On the Set of Images Modulo Viewpoint and Contrast Changes

Ganesh Sundaramoorthi∗  Peter Petersen†  Stefano Soatto‡

December 12, 2010

Abstract

We consider regions of images that exhibit smooth statistics, and pose the question of characterizing the “essence” of these regions that matters for visual recognition. Ideally, this would be a statistic (a function of the image) that does not depend on viewpoint and illumination, and yet is sufficient for the task. In this manuscript, we show that such statistics exist. 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 modulo diffeomorphisms 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. 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. As is known [22], any statistic of an image that is invariant to viewpoint is also invariant to shape, in the sense that images produced from scenes that are deformed versions of each other are lumped into the same class. As already pointed out in [22], this does not mean that one cannot recognize scenes that have different shape; it means that one cannot do so by means of invariant statistics. 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. We also assume that the images have infinite resolution. Both assumptions are unrealistic in practice; however, the analysis of this simplified case is important to understand the sparse nature of visual information (as we will see, the Attributed Reeb Tree is supported on a zero-measure subset of the image domain), and is relevant locally in the image, away from

∗ Department of Computer Science, UCLA
† Department of Mathematics, UCLA
‡ Department of Computer Science, UCLA
discontinuities. Lifting these assumptions is clearly important for engineering applications; this is discussed at length in [21].

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 conditions, occlusions and other visibility 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 there is a cogent need to factor out as many of these sources of variability as possible as part of the representation in a “pre-processing” phase. Ideally, one would want a representation of the data (images) that is invariant to nuisance factors, intrinsic or extrinsic and that represents a sufficient statistic for the task at hand. The most common nuisances in recognition are (a) viewpoint, (b) illumination, (c) visibility artifacts such as occlusions and cast shadows, (d) quantization and noise. 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 an image occludes another cannot be determined from an image alone, but can be ascertained as part of the matching process with another image [2]. What about the former two? Can one devise image representations that are invariant to both viewpoint and illumination, at least away from visibility artifacts such as occlusions and cast shadows?


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 a sufficient statistic, in the sense that it contains all and only the information necessary to accomplish the task, regardless of viewpoint and illumination. For the case of viewpoint, although earlier literature [5] suggested that general-case view-invariants do not exist, it has been shown that it is always possible to construct non-trivial viewpoint invariant image statistics for Lambertian objects of any shape [22]. 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 that general-case (global) illumination invariants do not exist, even for Lambertian objects. However, there is a considerable body of literature [1, 15, 3] dealing with more restricted illumination models that induce a monotonic continuous transformation of the image intensities, a.k.a. contrast transformation. It has been shown 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.

But even in this more constrained illumination model, 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. Even less hopeful that, if we find something, it would be a sufficient statistic! Yet, 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 one-dimensional subset of the image domain; finally, despite being thin, (iii) these invariants are sufficient statistics, in the sense that they are equivalent to the original data but for the effects of viewpoint and contrast.

---

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 [20].
4 The results of [5] refer to statistics of perspective measurements of point ensembles, although they have been subsequently misinterpreted as referring to image statistics.
5 Note that the image representation presented in these papers are contrast invariant, but are not viewpoint invariant.
6 This fact is exploited by the most successful local representations for recognition, such as the scale-invariant feature transform (SIFT) and the histogram of oriented gradients (HOG).
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 statistics? These would include shaded regions (Fig. 1) 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 statistics 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)\(^8\), a deterministic construction that is uniquely determined from an image and is a one-dimensional subset of the image.

4. We show that the set of viewpoint changes in space induce the entire set of diffeomorphisms on the domain of the image, when the three-dimensional shape of the scene is unknown.

5. We show that two images that have the same ART are related by a domain diffeomorphism and a contrast transformation.

6. We conclude that the \(\text{ART}\) is a viewpoint-illumination invariant\(^9\).

7. Finally, we show that the \(\text{ART}\) is a sufficient statistic, in the sense that it is equivalent to the original image up to an arbitrary domain diffeomorphism and contrast change\(^{10}\).

\(^7\) Many representations currently used for recognition involve combinations of these structures, such as extrema of difference-of-Gaussians (“blobs”), non-singularities of the second-moment-matrix (“corners”), sparse coding (“bases”) and segmentation or other processes to determine region boundaries.

\(^8\) 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 introduce this tree representation of the image, but to show that ART is a viewpoint/contrast sufficient statistic (see 4-8 above), and this is not considered in [17]. Also, we should mention that the ART of an image essentially characterizes the connected components of level set of an image treated as a function, and that (called extremal regions) has been considered in [13], but here the authors are unaware that extremal regions are viewpoint invariant, and indeed do not state nor prove the results of 4-8.

