \text{ in } f^{-1}(f(x)).$$

In other words, the Reeb graph of a function f is the set of connected components of level sets of f (with the additional information of the function value of each level set). We now recall some basic facts about Reeb graphs.

Lemma 1 (Reeb graph is connected)**.** *If* $f : \mathbb{S}^2 \to \mathbb{R}$ *is a function, then* $Reeb(f)$ *is connected.*

Proof. $Reeb(f)$ is the quotient space of $\mathbb{S}^2$ under the equivalence relation defined in Definition 7. Therefore, by definition we have a surjective continuous map $\pi : \mathbb{S}^2 \to Reeb(f)$, and connectedness is preserved under continuous maps. □

Lemma 2 (Reeb Tree)**.** *The Reeb graph of a surface in $\mathcal{S}$ that is the graph of a function f does not contain cycles.*

Proof. Let $\pi : \mathbb{S}^2 \to Reeb(f)$ be the quotient map. We prove that $Reeb(f)$ has no cycles. Assume $Reeb(f)$ has a cycle, i.e., there exists $\gamma : [0,1] \to Reeb(f)$, continuous with $\gamma(0) = \gamma(1)$, and we can assume that γ is one-to-one on $[0,1)$. We may then lift γ to a *continuous path*, $\hat{\gamma} : [0,1] \to \mathbb{S}^2$ that satisfies $\hat{\gamma}(0) = \hat{\gamma}(1)$ and $\pi \circ \hat{\gamma} = \gamma$:

1. If $\hat{\gamma}(0) \neq \hat{\gamma}(1)$, then since $(\hat{\gamma}(0), f(\hat{\gamma}(0))) \in [\hat{\gamma}(1), f(\hat{\gamma}(1))]$, we have that there must exist a continuous path $p : [1,2] \to \mathbb{S}^2$ such that $p(1) = p(0)$ and $f \circ p = f(\hat{\gamma}(0)) = f(\hat{\gamma}(1))$. Then $\tilde{\gamma} : [0,2] \to \mathbb{S}^2$ where
$$\tilde{\gamma}(t) = \begin{cases} \hat{\gamma}(t) & t \leq 1 \\ p(t) & t > 1 \end{cases}$$
satisfies $\tilde{\gamma}(0) = \tilde{\gamma}(2)$.

2. We show that $\hat{\gamma}$ can be chosen so that it is continuous. We may assume that γ passes through the critical points (of f), $\gamma(t_1), \ldots, \gamma(t_N)$ in that order. Thus, we divide the path γ into the sub-paths $\gamma(0) \to \gamma(t_1)$, $\gamma(t_1) \to \gamma(t_2)$, ..., that do not contain critical points in the intervals $(0,t_1), (t_1,t_2), \ldots, (t_N, 1)$. To construct $\hat{\gamma}$ in each interval $[t_i, t_{i+1}]$, we choose a point $x_i \in \pi^{-1}(\gamma((t_i + t_{i+1})/2)) \subset \mathbb{S}^2$. Then $\hat{\gamma}$ in $(t_i, (t_i + t_{i+1})/2)$ is defined as the path solving
$$\dot{y} = \nabla f(y), \quad y(0) = x_i \in \mathbb{S}^2$$
and in $((t_i + t_{i+1})/2, t_{i+1})$ as
$$\dot{y} = -\nabla f(y), \quad y(0) = x_i \in \mathbb{S}^2$$
clearly, these paths are continuous and we therefore have that $\hat{\gamma}$ is continuous, and $\pi(\hat{\gamma}) = \gamma$.

Now that we have a continuous loop $\hat{\gamma} : [0,1] \to \mathbb{S}^2$ we may contract $\hat{\gamma}$ to a point via a retraction, $F : [0,1] \times [0,1] \to \mathbb{S}^2$, such that $F(0,t) = \hat{\gamma}(t)$ and $F(1,t) = \widehat{\gamma(0)}$. Then $\pi \circ F$ is a retraction of γ to $\gamma(0)$, which is impossible unless $\gamma = \gamma(0)$, in which case we did not have a loop. A retraction of a loop (one-to-one path with endpoints the same) in $Reeb(f)$ is impossible. □

3.3 Attributed Reeb Trees (ART)

We now introduce the definition of Attributed Reeb Trees (ART), which we will show in the next section is the maximal invariant to viewpoint/contrast. To introduce the definition of ART, we must start with a series of intermediate definitions.

Definition 8 (Attributed Graph)**.** *Let $G = (V, E)$ be a graph (V is the vertex set and E is the edge set), and L be a set (called the* label set*). Let $a : V \to L$ be a function (called the* attribute *function). We define the* attributed graph *as $AG = (V, E, L, a)$.*

Definition 9 (Attributed Reeb Tree of a Function)**.** *Let $f \in \mathcal{F}$. Let V be the set of critical points of f. Define E to be*

$$E = \{(v_i, v_j) : i \neq j, \exists \text{ a continuous map } \gamma : [0,1] \to Reeb(f) \text{ such that } \gamma(0) = [(v_i, f(v_i))], \gamma(1) = [(v_j, f(v_j))] \text{ and } \gamma(t) \neq [(v, f(v))] \text{ for all } v \in V \text{ and all } t \in (0,1)\}. \quad (9)$$

Let $L = \mathbb{R}^+$, and
$$a(v) = f(v)$$
Note that the south pole $v_{sp} \in \mathbb{S}^2$, is a critical point, and we include that in our definition. We define
$$ART(f) := (V, E, L, a, v_{sp}).$$

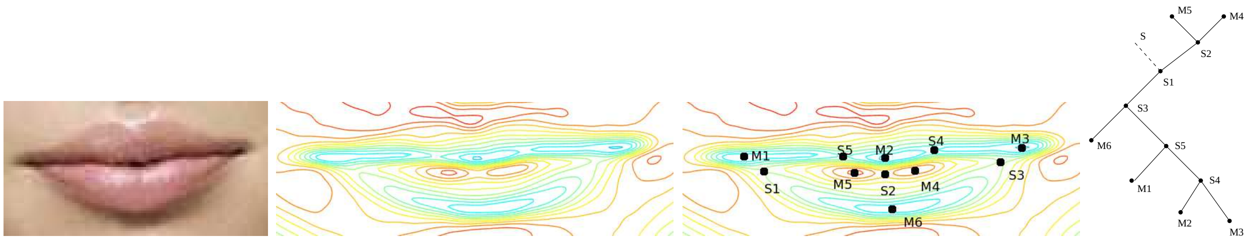

Figure 2: *The lip region of Fig. 1, its level lines, the level lines marked with extrema, and a graphical depiction of the ART (note that the height of the vertex is proportional to the attribute value).*

Note that the above definition encodes the type of critical point of each vertex $v \in V$:

Definition 10 (Index of a Vertex of an Attributed Tree)**.** *Let $T = (V, E, \mathbb{R}^+, a)$ be an attributed tree, we define the map $ind : V \to \{0, 1, 2\}$ as follows:*

1. *$ind(v) = 2$ if $a(v) < a(v')$ for any v' such that $(v, v') \in E$*
2. *$ind(v) = 0$ if $a(v) > a(v')$ for any v' such that $(v, v') \in E$*
3. *$ind(v) = 1$ if the above two conditions are not satisfied.*

Definition 11 (Equivalence of Attributed Trees)**.** *Let $T_1 = (V_1, E_1, \mathbb{R}^+, a_1, v_{sp,1})$ and $T_2 = (V_2, E_2, \mathbb{R}^+, a_2, v_{sp,2})$ be attributed trees. Then we say that T_1 is equivalent to T_2 denoted $T_1 \cong T_2$ if the trees (V_1, E_1) and (V_2, E_2) are isomorphic via a graph isomorphism, $\phi : V_1 \to V_2$, and the following properties are satisfied:*

- *if $a_1(v) > a_1(v')$ then $a_2(\phi(v)) > a_2(\phi(v'))$ for all $v, v' \in V_1$*
- *$\phi(v_{sp,1}) = v_{sp,2}$.*

Definition 12 (Degree of a Vertex)**.** *Let $G = (V, E)$ be a graph, and $v \in V$, then the degree of a vertex, $deg(v)$, is the number of edges that contain v.*

Definition 13 ($\mathcal{T}$, a Collection of Attributed Trees)**.** *Let $\mathcal{T}'$ denote the subset of attributed trees $(V, E, \mathbb{R}^+, a, v_{sp})$ satisfying the following properties:*

1. *(V, E) is a tree*
2. *If $v \in V$ and $ind(v) \neq 1$ then $deg(v) = 1$*
3. *If $v \in V$ and $ind(v) = 1$, then $deg(v) = 3$*
4. *$n_0 - n_1 + n_2 = 2$ where n_0, n_1 and n_2 are the number of vertices of index 0, 1, and 2.*

We define $\mathcal{T}$ to be the set $\mathcal{T}'$ under the equivalence defined in Definition 11.