\(^9\) Note that in [11], invariants to the constrained case of affine domain changes and contrast changes are considered. Our paper considers the general case of viewpoint changes, which correspond to general domain diffeomorphisms when the three-dimensional shape of the scene is unknown.

\(^{10}\) Note that this does not necessarily mean that a viewpoint-illumination invariant is a unique signature for an object. As
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 [9], and has been called “Actionable Information” in [20].

Clearly this is only a piece of the puzzle. It would be simplistic to argue that our key assumption, which we introduce in the next section, is made without loss of generality (Morse functions are dense in $C^2$, which is dense in $L^2$, and therefore they can approximate any discontinuous, square-integrable function to within an arbitrarily small error). Co-dimension one extrema (ridges, valleys, edges) in images are qualitatively different than regions with smooth statistics and should be treated as such, rather than generically approximated. This is beyond our scope in this paper, where we restrict our analysis away from such structures and only consider regions with smooth statistics. Our goal here is not to design another low-level image descriptor, but 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 [19] and shape coding [4] use a representation closely related to the proof-of-concept construction.

This paper’s contribution is theoretical and we do not make any claims on the practicality of this approach; however, we do believe the theory will be useful in designing better visual recognition systems. For example, many of the ideas presented in this paper have already inspired the visual recognition system [10] that has been implemented online on the iPhone (see http://www.youtube.com/watch?v=cMv-McHw660).

2 Image Invariants: Viewpoint and Illumination

2.1 Invariance to Viewpoint and Illumination

Let $S$ denote the set of closed, compact, smooth surfaces without boundary. The class $S$ is a representative of the space of all the boundaries of objects in the real world. We denote by $\rho_S : S \rightarrow \mathbb{R}^+$, $\rho_S \in A$ a function representing the albedo of $S \in S$. 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) and an illumination (contrast) $h \in H$ which is a monotonic function $h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, we denote the process of image formation as a function $F : S \times A \times SE(3) \times H \rightarrow I$ where $I = \{I : \Omega \rightarrow \mathbb{R}^+\}$ is the space of images:

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

We now define an invariant to viewpoint and illumination:

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

$$\mu(F(S, \rho_S, g, h)) = \mu(F(S, \rho_S, g', h'))$$

for all $g, g' \in SE(3)$, and $h, h' \in H$.

The set $V$ is called the set of invariants.

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

**Definition 3.** A maximal invariant $\mu$ 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 $g, g' \in SE(3)$, $h, h' \in H$ and $S, S' \in S$.

**Remark 1.** A maximal invariant is such that two images that are formed from different scenes do not have the same invariant representation.

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

[22] 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 [22].
2.2 Image Formation and Visibility

Our model for image formation will be simple: we assume our imaging device is a pinhole camera.

Definition 4. Given a viewpoint $g = (R, T) \in SE(3)$ ($R \in SO(3)$, $T \in \mathbb{R}^3$) and an object $S \in S$, the pinhole is at the origin in $\mathbb{R}^3$, the imaging plane $\Omega^g \subset \mathbb{R}^3$ (an embedding of $\Omega \subset \mathbb{R}^2$) is at $T$ and its orientation is determined by $R$. A point $X \in \text{Range}(S)$ is visible from viewpoint $g$ and the imaging plane $\Omega^g$ if the line segment from the origin to the point $X$ intersects $\Omega^g$ and (the line segment) does not intersect any point in $\text{Range}(S) \backslash \{X\}$. A camera projection $\pi$ from a viewpoint $g$ is a map from the visible points of the object $S$ to $\Omega$ given by the point of intersection described earlier. If the imaging plane, $\Omega$ lies on the $x−y$ plane (coordinates relative to the surface $S$), then $\pi$ is given by

$$\pi(X) = \frac{1}{X_3}(X_1, X_2), \text{ where } X = (X_1, X_2, X_3) \in \mathbb{R}^2 \times \mathbb{R}^+.$$

Now we may refine our definition of viewpoint/illumination invariance to take into account visibility.

Definition 5. Let $\mathcal{V}$ be a set. A functional $\mu : \text{Range}(F) \subset \mathcal{I} \to \mathcal{V}$ 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 $S \in S$, $g, g' \in SE(3)$ such that $S$ is visible from $g$ and $g'$.

Remark 4. 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 $S \in S$, $g, g' \in SE(3)$ such that $S$ is visible from $g$ and $g'$."

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 the Lambertian assumption we get that, if $\rho$ is the diffuse albedo, then an image $I(x) = \rho(p)$, were $x = \pi(p)$, is related to another image $J(x') = \rho(p)$, where $x' = \pi(gp) \equiv w(x)$. 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, in the absence of intrinsic calibration data,$^{11}$ $\rho \in A(3)$, 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 $p = \bar{x}Z(x)$ 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

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

---

$^{11}$Assuming calibrated data corresponds to assuming that the camera having captured the training image has the same calibration, whatever it is, of the camera that captured the test image.
where \( x \in \Omega \subset \mathbb{R}^2 \) is the domain for which no (self-)occlusions occur. This limits the range of motions \((R, T)\)

deciding on the shape \( Z(\cdot) \), which is unknown. If we call \( R_2 = [1 0 0]R \), \( R_2 = [0 1 0]R \),

and similarly \( R_3, T_1, T_2, T_3 \), we have, writing explicitly the above equation
\[
\begin{bmatrix}
  w_x(x) \\
  w_y(x)
\end{bmatrix} = \begin{bmatrix}
  R_1 & T_1 \\
  R_2 & T_2
\end{bmatrix} \begin{bmatrix}
  \bar{x}Z(x) + \frac{T_1}{R_3} \\
  \bar{x}Z(x) + \frac{T_3}{R_3}
\end{bmatrix}.
\]

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 \( R \in \mathbb{GL}(3) \) 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 \( T \):
\[
w_y(x) = \frac{R_2 \bar{x} T_3 - R_3 \bar{x} T_2}{R_1 \bar{x} T_3 - R_3 \bar{x} T_1} + \frac{R_1 \bar{x} T_2 - R_2 \bar{x} T_1}{R_1 \bar{x} T_3 - R_3 \bar{x} T_1}.
\]

So, of all diffeomorphisms \( w : \mathbb{R}^2 \to \mathbb{R}^2 \), we can consider the class implicitly defined by the constraint
\[
\langle \bar{w}(x), [R_2 \bar{x} T_3 - R_3 \bar{x} T_2, -(R_1 \bar{x} T_3 - R_3 \bar{x} T_1), R_1 \bar{x} T_2 - R_2 \bar{x} T_1]^T \rangle = 0.
\]

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
\[
w^\perp(x) \doteq \begin{bmatrix}
  R_2 \bar{x} T_3 - R_3 \bar{x} T_2 \\
  -(R_1 \bar{x} T_3 - R_3 \bar{x} T_1) \\
  R_1 \bar{x} T_2 - R_2 \bar{x} T_1
\end{bmatrix} = \hat{T} R \bar{x}
\]

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
\[
\mathcal{W} = \{ w : \mathbb{R}^2 \to \mathbb{R}^2 \mid \langle \bar{w}(x), \hat{T} R \bar{x} \rangle = 0, \text{ for some } (R, T) \in \mathbb{A}(3) \}.
\]

The 3 \( \times \) 3 matrix \( \hat{T} R \) is a fundamental matrix (it is an essential matrix when the cameras are calibrated and hence \((R, T) \in \mathbb{SE}(3))\).

**Remark 5.** Note that if \( \mathcal{W} \) is a group under composition, then the maximal image invariant to viewpoint/contrast is the orbit space, \( S(\mathcal{H} \times \mathcal{W}) \). We now note that, 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 \( w_3 = w_1 \circ w_2 \) may not be an element of \( \mathcal{W} \).

**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 \( w_3 \circ w_2 \), which can be written as \( \pi(R_2 R_1 \bar{x} Z_1(x) Z_2(x) \frac{Z_2(Z_2(x) + T_1)}{e_{\pi(x) + T_1}} + R_2 T_1 \frac{Z_2(x)}{e_{\pi(x) + T_1}} + T_2) \), where it can be seen that it is not possible to choose a constant \( T_3 \) unless \( \frac{Z_2}{e_{\pi(x) + T_1}} = 1 \) for all \( x \), which imposes a non-generic condition on \( Z_1 \) and \( Z_2 \), hence the contradiction. \( \square \)

**Remark 6.** 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 \( W \), 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 \( W \) under which visibility conditions are satisfied:

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

(7)

We now show that the group closure of \( \tilde{W} \) is the entire set of diffeomorphisms:

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

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

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

where \( v, w : [0,1] \times \mathbb{R}^2 \to \mathbb{R}^2 \), and \( w(1, \cdot) \) is the generated diffeomorphism. If \( w_{1,t}, w_{2,t} \in \tilde{W} \) 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 \( w \in \tilde{W} \) is all possible smooth vector fields, then the closure of \( \tilde{W} \) 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

\[
\text{span} \left( \left\{ \frac{\partial}{\partial t} w(\cdot|g_t, Z) : g_t \in SE(3), Z \text{ satisfies the condition in } (7) \right\} \right)
\]