Fig. 2 shows an example of constructing an ART from an image (in this case the lip part of the image in Fig. 1). We will show in the next section that $ART(\mathcal{F}) = \mathcal{T}$.

3.4 ART is the Maximal Viewpoint/Contrast Invariant

In this section, we show that $\mathcal{S}'' = \mathcal{T}$. Clearly $ART(f)$ is invariant with respect to domain diffeomorphisms and contrast changes, i.e. $h \circ f \circ w$, since the latter do not change the topology of the level curves. However, it is less immediate to see that the Attributed Reeb tree is a sufficient statistic, or that it is equivalent to the surface that generated it up to a domain diffeomorphism and contrast transformation.

We start by stating a fact from Morse theory [14] that we exploit in our argument:

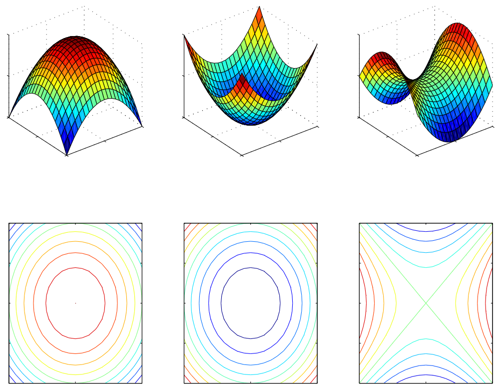

Figure 3: The Morse Lemma states that in a neighborhood of a critical point of a Morse function, the level sets are topologically equivalent to one of the three forms (left to right: maximum, minimum, and saddle critical point neighborhoods).

Lemma 3 (Morse Lemma)**.** *If $f : \mathbb{S}^2 \to \mathbb{R}$ is a Morse function, then for each critical point p_i of f, there is a neighborhood U_i of p_i and a chart $\psi_i : \tilde{U}_i \subset \mathbb{R}^2 \to U_i \subset \mathbb{S}^2$ so that*

$$f(\hat{x},\hat{y}) = f(p_i) + \begin{cases} -(\hat{x}^2+\hat{y}^2) & \text{if } p_i \text{ is a maximum point} \\ \hat{x}^2+\hat{y}^2 & \text{if } p_i \text{ is a minimum point} \\ \hat{x}^2-\hat{y}^2 & \text{if } p_i \text{ is a saddle point} \end{cases}$$

where $(\hat{x},\hat{y}) = \psi_i(x,y)$ and $(x,y) \in \mathbb{S}^2$ are the natural arguments of f.

Figure 3 shows the three canonical forms stated in the previous lemma.

Lemma 4 (Degree of Vertices in ART)**.** *Let $f \in \mathcal{F}$, and $ART(f) = (V, E, L, a, v_{sp})$, then*

1. *if $v \in V$ and $ind(v) \neq 1$, then $deg(v) = 1$*
2. *if $v \in V$ and $ind(v) = 1$, then $deg(v) = 3$.*

Proof. The first assertion (the case when v is a maximum or minimum) follows directly from the Morse Lemma. The second may be proved using the two relations

$$n_{0,2} - n_1 = 2 \quad \text{and} \quad n_{0,2} + n_1 - |E| = 1 \tag{10}$$

where $n_{0,2}$ denotes the number of vertices of degree 0 or 2, n_1 is the number of vertices of degree 1, and $|E|$ is the number of edges. The first is relation is a fact from Morse Theory [14], and the second is simply the relation for trees that $|V| - |E| = 1$. Noting that for any graph,

$$\sum_{v\in V} deg(v) = 2|E| \quad \text{or} \quad n_{0,2} + \sum_{v\in V, ind(v)=1} deg(v) = 2|E|, \tag{11}$$

and combining with (10), we find that

$$\sum_{v\in V, ind(v)=1} deg(v) = 3n_1, \tag{12}$$

but according to the Morse Lemma and the fact that critical points have distinct values (by definition of $\mathcal{F}$), $deg(v) > 2$ if $ind(v) = 1$. These facts and (12) mean that $deg(v) = 3$ if $ind(v) = 1$. ☐

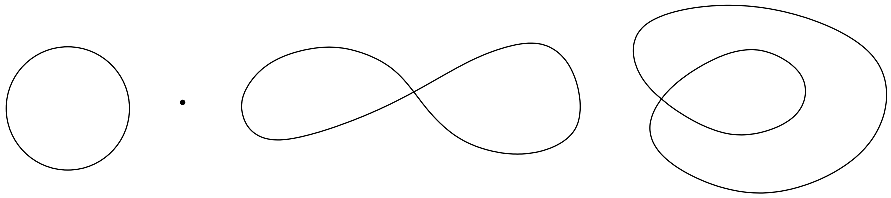

Figure 4: The possible connected components of a level set of a function. Left to right: a regular point's level set, a minimum or maximum point, a Type 1 saddle point, and a Type 2 saddle point level set. Note that the last two are indistinguishable on the sphere, but not on the plane (as in the case of interest).

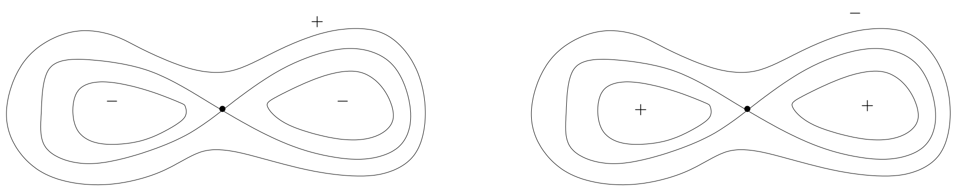

Figure 5: Level sets in a thickening of a Type 1 saddle connected component, $\pi_f^{-1}([x, f(x)])$. The plus/minus indicates that the level sets are above/below the value of the saddle point. An example of this type of saddle point arises from a pair of shorts.

Lemma 5 (Global Topology of Connected Level Sets)**.** *Let $f \in \mathcal{F}$, and $\pi_f : \mathbb{R}^2 \to Reeb(f)$ be the natural quotient map. Then $\pi_f^{-1}([x, f(x)])$ for each $x \in \mathbb{R}^2$ is topologically the same as one of the following:*

Proof. There are three cases: either $x \in \mathbb{R}^2$ is a critical point (saddle or min/max) or a regular point. Note that because we are working with the class $\mathcal{F}$ of functions, $\pi_f^{-1}([x, f(x)])$ is compact, and not other critical point may have the value $f(x)$. By the Morse Lemma, if x is a regular point, then $\pi_f^{-1}([x, f(x)])$ is topologically a circle, and if x is a min/max, then $\pi_f^{-1}([x, f(x)])$ is a point. The only case that remains is the saddle. For x a saddle $\pi_f^{-1}([x, f(x)])$ is compact and must cross at an 'X', there are only two possible topologies for $\pi_f^{-1}([x, f(x)])$, and they are the latter two cases. □

By the previous Lemma and the Morse Lemma, it is easy to see that in thickening around $\pi_f^{-1}([x, f(x)])$ (x a saddle), the level sets are topologically equivalent to the cases in Fig. 5 for Type 1 saddles, and in Fig. 6 for Type 2 saddles.

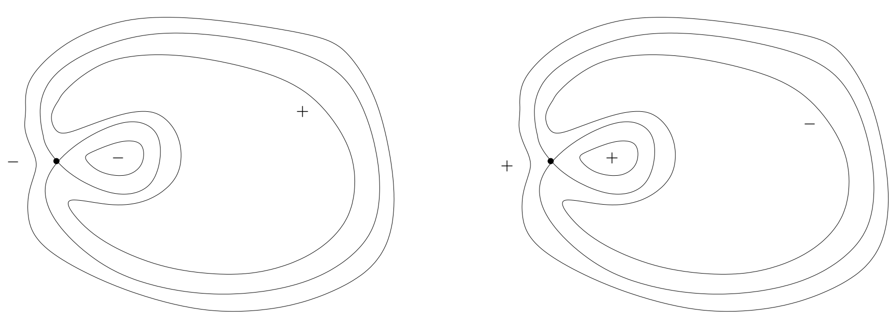

Figure 6: Level sets in a thickening of a Type 2 saddle connected component, $\pi_f^{-1}([x, f(x)])$. The plus/minus indicates that the level sets are above/below the value of the saddle point. An example of this type of saddle arises from a hill with a pit on the side.

Lemma 6. *Let f_1, $f_2 \in \mathcal{F}$ and $ART(f_1) \cong ART(f_2)$. Let ϕ be a graph isomorphism between the trees in $ART(f_1)$ and $ART(f_2)$ satisfying Def. 11. If $v \in V_1$ and $v' \in V_2$ where v is a Type 1 saddle and v' is a Type 2 saddle, then $\phi(v) \neq v'$.*

Proof. We proceed by induction on n, the number of saddles of f_1 (or f_2). If $n = 1$, then the Attributed Reeb Trees must have one of the forms in Fig. 7. Note that v_{sp} is the south pole vertex (of $\mathbb{S}^2$), which is

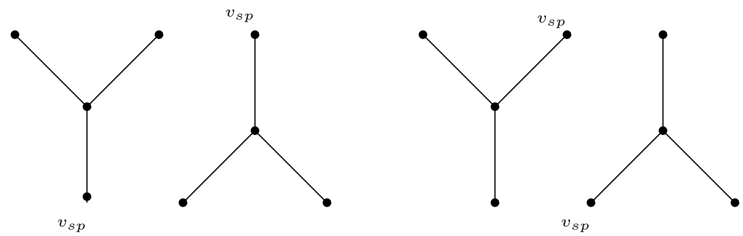

Figure 7: If $n = 1$, then the $ART(f)$ must be equivalent to the Type 1 saddles (left) or the Type 2 saddles (right), and the two types are not equivalent since v_{sp} must be preserved under ϕ.

equivalent to the point at infinity in $\mathbb{R}^2$. Because v_{sp} must be preserved by ϕ (that is, the points at infinity in the domains of f_1 and f_2 must be mapped to each other), a Type 1 saddle (on the left in Fig. 7) may not be mapped to a Type 2 saddle (on the right in Fig. 7).

Next assume that for all f_1', f_2' that have $n-1$ saddles, we have that $\phi'(v) \neq v'$ where $v \in V_1$ and $v' \in V_2$ are different saddle types for any valid graph isomorphism ϕ'. Now let f_1, f_2 have n saddles. Choose a saddle point v_s of f_1 that is adjacent to two vertices that are not saddle points, and let $v_s' = \phi(v_s)$. We claim that v_s and v_s' are saddles of the same type. Indeed, the Attributed Reeb trees around the v_s and v_s' are in Figure 8, where the label S denotes a vertex that is a saddle point and the others denote maxima

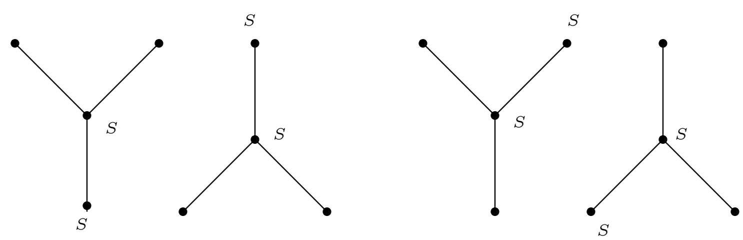

Figure 8: Attributed Reeb trees of Type 1 (left) and Type 2 (right) saddles which are adjacent to two vertices that are not saddles.

or minima. Clearly, ϕ may not map v_s to v_s' if they are of different types. Now we reduce $ART(f_1)$ and $ART(f_2)$ to have trees with $n-1$ saddles by removing the maxima/minima adjacent to v_s and v_s' (and their edges). Note that v_s and v_s' now become a maximum or minimum. The resulting attributed trees have $n-1$ saddles and result from functions f_1' and f_2' that are obtained by coarsening f_1 and f_2 near v_s and v_s' (note that we may also apply Lemma 8 to obtain f_1' and f_2'). Now the restriction of ϕ to $ART(f_1')$ and $ART(f_2')$ is a valid equivalence. But by the inductive hypothesis, ϕ does not map different types of saddles to each other. □

We now move to the core part of our argument:

Lemma 7. *Let f_1, $f_2 \in \mathcal{F}$ be functions that generate two surfaces. Then*

$$ART(f_1) \cong ART(f_2) \Leftrightarrow \ \exists\ h \in \mathcal{H}, w \in \mathcal{W} \text{ such that } f_1 = h \circ f_2 \circ w. \tag{13}$$

Note that the diffeomorphism w and contrast function h are not necessarily unique.

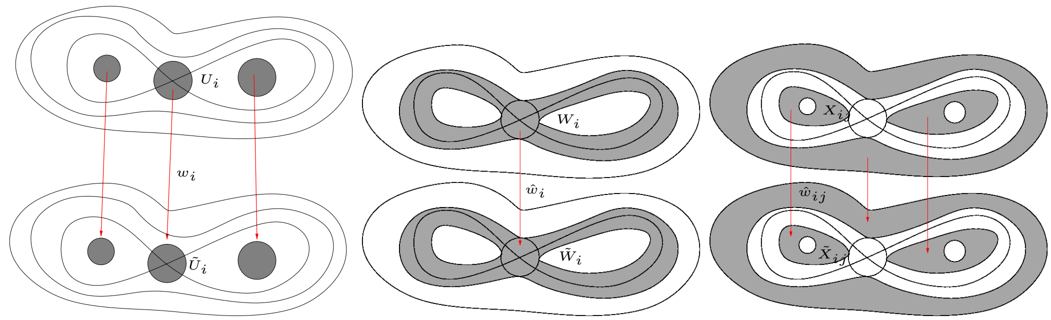

Figure 9: Illustration of Steps 2, 3, and 4, respectively, of the proof of Lemma 7. This figure is the proof of Lemma 7 in pictures: it details the construction of the domain diffeomorphism w such that $f_1 = h \circ f_2 \circ w$ where h is a contrast change and $ART(f_1) \cong ART(f_2)$. The Morse Lemma (Lemma 3) guarantees the existence of a diffeomorphism w_i that maps a neighborhood, U_i of critical point p_i of f_1 to a neighborhood $\tilde{U}_i$ of a corresponding critical point $\tilde{p}_i$ of f_2 and satisfying the desired conditions (left most image). The map w_i can then be extended to a map $\tilde{w}_i$ defined on W_i, the level sets of f_1 that intersect U_i, and range $\tilde{W}_i$, the level sets of f_2 that intersect $\tilde{U}_i$ (middle image). This is done by mapping the level set of f_1 that contains $x \in U_i$ to the level set of f_2 that contains $w_i(x) \in \tilde{U}_i$. Finally, the map is extended to the region X_{ij} between adjacent critical points p_i, p_j on $ART(f_1)$ and range in the region, $\tilde{X}_{ij}$, between corresponding adjacent critical points $\tilde{p}_i$, $\tilde{p}_j$ (right image). Such regions are diffeomorphic to a disc with a hole in the center, and therefore there is a diffeomorphism between X_{ij} and $\hat{X}_{ij}$ that extends $\tilde{w}_i$ and $\tilde{w}_j$. Note that by Lemma 6 Type 1 and 2 saddles are preserved under the map w, and thus similar pictures would follow for Type 2 saddles.

Proof. Let $ART(f_1) = (V_1, E_1, \mathbb{R}^+, a_1)$ and $ART(f_2) = (V_2, E_2, \mathbb{R}^+, a_2)$. We construct w to be a C^1 diffeomorphism, but similar reasoning can be used to obtain a C^2 diffeomorphism. We prove the forward direction in steps (the steps are pictorially shown in Fig. 9):

1. We may associate critical points p_i of f_1 to corresponding critical points $\tilde{p}_i$ of f_2 via the graph isomorphism $\phi : V_1 \to V_2$.