(8)

is the set of smooth vector fields. Indeed,

\[
\frac{\partial}{\partial t} w(\cdot|g_t, Z) = \left( \partial_t R_{t} \tilde{x}Z(x) + \partial_t T_{t} (R_{3,t} \cdot \tilde{x}Z(x) + T_{3,t}) + (R_{t} \tilde{x}Z(x) + T_{t}) (\partial_t R_{3,t} \cdot \tilde{x}Z(x) + \partial_t T_{3,t}) \right) \left( R_{3,t} \cdot \tilde{x}Z(x) + T_{3,t} \right)^2,
\]

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

\[
\left. \frac{\partial}{\partial t} w(x, z | g_t, Z) \right|_{t=0} = \frac{1}{d_1 x_1 Z(x) + d_2 x_2 Z(x) + d_3} \left[ (a_1 x_1^2 + a_2 x_1 x_2 + a_3 x_2^2) Z^2(x) + (b_1 x_1 + b_2 x_2 + b_3) Z(x) + c_1 \right]
\]

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 (8) contains both the sets

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

which establishes our claim that (8) is the set of smooth vector fields. \( \square \)

4 **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 [22]), 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).

### 4.1 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 6** (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 7.** 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 16 in Section 4.4.

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

**Definition 7** (F). A function \( f : \mathbb{R}^2 \to \mathbb{R}^+ \) is in class \( F \) if \( f \in F \) 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 8.** If \( f \in F \), 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 \( f \in F \) 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 \( F \) stable under small perturbations (e.g. noise in images); we will make this notion of stability more precise in Remark 13 in Section 4.4.

**Remark 9.** 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 \( L^\infty \) distance) to the boundary \( \partial([0,1] \times [0,1]) \). Define \( b : \mathbb{R} \to \mathbb{R} \) as

\[
b_c(x) = \begin{cases} 
\exp \left( -\frac{x^2}{2} \right) & x > 0 \\
x \exp \left( -\frac{x^2}{2} \right) & x < 0.
\end{cases}
\]
Then the extended function \( f : \mathbb{R}^2 \to \mathbb{R}^+ \) is
\[
f(x) = \begin{cases} 
  g(x) b_c(\text{dist}_c(x)) & x \text{ is inside } c \\
  b_c(-\text{dist}_c(x)) & x \text{ is outside } c 
\end{cases}
\]
where \( \text{dist}_c(x) \) is the distance from \( x \) to the curve \( c \).

Now consider the set of surfaces that are the graph of a function in \( \mathcal{F} \),
\[
\mathcal{S} = \{(x, f(x)) | x \in \mathbb{S}^2 \} \mid f \in \mathcal{F} \}.
\]
The set of monotonic continuous functions, also called contrast functions in \([6]\), is indicated by
\[
\mathcal{H} = \{ h \in C^2(\mathbb{R}^+; \mathbb{R}^+) \mid 0 < \frac{dh}{dt} < \infty, \ t \in \mathbb{R}^+ \}.
\]
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, \( \mathcal{S}' = \{ x \mid f(x) = \lambda, \lambda \in \mathbb{R}^+ \} \); it follows from Proposition 1 and Theorem 1 on page 11 of \([6]\) that the orbit space of surfaces \( \mathcal{S} \) modulo \( \mathcal{H} \) is given by their topographic map,
\[
\mathcal{S}' = \mathcal{S}/\mathcal{H}.
\]
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 \([6]\).

**Remark 10.** 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 \( \nabla f \), 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 \( \mathcal{F} \):
\[
\mathcal{W} = \{ w \in C^2(\mathbb{R}^2; \mathbb{R}^2) \mid \text{ a diffeomorphism} \} \cong \{ w \in C^2(\mathbb{S}^2; \mathbb{S}^2) \mid \text{ a diffeomorphism s.t. } w(\sigma) = \sigma, \sigma \text{ is the south pole} \}
\]
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
\[
\mathcal{S}'' = \mathcal{S}'/\mathcal{W} = \mathcal{S}/(\mathcal{H} \times \mathcal{W})
\]
of surfaces that are equivalent up to domain diffeomorphisms and contrast functions.

**Remark 11.** 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 12.** 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 13.** 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.
4.2 Reeb Graphs: Towards Viewpoint/Contrast Invariants

We now introduce Reeb graphs [10], 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 8 (Reeb Graph of a Function).** Let \( f : \mathbb{S}^2 \to \mathbb{R} \) be a continuous function. We define

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

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 \( \text{Reeb}(f) \) is connected.