2. Using the Morse Lemma, there exist neighborhoods $U_i, \tilde{U}_i \subset \mathbb{S}^2$ and diffeomorphisms $w_i : U_i \to \tilde{U}_i$ where $p_i \in U_i$ is a critical point of f_1 and $\tilde{p}_i \in \tilde{U}_i$ is the corresponding critical point of f_2 such that

$$f_1|U_i = h_i \circ f_2 \circ w_i|U_i$$

for the contrast change $h_i : f_2(\tilde{U}_i) \to f_1(U_i)$, $h_i(x) = f_1(p_i) - f_2(\tilde{p}_j) + x$. We may assume that $\{U_i\}$ are disjoint as are $\{\tilde{U}_i\}$. We may also assume that $f_1(U_i) \cap f_1(U_j) = \emptyset$ and $f_2(\tilde{U}_i) \cap f_2(\tilde{U}_j) = \emptyset$ for $i \neq j$ since critical values are assumed to be distinct (by definition of $\mathcal{F}$). Note that $w_i = \tilde{\psi}_i^{-1} \circ \psi_i$ where ψ_i and $\tilde{\psi}_i$ given from applying the Morse Lemma to f_1 and f_2 around the critical points p_i and $\tilde{p}_i$, respectively.

3. Let $\pi_1 : \mathbb{S}^2 \to Reeb(f_1)$ and $\pi_2 : \mathbb{S}^2 \to Reeb(f_2)$ be the natural quotient maps. For each p_i and $\tilde{p}_i$, that correspond to minima or maxima (i.e., $ind(p_i) = ind(\tilde{p}_i) \neq 1$), we may choose $W_i \subset U_i$ and $\tilde{W}_i \subset \tilde{U}_i$ that are open such that $\partial(W_i) = \pi_1^{-1}([q, f_1(q)])$, $\partial(\tilde{W}_i) = \pi_2^{-1}([w_i(q), f_2(w_i(q))])$ for some $q \in U_i$, and $w_i(W_i) = \tilde{W}_i$. We define $\hat{w}_i = w_i|W_i$.

   Now we consider each p_i that is a saddle point (i.e., $ind(p_i) = 1$). By choosing an appropriate subset of U_i and $\tilde{U}_i$ (which for simplicity are denoted by U_i and $\tilde{U}_i$), we may assume that $\pi_1^{-1}([q, f_1(q)]) \cap U_i$ and $\pi_2^{-1}([w_i(q), f_2(w_i(q))]) \cap \tilde{U}_i$ each have at most two connected components for $q \in U_i$. For example, we can choose $U_i = \psi_i^{-1}(B_\epsilon(0))$ and $\tilde{U}_i = \tilde{\psi}_i^{-1}(B_\epsilon(0))$ for $\epsilon < 0$ small and B denotes the disc in $\mathbb{R}^2$.

We now extend each $w_i : U_i \to \tilde{U}_i$ to $\hat{w}_i : W_i \to \tilde{W}_i$ where

$$W_i = \bigcup_{q \in U_i \backslash \{p_i\}} \pi_1^{-1}([q, f_1(q)])$$

$$\tilde{W}_i = \bigcup_{q \in \tilde{U}_i \backslash \{\tilde{p}_i\}} \pi_2^{-1}([q, f_2(q)])$$

We define $\hat{w}_i$ as follows:

- Note that each $\pi_1^{-1}([q, f_1(q)])$ $(q \in \tilde{U}_i \backslash \{\tilde{p}_i\})$ and $\pi_2^{-1}([w_i(q), f_2(w_i(q))])$ are both diffeomorphic to the circle (since q is not a critical point), and therefore diffeomorphic to themselves.
- Let us consider the case when $\pi_1^{-1}([q, f_1(q)]) \cap U_i$ consists of two connected components (the case of one connected component is done similarly). Let A, B, C, D denote points of $\partial(\pi_1^{-1}([q, f_1(q)]) \cap U_i)$ and let $A' = w_i(A), B' = w_i(B), C' = w_i(C), D' = w_i(D)$. We assume that $A \to B \to C \to D \to A$ traverses $\pi_1^{-1}([q, f_1(q)])$. Assume $A \to B$ and $C \to D$ specifies the parts of $\pi_1^{-1}([q, f_1(q)])$ where w_i is defined. Let $c_1, c_2 : [0, 1] \to \mathbb{R}^2$ be parameterized by arc-length parameter (and whose orientation is consistent with the orientation of $A \to B \to C \to D$ and $A' \to B' \to C' \to D'$) of $\pi_1^{-1}([q, f_1(q)])$ and $\pi_2^{-1}([w_i(q), f_2(w_i(q))])$. We define $\varphi : [0, 1] \to [0, 1]$ to be such that
  - $\varphi(0) = 0, \varphi(1) = 1$ and $\varphi'(0) = \varphi'(1)$
  - Define $\varphi(\xi)$ so that $\Xi = c_1(\xi)$ and $\Xi' = c_2(\varphi(\xi))$ for $\xi = 0, b, c, d, 1$, $\Xi = A, B, C, D, A$, resp.
  - Define $\varphi'(\xi)$ so that $\nabla w_i(c_1(\xi)) \cdot c_1'(\xi) = c_2'(\varphi(\xi))\varphi'(\xi)$ where $\xi = 0, b, c, d, 1$.
  - Naturally, we may define φ in the intervals $[0, b]$ and $[c, d]$ as satisfying $w_i(c_1(\xi)) = c_2(\varphi(\xi))$.
  - We define

    $$\varphi(x) = \varphi(b) + \int_b^x g(\xi)\, \mathrm{d}\xi, \quad \text{for } x \in (b, c)$$

    where $g : [b, c] \to \mathbb{R}^+$ satisfies

    $$\int_b^c g(x)\, \mathrm{d}x = \varphi(c) - \varphi(b), \quad g(b) = \varphi'(b), \quad g(c) = \varphi'(c)$$

    and is continuous with respect to b, c, $\varphi'(b), \varphi'(c)$ and x. We may similarly define $\varphi|_{[d,1]}$.

  Next we define $\hat{w}_i$ by setting

  $$\hat{w}_i(c_1(\xi)) = c_2(\varphi(\xi)).$$

- Note that $\hat{w}_i : W_i \to \tilde{W}_i$ is a diffeomorphism because
  - $\hat{w}_i | U_i = w_i$ is a diffeomorphism by the previous step
  - By Lemma 6, w_i does not map a type 1 saddle to a type 2 saddle and vice-versa, and so $\hat{w}_i | (W_i \backslash U_i)$ will be a diffeomorphism, details of which follow.
  - $\hat{w}_i | (W_i \backslash U_i)$ is a diffeomorphism: for the region

    $$\{\pi_1^{-1}([q, f_1(q)]) : q \in U_i \backslash \{p_i\}, \pi_1^{-1}([q, f_1(q)]) \cap U_i \text{ has 2 connected components}\}$$

    and (each connected component of) the region

    $$\{\pi_1^{-1}([q, f_1(q)]) : q \in U_i \backslash \{p_i\}, \pi_1^{-1}([q, f_1(q)]) \cap U_i \text{ has 1 connected component}\}$$

    the parameterization of these regions by the family of c_1 and c_2 are differentiable, and so is the family of φ. Therefore, $\hat{w}_i$ is a differentiable as is its inverse.
  - $Dw_i | \partial U_i = D\hat{w}_i | \partial(W_i \backslash U_i)$: this is by construction of φ in the previous step to be differentiable, and differentiable in its boundary conditions.

4. Finally, we extend the diffeomorphisms $\hat{w}_i$ to form a diffeomorphism $w : \mathbb{S}^2 \to \mathbb{S}^2$. Define w on the neighborhoods W_i so that $w|W_i = \hat{w}_i$. In the following, we define w in the region $\mathbb{S}^2 \setminus \cup_i W_i$.