**Proof.** \( \text{Reeb}(f) \) is the quotient space of \( \mathbb{S}^2 \) under the equivalence relation defined in Definition 8. Therefore, by definition we have a surjective continuous map \( \pi : \mathbb{S}^2 \to \text{Reeb}(f) \), and connectedness is preserved under continuous maps. \( \square \)

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

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

\[
\hat{\gamma}(t) = \begin{cases} 
\hat{\gamma}(t) & t \leq 1 \\
p(t) & t > 1
\end{cases}
\]

satisfies \( \hat{\gamma}(0) = \hat{\gamma}(2) \).

2. We show that \( \hat{\gamma} \) can be chosen so that it is continuous. We may assume that \( \gamma \) passes through the critical points of \( f \), \( \gamma(t_1), \ldots, \gamma(t_N) \) in that order. Thus, we divide the path \( \gamma \) into the sub-paths \( \gamma(0) \to \gamma(t_1) \), \( \gamma(t_1) \to \gamma(t_2) \), \ldots, \( \gamma(t_N, 1) \), 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 \( x_i \in f^{-1}(\gamma((t_i + t_{i+1})/2)) \subset \mathbb{S}^2 \). Then \( \hat{\gamma} \) in \((t_i, (t_i + t_{i+1})/2)\) is defined as the path solving

\[
y = \nabla f(y), \quad y(0) = x_i \in \mathbb{S}^2
\]

and in \((t_i + t_{i+1})/2, t_{i+1})\) as

\[
y = -\nabla f(y), \quad y(0) = x_i \in \mathbb{S}^2
\]

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 \( F(1, t) = \gamma(0) \). Then \( \pi \circ F \) is a retraction of \( \gamma \) 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 \( \text{Reeb}(f) \) is impossible. \( \square \)
4.3 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 9** (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 $a : V → L$ be a function (called the attribute function). We define the attributed graph as $AG = (V, E, L, a)$.

**Definition 10** (Attributed Reeb Tree of a Function). Let $f ∈ F$. Let $V$ be the set of critical points of $f$. Define $E$ to be

$$E = \{(v_i, v_j) : i \neq j, \exists \text{ a continuous map } \gamma : [0, 1] → \text{Reeb}(f) \text{ such that } \gamma(0) = [(v_i, f(v_i))], \gamma(1) = [(v_j, f(v_j))], \gamma(t) \neq [(v, f(v))] \text{ for all } v \in V \text{ and all } t \in (0, 1)\}. \quad (14)$$

Let $L = \mathbb{R}^+$, and

$$a(v) = f(v)$$

Note that the south pole $v_{sp} ∈ S^2$, is a critical point, and we include that in our definition. We define

$$\text{ART}(f) := (V, E, L, a, v_{sp}).$$

Note that the above definition encodes the type of critical point of each vertex $v ∈ V$.

**Definition 11** (Index of a Vertex of an Attributed Tree). Let $T = (V, E, \mathbb{R}^+, a)$ be an attributed tree, we define the map $\text{ind} : V → \{0, 1, 2\}$ as follows:

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

**Definition 12** (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, $φ : V_1 → V_2$, and the following properties are satisfied:

- if $a_1(v) > a_1(v')$ then $a_2(φ(v)) > a_2(φ(v'))$ for all $v, v' ∈ V_1$
- $φ(v_{sp, 1}) = v_{sp, 2}$

**Definition 13** (Degree of a Vertex). Let $G = (V, E)$ be a graph, and $v ∈ V$, then the degree of a vertex, $\text{deg}(v)$, is the number of edges that contain $v$.

**Definition 14** ($T$, a Collection of Attributed Trees). Let $T'$ 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 $v ∈ V$ and $\text{ind}(v) \neq 1$ then $\text{deg}(v) = 1$
3. If $v ∈ V$ and $\text{ind}(v) = 1$, then $\text{deg}(v) = 3$
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 $T$ to be the set $T'$ under the equivalence defined in Definition 12.

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 $\text{ART}(F) = T$. 

11
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).

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).

4.4 ART is the Maximal Viewpoint/Contrast Invariant

In this section, we show that $S'' = T$. 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:

Lemma 3 (Morse Lemma). If $f : S^2 \rightarrow \mathbb{R}$ 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 $\psi_i : \tilde{U}_i \subset \mathbb{R}^2 \rightarrow U_i \subset S^2$ so that

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

where $(\hat{x}, \hat{y}) = \psi_i(x, y)$ and $(x, y) \in S^2$ 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 $f \in F$, and ART$(f) = (V, E, L, a, v_{sp})$, then

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

12
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
\[
n_{0,2} - n_1 = 2 \quad \text{and} \quad n_{0,2} + n_1 - |E| = 1
\]
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,
\[
\sum_{v \in V} \deg(v) = 2|E| \quad \text{or} \quad n_{0,2} + \sum_{v \in V, \text{ind}(v)=1} \deg(v) = 2|E|,
\]
and combining with (15), we find that
\[
\sum_{v \in V, \text{ind}(v)=1} \deg(v) = 3n_1,
\]
but according to the Morse Lemma and the fact that critical points have distinct values (by definition of \( \mathcal{F} \)), \( \deg(v) > 2 \) if \( \text{ind}(v) = 1 \). These facts and (17) mean that \( \deg(v) = 3 \) if \( \text{ind}(v) = 1 \).

**Lemma 5** (Global Topology of Connected Level Sets). Let \( f \in \mathcal{F} \), and \( \pi_f : \mathbb{R}^2 \to \text{Reeb}(f) \) 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:

![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).](image)

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.

**Lemma 6.** Let \( f_1, f_2 \in \mathcal{F} \) and \( \text{ART}(f_1) \cong \text{ART}(f_2) \). Let \( \phi \) be a graph isomorphism between the trees in \( \text{ART}(f_1) \) and \( \text{ART}(f_2) \) satisfying Def. 12. 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 \( S^2 \)), which is equivalent to the point at infinity in \( \mathbb{R}^2 \). Because \( v_{sp} \) must be preserved by \( \phi \) (that is, the points at infinity...
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 $\phi'$. 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 or minima. Clearly, $\phi$ may not map $v_s$ to $v'_s$ if they are of different types. Now we reduce $\text{ART}(f_1)$ and $\text{ART}(f_2)$ 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 $\phi$ to $\text{ART}(f_1')$ and $\text{ART}(f_2')$ is a valid equivalence. But by the inductive hypothesis, $\phi$ does not map different types of saddles to each other. \hfill \Box

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

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

(18)

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

**Proof.** Let $\text{ART}(f_1) = (V_1, E_1, \mathbb{R}^+, a_1)$ and $\text{ART}(f_2) = (V_2, E_2, \mathbb{R}^+, a_2)$. 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 $\phi : V_1 \to V_2$. 

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure5}
\caption{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.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure6}
\caption{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.}
\end{figure}
Figure 7: If \( n = 1 \), then the \( \text{ART}(f) \) 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 \( \phi \).

\[
\begin{align*}
\text{Figure 8: Attributed Reeb trees of Type 1 (left) and Type 2 (right) saddles which are adjacent to two vertices that are not saddles.}
\end{align*}
\]

2. Using the Morse Lemma, there exist neighborhoods \( U_i, \tilde{U}_i \subset S^2 \) and diffeomorphisms \( w_i : U_i \rightarrow \tilde{U}_i \) 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 \( h_i : f_2(\tilde{U}_i) \rightarrow f_1(U_i), \) \( h_i(x) = f_1(p_i) - f_2(\tilde{p}_j) + x \). 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 \( w_i = \psi_i^{-1} \circ \psi_i \) where \( \psi_i \) 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 \( \pi_1 : S^2 \rightarrow \text{Reeb}(f_1) \) and \( \pi_2 : S^2 \rightarrow \text{Reeb}(f_2) \) be the natural quotient maps. For each \( p_i \) and \( \tilde{p}_i \) that correspond to minima or maxima (i.e., \( \text{ind}(p_i) = \text{ind}(\tilde{p}_i) \neq 1 \)), 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 \( \tilde{w}_i = w_i|_{W_i} \).

Now we consider each \( p_i \) that is a saddle point (i.e., \( \text{ind}(p_i) = 1 \)). 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 \( U_i = \psi_i^{-1}(B_{\varepsilon}(0)) \) and \( \tilde{U}_i = \tilde{\psi}_i^{-1}(B_{\varepsilon}(0)) \) for \( \varepsilon < 0 \) small and \( B \) denotes the disc in \( \mathbb{R}^2 \).