   Let p_i and p_j be critical points of f_1 with corresponding vertices $v_i, v_j \in V_1$ such that $(v_i, v_j) \in E_1$; also let $\tilde{p}_i, \tilde{p}_j$ be the corresponding critical points of f_2 and $v_i', v_j' \in V_2$ (with $(v_i', v_j') \in E_2$) corresponding vertices. Let $\gamma_{ij} : [0,1] \to Reeb(f_1)$ be a continuous path such that $\gamma_{ij}(0) = [(p_i, f_1(p_i))]$ and $\gamma_{ij}(1) = [(p_j, f_1(p_j))]$. Similarly, let $\tilde{\gamma}_{ij} : [0,1] \to Reeb(f_2)$ be a continuous path such that $\tilde{\gamma}_{ij}(0) = [(\tilde{p}_i, f_2(\tilde{p}_i))]$ and $\tilde{\gamma}_{ij}(1) = [(\tilde{p}_j, f_2(\tilde{p}_j))]$. We define

   $$X_{ij} = \pi_1^{-1}(\gamma_{ij}([0,1])) \setminus (W_i \cup W_j)$$
   $$\tilde{X}_{ij} = \pi_2^{-1}(\tilde{\gamma}_{ij}([0,1])) \setminus (\tilde{W}_i \cup \tilde{W}_j).$$

   Note that X_{ij} and $\tilde{X}_{ij}$ are both diffeomorphic to an annular region in $\mathbb{R}^2$. Therefore, $\partial X_{ij} = \partial_{in} X_{ij} \cup \partial_{out} X_{ij}$ where $\partial_{in} X_{ij}$ denotes the inner boundary of X_{ij} and $\partial_{out} X_{ij}$ denotes the outer boundary.[10]

   We define $\hat{w}_{ij}, w_{ij} : X_{ij} \to \tilde{X}_{ij}$ as follows:

   - We define $\zeta_{ij} : \partial_{in} X_{ij} \times \mathbb{R}^+ \to \mathbb{S}^2$ and $\hat{\zeta}_{ij} : \partial_{in} \tilde{X}_{ij} \times \mathbb{R}^+ \to \mathbb{S}^2$ as

     $$\partial_t \zeta_{ij}(x,t) = \pm \nabla f_1(\zeta_{ij}(x,t)),\ \zeta_{ij}(x,0) = x \in \partial_{in} X_{ij}$$
     $$\partial_t \hat{\zeta}_{ij}(x,t) = \pm \nabla f_2(\hat{\zeta}_{ij}(x,t)),\ \hat{\zeta}_{ij}(x,0) = x \in \partial_{in} \hat{X}_{ij}$$

     where we use the positive gradient direction if $f_1(\partial_{in} X_{ij}) < f_1(\partial_{out} X_{ij})$ otherwise negative. Note that $\zeta_{ij}(\partial_{in} X_{ij}, t)$ ($\tilde{\zeta}_{ij}(\partial_{in} \tilde{X}_{ij}, t)$) is a level set of f_1 (f_2) for each t since $\partial_{in} X_{ij}$ ($\partial_{in} \tilde{X}_{ij}$) is a level set of f_1 (f_2). Also in finite time, T ($\tilde{T}$), $\zeta_{ij}(\partial_{in} X_{ij}, T) = \partial_{out} X_{ij}$ ($\tilde{\zeta}_{ij}(\partial_{in} \tilde{X}_{ij}, \tilde{T}) = \partial_{out} \tilde{X}_{ij}$).
   - Note that $\zeta_{ij}(\partial_{in} X_{ij}, [0,T]) = X_{ij}$ and $\tilde{\zeta}_{ij}(\partial_{in} \tilde{X}_{ij}, [0,\tilde{T}]) = \tilde{X}_{ij}$. We define $w_{ij} : X_{ij} \to \tilde{X}_{ij}$ as

     $$w_{ij}(\zeta_{ij}(x,t)) = \begin{cases} \tilde{\zeta}_{ij}(w_i(x), h_{ij}(t)) & x \in \mathrm{cl}(W_i) \\ \tilde{\zeta}_{ij}(w_j(x), h_{ij}(t)) & x \in \mathrm{cl}(W_j) \end{cases},\ \text{for } x \in \partial_{in} X_{ij},\ t \in [0,T]. \tag{14}$$

     where $h_{ij} : [0,T] \to [0,\tilde{T}]$ is chosen to be smooth, satisfies the conditions

     $$h_{ij}(0) = 0,\ h_{ij}(T) = \tilde{T},\ h_{ij}'(0) = h_i'(f_2 \circ w_i(\partial_{in} X_{ij})),\ h_{ij}'(T) = h_j'(f_2 \circ w_j(\partial_{out} X_{ij})),$$

     and is such that $h : f_2(\mathbb{S}^2) \to f_1(\mathbb{S}^2)$ with the conditions

     $$h(f_1(\partial_{in} X_{ij})) = f_2(\partial_{in} \tilde{X}_{ij}),\ h'(f_1(\partial_{in} X_{ij})) = h_{ij}'(0)$$
     $$h(f_1(\partial_{out} X_{ij})) = f_2(\partial_{out} \tilde{X}_{ij}),\ h'(f_1(\partial_{in} X_{ij})) = h_{ij}'(T)$$
     $$h(v) = h_i(v) \text{ for } v \in f_2(\tilde{U}_i)$$

     is smooth. Note that h is the contrast change that we have been seeking in (13).
   - It is clear that $w_{ij} : X_{ij} \to \tilde{X}_{ij}$ is a diffeomorphism; however it may not be the case that

     $$Dw_{ij}|\partial X_{ij}(x) = \begin{cases} Dw_i|\partial W_i(x) & x \in \partial W_i \\ Dw_j|\partial W_j(x) & x \in \partial W_j \end{cases}. \tag{15}$$

     Indeed by Step 3, recall that we have

     $$f_1(x) = h_i \circ f_2 \circ w_i(x) \text{ for } x \in U_i$$

[10] Note that a simple curve in $\mathbb{S}^2$ does not define an inside and outside; however, we are identifying $\mathbb{S}^2$ with $\mathbb{R}^2$ by specifying that the south pole of $\mathbb{S}^2$ is mapped to infinity.

and so by differentiating, we have

$$\nabla f_1(x) = h_i(f_2 \circ w_i(x)) Dw_i(x) \cdot \nabla f_2(w_i(x)),$$

or

$$Dw_i(x) \cdot \nabla f_1(x) = h_i(f_2 \circ w_i(x)) Dw_i(x) Dw_i^T(x) \nabla f_2(w_i(x)). \tag{16}$$

Next by differentiating (14), we have that

$$Dw_{ij} \cdot \partial_t \zeta_{ij}(x,t) = \partial_t \tilde{\zeta}_{ij}(w_i(x), h_{ij}(t)) h'_{ij}(t)$$

that is

$$Dw_{ij} \cdot \nabla f_1(\zeta_{ij}(x,t)) = h'_{ij}(t) \nabla f_2(\tilde{\zeta}_{ij}(w_i(x), h_{ij}(t))).$$

In order to "adjust" w_{ij} so that (15) holds, we define a new map $\hat{w}_{ij}$ as follows. Let us abuse the notation and let $\partial_{in} X_{ij}, \partial_{in} \tilde{X}_{ij} : \mathbb{S}^1 \to \mathbb{R}^2$ denote smooth parameterizations of the corresponding sets so that $w_i(\partial_{in} X_{ij}(u)) = \partial_{in} \tilde{X}_{ij}(u)$ for all $u \in \mathbb{S}^1$. Define $c_1, c_2 : \mathbb{S}^1 \times [0,1] \to \mathbb{R}^2$ as

$$\begin{aligned} c_1(u,v) &= \zeta(\partial_{in} X_{ij}(u), vT) \\ c_2(u,v) &= \tilde{\zeta}(\partial_{in} \tilde{X}_{ij}(u), h(vT)). \end{aligned}$$

Observe that $w_{ij}(c_1(u,v)) = c_2(u.v)$ for all $(u,v) \in \mathbb{S}^1 \times [0,1]$. We now define $\varphi : \mathbb{S}^1 \times [0,1] \to \mathbb{S}^1$ so that the map $\hat{w}_{ij} : X_{ij} \to \tilde{X}_{ij}$ defined by

$$\hat{w}_{ij}(c_1(u,v)) = c_2(\varphi(u,v), v) \tag{17}$$

satisfies (15). Computing derivatives of (17) we have

$$\frac{\partial}{\partial v} \hat{w}_{ij}(c_1(u,v)) = \partial_u c_2(\varphi(u,v), v) \varphi_v(u,v) + \partial_v c_2(\varphi(u,v), v).$$