We now extend each \( w_i : U_i \rightarrow \tilde{U}_i \) to \( \tilde{w}_i : \tilde{W}_i \rightarrow \tilde{W}_i \) where

\[
\begin{align*}
W_i &= \bigcup_{q \in U_i \setminus \{p_i\}} \pi_1^{-1}([q, f_1(q)]) \\
\tilde{W}_i &= \bigcup_{q \in \tilde{U}_i \setminus \{\tilde{p}_i\}} \pi_2^{-1}([q, f_2(q)])
\end{align*}
\]

We define \( \tilde{w}_i \) as follows:

- Note that each \( \pi_1^{-1}([q, f_1(q)]) \) (\( q \in \tilde{U}_i \setminus \{\tilde{p}_i\} \)) 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 \( \pi^{-1}([q, f_1(q)]) \cap U_i \) consists of two connected components (the case of one connected component is done similarly). Let \( A, B, C, D \) denote points of \( \partial(\pi^{-1}([q, f_1(q)]) \cap U_i) \) 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 \( \pi^{-1}([q, f_1(q)]) \). Assume \( A \to B \) and \( C \to D \) specifies the parts of \( \pi^{-1}([q, f_1(q)]) \) 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 \( \pi^{-1}([q, f_1(q)]) \)) and \( \pi^{-1}([w_i(q), f_2(w_i(q))]) \). 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 \( \varphi \) in the intervals \([0, b]\) and \([c, d]\) as satisfying \( w_i(c_1(\xi)) = c_2(\varphi(\xi)) \).
- We define

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

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

\[
\int_b^c g(x) \, dx = \varphi(c) - \varphi(b), \quad g(b) = \varphi'(b), \quad g(c) = \varphi'(c)
\]

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 \( \hat{w}_i : W_i \to \hat{W}_i \) is a diffeomorphism because
  - \( \hat{w}_i|_{U_i} = w_i \) is a diffeomorphism by the previous step
- By Lemma 9, \( w_i \) does not map a type 1 saddle to a type 2 saddle and vice-versa, and so \( \tilde{w}_i | (W_i \setminus U_i) \) will be a diffeomorphism, details of which follow.
- \( \tilde{w}_i | (W_i \setminus U_i) \) is a diffeomorphism: for the region
  \[ \{ \pi_1^{-1}(q, f_1(q)) : q \in U_i \setminus \{ p_i \}, \pi_1^{-1}(q, f_1(q)) \cap U_i \text{ has 2 connected components} \} \]
  and (each connected component of) the region
  \[ \{ \pi_1^{-1}(q, f_1(q)) : q \in U_i \setminus \{ p_i \}, \pi_1^{-1}(q, f_1(q)) \cap U_i \text{ has 1 connected component} \} \]
  the parameterization of these regions by the family of \( c_1 \) and \( c_2 \) are differentiable, and so is the family of \( \phi \). Therefore, \( \tilde{w}_i \) is a differentiable as is its inverse.
- \( D\tilde{w}_i | \partial U_i = D\tilde{w}_i | \partial (W_i \setminus U_i) \): this is by construction of \( \phi \) in the previous step to be differentiable, and differentiable in its boundary conditions.

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

Let \( p_i \) and \( p_j \) be critical points of \( f_1 \) with corresponding vertices \( v_i, v_j \in V_i \) such that \( (v_i, v_j) \in E_1 \); also let \( p_i, p_j \) 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 \( \gamma_{ij} : [0, 1] \to \text{Reeb}(f_1) \) 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 \( \tilde{\gamma}_{ij} : [0, 1] \to \text{Reeb}(f_2) \) 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.\footnote{Note that a simple curve in \( S^2 \) does not define an inside and outside; however, we are identifying \( S^2 \) with \( \mathbb{R}^2 \) by specifying that the south pole of \( S^2 \) is mapped to infinity.}