Note that by definition of c_2

$$\partial_u c_2(\varphi(u,v), v) = A(u,v)(\nabla f_2(c_2(\varphi(u,v), v)))^\perp$$

where $x^\perp$ means counterclockwise rotation by $\pi/2$, and A is a scalar-valued function. Next, we have that

$$\partial_v c_2(\varphi(u,v), v) = B(u,v) \nabla f_2(c_2(\varphi(u,v), v))$$

for a scalar-valued function B. Now for $v \in \{0,1\}$ we must have that φ satisfies the conditions

$$\begin{aligned} &\varphi(u,0) = u,\ \varphi(u,1) = u \\ &A(u,v)(\nabla f_2(c_2(\varphi(u,v), v)))^\perp \varphi_v(u,v) + B(u,v) \nabla f_2(c_2(\varphi(u,v), v)) = \frac{1}{T} Dw_i(c_1(u,v)) \cdot \nabla f_1(c_1(u,v)) \end{aligned}$$

where $Dw_i(c_1(u,v)) \cdot \nabla f_1(c_1(u,v))$ is specified in (16). In other words, we must choose φ to satisfy the boundary conditions

$$\begin{aligned} \varphi(u,0) = u,\ \varphi(u,1) = u \\ \varphi_v(u,0) = E(u),\ \varphi_v(u,1) = F(u) \end{aligned}$$

where $E, F : \mathbb{S}^1 \to \mathbb{R}^+$ are specified. Note that in the interior of $\mathbb{S}^1 \times [0,1]$, we need the monotonicity condition that

$$\varphi_u > 0.$$

We may specify φ in the interior of $\mathbb{S}^1 \times [0,1]$ to, for example, satisfy:

$$\varphi_{uuuu} + \varphi_{vvvv} = 0.$$

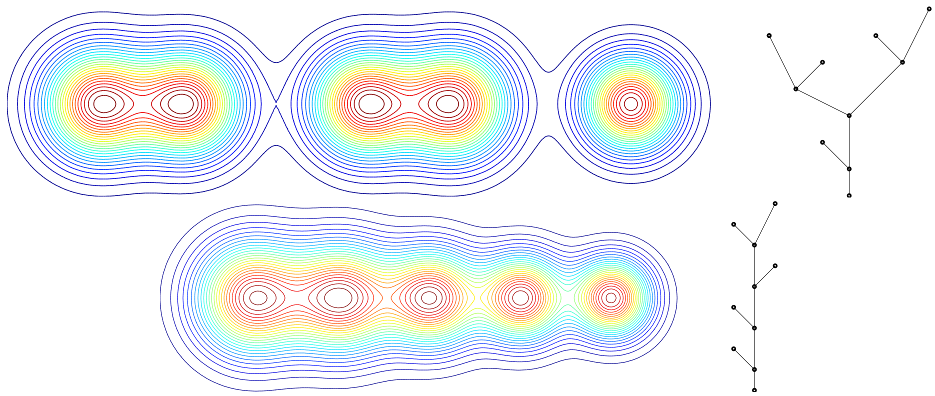

Figure 10: This figure shows the importance of the structure of the ART in determining whether two functions are in the same equivalence class. The figure shows the level sets of two functions and their corresponding Reeb trees. In this case, each function has the same number of min/max/saddles, and values, but the ARTs are different and the functions are not equivalent via a viewpoint/contrast change.

Now $w|X_{ij} = \hat{w}_{ij}$ and $w|W_i = \hat{w}_i$ specifies a diffeomorphism $w : \mathbb{S}^2 \to \mathbb{S}^2$.

□

Remark 12. *Note that there is no subset (in general) of the attributed Reeb tree that is sufficient to determine the domain diffeomorphism w. In other words the vertices, their values and their indices are not a sufficient statistic to determine a domain diffeomorphism, w. To see this, we give an example of two attributed Reeb trees that have the same number and types of critical points and values, but are not equivalent (see Figure 10).*

Remark 13. *Condition 2 in Definition 6 ensures that $ART(f)$ does not change under small perturbations of f, e.g., $f + \epsilon g$ for small ϵ. This property is important in image analysis since the presence of noise in images is common, and thus, we are interested in a class of functions that are stable under small amounts of noise.*

To demonstrate this point, consider the following function with two saddle points that have the same function value and belong to the same connected component of a level set:

$$f(x, y) = \exp\left[-(x^2 + y^2)\right] + \exp\left[-((x - 3)^2 + y^2)\right] + \exp\left[-((x + 3)^2 + y^2)\right];$$

the function and its attributed Reeb tree is plotted in the top of Figure 11. Now consider a slightly perturbed version of f:

$$g(x, y) = \exp\left[-(x^2 + y^2)\right] + \exp\left[-(1 + 2\epsilon)((x - 3)^2 + y^2)\right] + \exp\left[-(1 + \epsilon)((x + 3)^2 + y^2)\right],$$

where $\epsilon > 0$; the function is plotted in the bottom of Figure 11. Although f only differs from g by a slight perturbation, the attributed Reeb trees are not equivalent. Indeed f is not a stable function under small perturbations, while the function g is stable.

Further, Condition 2 simplifies our classification of the equivalence of functions under contrast and viewpoint changes. Indeed, the attributed Reeb tree may not contain enough information to determine a domain diffeomorphism w between two functions with same Reeb tree in the case of multiple saddles belonging to the same connected component of a level set. In such a case, multiple saddle points of a function coalesce to a single point in the ART. The graph isomorphism ϕ in the proof of Lemma 7 may not be enough to determine the correspondence between saddles of f_1 and those of f_2 in this case since ϕ only associates the group of coalesced saddles of f_1 to the group of coalesced saddles of f_2.

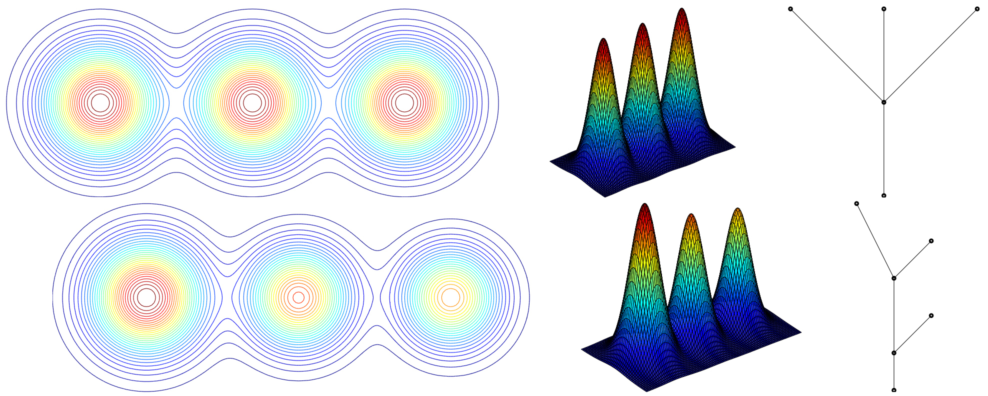

Figure 11: Top: A Morse function (its level sets, surface, and attributed Reeb tree, respectively) of a function with multiple saddles on the same connected component of a level set. Bottom: a slightly perturbed version of the above Morse function. The attributed Reeb tree of the function on the top is not stable under small perturbations; while the one on the bottom is stable.

Lemma 8. *For each $T \in \mathcal{T}$, there exists a Morse function $f \in \mathcal{F}$ so that $ART(f) = T$.*

Proof. Let $T' \in \mathcal{T}'$ be any representative of T. We apply the following algorithm to construct the level sets of f in $\mathbb{R}^2$ so that $ART(f) = T$. The algorithm recursively traverses the tree T starting from v_{sp}, constructing the level sets of f out from infinity in $\mathbb{R}^2$ (equivalently, the south-pole of $\mathbb{S}^2$) inward.

- Let $R = \{x \in \mathbb{R}^2 : |x| \leq 1\}$. We define f on $\mathbb{R}^2 \backslash R$ so that the level sets of f inside the region R are $L_\delta = \{x \in \mathbb{R}^2 : |x| = \delta\}$ for each $\delta > 1$.
- Set v to be the vertex adjacent to v_{sp}.
- `SubAlgorithm(`v`,`R`)`
  - If there are no vertices adjacent to v that have not been visited, then v corresponds to a minimum or maximum of a function with ART T. We define f in R is defined to be diffeomorphically equivalent to $g : B_1(0) \to \mathbb{R}$ ($B_1(0)$ is the ball of radius 1 centered at 0) $g(x) = \pm(x_1^2 + x_2^2)$ (+ if v corresponds to a minimum, and $-$ if v corresponds to a maximum) and consistent with f already constructed on ∂R.
  - Otherwise, let v_1, v_2 be the two vertices adjacent to v that have not been visited.
    - Let $R' \subset R$ such that $\text{cl}(R') \subset R$ and let R' be a 2-fold connected closed set (i.e., a region with two holes).
    - If $a(v_1), a(v_2) > a(v)$ or $a(v) > a(v_1), a(v_2)$ then v must correspond to a Type 1 saddle point [11]. Define f on R' to be diffeomorphically equivalent to a function that has level sets illustrated in the left of (in case $a(v) > a(v_1), a(v_2)$) or in the right of Figure 5 (in case $a(v) > a(v_1), a(v_2)$).
    - If $a(v_1) > a(v) > a(v_2)$ or $a(v_1) < a(v) < a(v_2)$, then v must be a Type 2 saddle point. Let v_{prev} be the vertex that was previously visited. Define f on R' to be diffeomorphically equivalent to a function that has level sets illustrated in the left of (in case $a(v_{prev}) < a(v)$) or in the right of Figure 6 (in case $a(v_{prev}) > a(v)$).

[11] Distinguishing between Type 1 and Type 2 saddles is based on the order of the traversal of the tree, T, i.e., the order of vertices visited before and after saddle vertices. Note that the algorithm constructs the function f from outward regions inwards. Thus, level sets of a function corresponding to the interior of the edge (v_{prev}, v) (where v_{prev} is the vertex visited prior to v) enclose the domain of the function corresponding the portion of the tree containing vertices v_1, v_2, v. By looking at Figs. 5, 6, Type 1 saddles are such that v_1, v_2 have attributes that are either both less or greater than v, otherwise they are Type 2 saddles.

* The region, $R'' = R \backslash cc(R')$ (where $cc(R')$ indicates the intersection of all simply connected regions containing R') is diffeomorphic to an annulus. The function f in R'' is constructed so that it is diffeomorphic to $g : B_2(0) \backslash B_1(0) \to \mathbb{R}^+$ defined by $g(x_1, x_2) = \pm(x_1^2 + x_2^2)$ (+ if $a(v) < a(v_{prev})$, − otherwise) and maintaining any boundary conditions in $\partial R''$ imposed by the previous steps of the algorithm. This is guaranteed by the Collar Theorem of differential topology (see [14]).
* Repeat `SubAlgorithm(`v_1`,`R_1`)`, `SubAlgorithm(`v_2`,`R_2`)` where R_1 and R_2 are the two connected components of $cc(R') \backslash R$.

- The values of the level sets are chosen so as to be consistent with the attributes of the ART.

□

Collecting all these results together, we have the following result.

Theorem 3. *The attributed Reeb tree of a surface uniquely determines it up to a contrast change and domain diffeomorphism. Equivalently, the orbit space of surfaces that are graphs of Morse functions, $\mathcal{F}$, under the action of contrast and domain diffeomorphisms, $\mathcal{H} \times \mathcal{W}$, is*

$$\boxed{\mathcal{S}'' = \mathcal{T}} \tag{18}$$

Proof. We can define the mapping $ART : \mathcal{S}/(\mathcal{H} \times W) \to \mathcal{T}$ by

$$ART([f]) := ART(f), \text{ where } [f] = \{h \circ f \circ w \in \mathcal{F} : (h, w) \in \mathcal{H} \times \mathcal{W}\}$$

The function above is well-defined since any representative $g \in [f]$ will have the same Attributed Reeb Tree. Note

- Lemma 7 states that $ART : \mathcal{S}/(\mathcal{H} \times W) \to \mathcal{T}$ is injective.
- Lemma 8 states that $ART : \mathcal{S}/(\mathcal{H} \times W) \to \mathcal{T}$ is surjective.
- Therefore, $ART : \mathcal{S}/(\mathcal{H} \times W) \to \mathcal{T}$ is a bijection and therefore, $\mathcal{S}/(\mathcal{H} \times W) = \mathcal{T}$.

□

Remark 14. *The results above do not cover the case of surfaces that are not graphs of Morse functions. In the context of image analysis we always deal with surfaces that are graphs (the intensity values), but in general they are neither smooth nor have isolated extrema. Lack of smoothness is caused by discontinuities for instance due to occlusions and material boundaries. Therefore, the analysis above applies only to a* segment *(a sub-set) of the image domain, which can be mapped without loss of generality to the unit square. Non-isolated extrema such as ridges and valleys are also commonplace in images, but they are accidental in the sense that a ridge with constant height can be turned into a Morse function by slightly perturbing it, thus generating a maximum along the ridge. The ART is stable with respect to such perturbations, although one could question the loss of discriminative power of the representation of ridges as "thin blobs" that renders them indistinguishable from other blobs, regardless of their shape.*

4 Structural Stability to Non-Invertible Nuisances

Definition 14 (Structural Stability to Scale)**.**

1. *An image $I : \Omega \to \mathbb{R}^+$ is* structurally stable to scale σ_0 *if $ART(I * G_\sigma) = ART(I)$ for all $\sigma \leq \sigma_0$.*
2. *An image $I : \Omega \to \mathbb{R}^+$ is* structurally stable at scale σ to scale σ_0 *if $G_\sigma * I$ is the smallest σ for which structurally stable to scale σ_0.*

5 Conclusion

In this manuscript we have focused on analyzing portions of the image that exhibit smooth shading or smooth texture variations. Such regions of the image would be discarded by most feature selectors used in the recognition literature as they contain no discontinuities (edges or corners). They would also be "misinterpreted" by any segmentation algorithm, as the smooth gradient would generate spurious boundaries that are unstable with respect to perturbations of the image [7]. And yet, smoothly shaded regions convey "information" that can be useful for recognition. We have shown that

- It is possible to compute functions of an image region that exhibits smooth variation that are invariant to both viewpoint and a coarse illumination model (contrast transformations), called ARTs.
- Such functions are sufficient for recognition of objects and scenes under changes of viewpoint and illumination, in the sense that they are equivalent to the image up to an arbitrary change of viewpoint (domain diffeomorphism) and contrast transformation (a first-order approximation of illumination changes).
- Such functions have support on a set of measure zero of the image domain.

These results do not cover the case of image surfaces that are not graphs of Morse functions. These include discontinuities and ridges/valleys. Therefore, the analysis above applies only to a *segment* (a subset) of the image domain, which can be mapped without loss of generality to the unit square. Non-isolated extrema such as ridges and valleys are also commonplace in images; they can be turned into a Morse function by an infinitesimal perturbation. One could question the loss of discriminative power of the representation of ridges as "thin blobs" that renders them indistinguishable from other blobs, regardless of their shape. Contrast transformations are only a coarse model of the complex effects that illumination changes induce in an image. Devising illumination models that are phenomenologically consistent and yet amenable to analysis is an open research topic in computer vision.

Appendix

A Epipolar diffeomorphisms

In this section we derive Equation (2) from (1).

If we call $R_1 \doteq [1\ 0\ 0]R$, $R_2 = [0\ 1\ 0]R$, and similarly R_3, T_1, T_2, T_3, we have, writing explicitly (1)

$$w(x|R,T,Z(\cdot)) = \frac{\begin{bmatrix} R_1 \\ R_2 \end{bmatrix} \bar{x} Z(x) + \begin{bmatrix} T_1 \\ T_2 \end{bmatrix}}{R_3 \bar{x} Z(x) + T_3}. \tag{19}$$

This equation specifies the class of allowable domain diffeomorphisms under changes of viewpoint away from occlusions, when the scene is rigid and Lambertian, $x \mapsto w(x|R,T,Z(\cdot))$. Thus, once the (positive, scalar-valued) function $Z(\cdot)$, the matrix $R \in \mathbb{G}L(3)$ and the vector $T \in \mathbb{R}^3$ are determined, so is the diffeomorphism w.

To make more explicit the dependency between w_x and w_y, we can imagine choosing w_x arbitrarily, which in turn determines

$$Z(x) = \frac{w_x(x)T_3 - T_1}{R_1\bar{x} - w_x(x)R_3\bar{x}},$$

and after substituting and simplifying, this uniquely determines $w_2(x)$ as a function of R and T:

$$\boxed{w_y(x) = w_x(x)\frac{R_2\bar{x}T_3 - R_3\bar{x}T_2}{R_1\bar{x}T_3 - R_3\bar{x}T_1} + \frac{R_1\bar{x}T_2 - R_2\bar{x}T_1}{R_1\bar{x}T_3 - R_3\bar{x}T_1}.} \tag{20}$$

So, of all diffeomorphisms $w : \mathbb{R}^2 \to \mathbb{R}^2$, we can consider the class implicitly defined by the constraint

$$\langle \bar{w}(x), [R_2\bar{x}T_3 - R_3\bar{x}T_2,\ -(R_1\bar{x}T_3 - R_3\bar{x}T_1),\ R_1\bar{x}T_2 - R_2\bar{x}T_1]^T\rangle = 0. \tag{21}$$

Equivalently, the diffeomorphism w, written in homogeneous coordinates $\bar{w}(x) = [w_1(x),\ w_2(x),\ 1]$ has to be orthogonal, for all $x \in \mathbb{R}^2$, to the function

$$w^{\perp}(x) \doteq \begin{bmatrix} R_2\bar{x}T_3 - R_3\bar{x}T_2 \\ -(R_1\bar{x}T_3 - R_3\bar{x}T_1) \\ R_1\bar{x}T_2 - R_2\bar{x}T_1 \end{bmatrix} = \widehat{T}R\bar{x} \tag{22}$$

where the reader will recognize the latter expression from epipolar geometry [12]. The set of allowable diffeomorphisms, under no occlusions, Lambertian reflection and rigidity, is therefore

$$\boxed{\mathcal{W} \doteq \{w : \mathbb{R}^2 \to \mathbb{R}^2\ | \langle \bar{w}(x), \widehat{T}R\bar{x}\rangle = 0,\ \text{for some } (R,T) \in \mathbb{A}(3)\}.} \tag{23}$$

which is (2).

Proof of Theorem 1

Proof. Assume $w_3 \in \mathcal{W}$, and therefore there exist R_3, T_3, Z_3 such that $w_3 = w(x|R_3, T_3, Z_3)$. Now consider $w_1 \circ w_2$, which can be written as $\pi(R_2R_1\bar{x}Z_1(x)\frac{Z_2(\bar{\pi}(R_1\bar{x}Z_1(x)+T_1))}{e_3\cdot(R_1\bar{x}Z_1(x)+T_1)} + R_2T_1\frac{Z_2(x)}{e_3\cdot(R_1\bar{x}Z_1(x)+T_1)} + T_2)$, where it can be seen that it is not possible to choose a constant T_3 unless $\frac{Z_2}{e_3\cdot(R_1\bar{x}Z_1(x)+T_1)} = 1$ for all x, which imposes a non-generic condition on Z_1 and Z_2, hence the contradiction. □

Proof of Theorem 2

Proof. We note that orientation preserving diffeomorphisms of the plane can be generated by integrating time-varying vector fields:

$$\begin{cases} \dot{w}(t,x) = v(t, w(t,x)) & t \in [0,1],\ x \in \mathbb{R}^2 \\ w(0,x) = x & x \in \mathbb{R}^2 \end{cases}$$

where $v, w : [0,1] \times \mathbb{R}^2 \to \mathbb{R}^2$, and $w(1,\cdot)$ is the generated diffeomorphism. If $w_{1,t}, w_{2,t} \in \tilde{\mathcal{W}}$ is a family of diffeomorphisms, then

$$\frac{\partial}{\partial t} w_{1,t} \circ w_{2,t} = (\partial_t w_{1,t}) \circ w_{2,t} + (Dw_{1,t} \circ w_{2,t}) \cdot \partial_t w_{2,t} = v_{1,t} \circ w_{2,t} + (Dw_{1,t} \circ w_{2,t}) \cdot v_{2,t}.$$

Therefore from the previous expression, it is apparent that if the linear span of the vector fields generated by $w \in \tilde{\mathcal{W}}$ is all possible smooth vector fields, then the closure of $\tilde{\mathcal{W}}$ is the set of orientation preserving diffeomorphisms.

Let $w(\cdot|g_t, Z)$ be a family of diffeomorphisms where $t \mapsto g_t$ is such that $g_t \in SE(3)$ corresponds to a path of viewpoint changes and Z is a fixed surface. We show that

$$\text{span}\left(\left\{\frac{\partial}{\partial t} w(\cdot|g_t, Z) : g_t \in SE(3), Z \text{ satisfies the condition in (3)}\right\}\right) \tag{24}$$

is the set of smooth vector fields. Indeed,

$$\frac{\partial}{\partial t} w(\cdot|g_t, Z) = \frac{(\partial_t R_t\bar{x}Z(x) + \partial_t T_t)(R_{3,t}\cdot\bar{x}Z(x) + T_{3,t}) + (R_t\bar{x}Z(x) + T_t)(\partial_t R_{3,t}\cdot\bar{x}Z(x) + \partial_t T_{3,t})}{(R_{3,t}\cdot\bar{x}Z(x) + T_3)^2},$$

where $g_t = ((R_t, R_{3,t}), (T_t, T_{3,t}))$, and that may be expressed in the form

$$\left.\frac{\partial}{\partial t} w(x_1, x_2|g_t, Z)\right|_{t=0} = \frac{1}{d_1x_1Z(x) + d_2x_2Z(x) + d_3}\left[(a_1x_1^2 + a_2x_1x_2 + a_3x_2^2)Z^2(x) + (b_1x_1 + b_2x_2 + b_3)Z(x) + c_1\right]$$

where $x = (x_1, x_2)$, $d_i \in \mathbb{R}$ and $a_i, b_i, c_i \in \mathbb{R}^2$. By choosing $g_t(0)$ and $\partial g_t(0)$ appropriately, we may obtain arbitrary coefficients. Therefore, it is apparent that the span in (24) contains both the sets

$$\left\{ \begin{pmatrix} Z_1(x_1, x_2) \\ 0 \end{pmatrix} : Z_1 : \mathbb{R}^2 \to \mathbb{R} \right\} \text{ and } \left\{ \begin{pmatrix} 0 \\ Z_2(x_1, x_2) \end{pmatrix} : Z_2 : \mathbb{R}^2 \to \mathbb{R} \right\},$$

which establishes our claim that (24) is the set of smooth vector fields.

□

References

[1] L. Alvarez, F. Guichard, P. L. Lions, and J. M. Morel. Axioms and fundamental equations of image processing. *Arch. Rational Mechanics*, 123, 1993.

[2] A. Ayvaci, M. Raptis, and S. Soatto. Optical flow and occlusion detection with convex optimization. In *Proc. of Neuro Information Processing Systems (NIPS)*, December 2010.

[3] Coloma Ballester, Vicent Caselles, and Pascal Monasse. The tree of shapes of an image. *ESAIM: Control, Optimization and Calculus of Variations*, 9:1–18, 2003.

[4] J. B. Burns, R. S. Weiss, and E. M. Riseman. The non-existence of general-case view-invariants. In *Geometric Invariance in Computer Vision*, pages 120–131, 1992.

[5] V. Caselles, B. Coll, and J.-M. Morel. Topographic maps and local contrast changes in natural images. *Int. J. Comput. Vision*, 33(1):5–27, 1999.

[6] H. F. Chen, P. N. Belhumeur, and D. W. Jacobs. In search of illumination invariants. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, 2000.

[7] C. Galleguillos, B. Babenko, A. Rabinovich, and S. Belongie. Weakly supervised object localization with stable segmentations. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, 2007.

[8] J. J. Gibson. *The ecological approach to visual perception.* LEA, 1984.

[9] T. Lee and S. Soatto. Video-based descriptors for object recognition. *Image and Vision Computing*, 2011.

[10] J. L. Lisani, L. Moisan, P. Monasse, and J. M. Morel. Affine invariant mathematical morphology applied to a generic shape recognition algorithm. *Computational Imaging and Vision*, 18, 2000.

[11] D. G. Lowe. Object recognition from local scale-invariant features. In *ICCV*, 1999.

[12] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry. *An invitation to 3D vision, from images to geometric models.* Springer Verlag, 2003.

[13] Jiri Matas, Ondrej Chum, Martin Urban, and Tomás Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In Paul L. Rosin and A. David Marshall, editors, *BMVC*. British Machine Vision Association, 2002.

[14] J. Milnor. *Morse Theory.* Annals of Mathematics Studies no. 51. Princeton University Press, 1969.

[15] Pascal Monasse and Frederic Guichard. Fast computation of a contrast-invariant image representation. *IEEE Transactions on Image Processing*, 9(5):860–872, 2000.

[16] G. Reeb. Sur les points singuliers d'une forme de pfaff completement integrable ou d'uun function numerique. *Comp. Rend. Acad. Science Paris*, 222:847–849, 1946.

[17] Philippe Salembier and Luis Garrido. Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. *IEEE Transactions on Image Processing*, 9(4):561–576, 2000.

[18] Y. Shinagawa. Homotopic image pseudo-invariants for openset object recognition and image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(11):1891–1901, Nov. 2008.

[19] S. Soatto. Actionable information in vision. In *Proc. of the Intl. Conf. on Comp. Vision*, October 2009.

[20] S. Soatto. *Steps Toward a Theory of Visual Information.* UCLA Technical Reports: http://www.cs.ucla.edu/csd/research/archives.html, Technical Report UCLA-CSD100028, September 13, 2010.

[21] A. Vedaldi and S. Soatto. Features for recognition: viewpoint invariance for non-planar scenes. In *Proc. of the Intl. Conf. of Comp. Vision*, pages 1474–1481, October 2005.