We define \( \tilde{w}_{ij}, w_{ij} : X_{ij} \to \tilde{X}_{ij} \) as follows:
- We define \( \zeta_{ij} : \partial_{in} X_{ij} \times \mathbb{R}^+ \to S^2 \) and \( \tilde{\zeta}_{ij} : \partial_{in} \tilde{X}_{ij} \times \mathbb{R}^+ \to S^2 \) as
  \[
  \partial_{\zeta_{ij}}(x, t) = \pm \nabla f_1(\zeta_{ij}(x, t)), \quad \zeta_{ij}(x, 0) = x \in \partial_{in} X_{ij}
  \]
  \[
  \partial_{\tilde{\zeta}_{ij}}(x, t) = \pm \nabla f_2(\tilde{\zeta}_{ij}(x, t)), \quad \tilde{\zeta}_{ij}(x, 0) = x \in \partial_{in} \tilde{X}_{ij}
  \]
  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 \( \zeta_{ij}(\partial_{in} \tilde{X}_{ij}, [0, \tilde{T}]) = \tilde{X}_{ij} \). We define \( w_{ij} : X_{ij} \to \tilde{X}_{ij} \) as
  \[
  w_{ij}(\zeta_{ij}(x, t)) = \begin{cases} \tilde{\zeta}_{ij}(w_{ij}(x), h_{ij}(t)) & x \in \text{cl}(W_i) \\ \tilde{\zeta}_{ij}(w_{ij}(x), h_{ij}(t)) & x \in \text{cl}(W_j) \end{cases}, \text{ for } x \in \partial_{in} X_{ij}, t \in [0, T].
  \]
where \( h_{ij} : [0, T] \to [0, \tilde{T}] \) is chosen to be smooth, satisfies the conditions
\[
\begin{align*}
h_{ij}(0) &= 0, \quad h_{ij}(T) = \tilde{T}, \\
h'_{ij}(0) &= h'_{ij}(T) = h'_{ij}(f_2 \circ w_{ij}(\partial_{out} X_{ij})),
\end{align*}
\]
and is such that \( h : f_2(S^2) \to f_1(S^2) \) with the conditions
\[
\begin{align*}
h(f_1(\partial_{in} X_{ij})) &= f_2(\partial_{in} \tilde{X}_{ij}), \quad h'(f_1(\partial_{in} X_{ij})) = h'_{ij}(0) \\
h(f_1(\partial_{out} X_{ij})) &= f_2(\partial_{out} \tilde{X}_{ij}), \quad h'(f_1(\partial_{out} X_{ij})) = h'_{ij}(T) \\
h(v) &= h_i(v) \text{ for } v \in f_2(\tilde{U}_i)
\end{align*}
\]
is smooth. Note that \( h \) is the contrast change that we have been seeking in [18].
• It is clear that \(w_{ij} : X_{ij} \to \hat{X}_{ij}\) is a diffeomorphism; however it may not be the case that

\[
D\omega_{ij}\partial X_{ij}(x) = \begin{cases} 
D\omega_{i1}\partial W_i(x) & x \in \partial W_i \\
D\omega_{j2}\partial W_j(x) & x \in \partial W_j 
\end{cases}.
\tag{20}
\]

Indeed by Step 3, recall that we have

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

and so by differentiating, we have

\[\nabla f_1(x) = h_i(f_2 \circ w_i(x))D\omega_i(x) \cdot \nabla f_2(w_i(x)),\]

or

\[D\omega_i(x) \cdot \nabla f_1(x) = h_i(f_2 \circ w_i(x))D\omega_i(x)D\omega_i^T(x)\nabla f_2(w_i(x)).\tag{21}\]

Next by differentiating \([19]\), we have that

\[D\omega_{ij} \cdot \partial t\zeta_{ij}(x,t) = \partial t\tilde{\zeta}_{ij}(w_i(x), h_{ij}(t))h'_{ij}(t)\]

that is

\[D\omega_{ij} \cdot \nabla f_1(\zeta_{ij}(x,t)) = h'_{ij}(t)\nabla f_2(\tilde{\zeta}_{ij}(w_i(x), h_{ij}(t))).\]

In order to “adjust” \(w_{ij}\) so that \([20]\) holds, we define a new map \(\hat{w}_{ij}\) as follows. Let us abuse the notation and let \(\partial_{\hat{m}}X_{ij}, \partial_{\hat{m}}\hat{X}_{ij} : S^1 \to \mathbb{R}^2\) denote smooth parameterizations of the corresponding sets so that \(w_i(\partial_{\hat{m}}X_{ij}(u)) = \partial_{\hat{m}}\hat{X}_{ij}(u)\) for all \(u \in S^1\). Define \(c_1, c_2 : S^1 \times [0,1] \to \mathbb{R}^2\) as

\[c_1(u,v) = \zeta(\partial_{\hat{m}}X_{ij}(u), vT)\]

\[c_2(u,v) = \tilde{\zeta}(\partial_{\hat{m}}\hat{X}_{ij}(u), h(vT)).\]

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

\[\hat{w}_{ij}(c_1(u,v)) = c_2(\varphi(u,v), v)\tag{22}\]

satisfies \([20]\). Computing derivatives of \([22]\), 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 \(\varphi\) satisfies the conditions

\[\varphi(u,0) = u, \varphi(u,1) = u\]

\[A(u,v)(\nabla f_2(c_2(\varphi(u,v), v)))^\perp \varphi_v(u,v) + B(u,v)\nabla f_2(c_2(\varphi(u,v), v)) = \frac{1}{T}D\omega_i(c_1(u,v)) \cdot \nabla f_1(c_1(u,v))\]

where \(D\omega_i(c_1(u,v)) \cdot \nabla f_1(c_1(u,v))\) is specified in \([21]\). In other words, we must choose \(\varphi\) to satisfy the boundary conditions

\[\varphi(u,0) = u, \varphi(u,1) = u\]

\[\varphi_v(u,0) = E(u), \varphi_v(u,1) = F(u)\]
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.

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

$$\varphi_u > 0.$$  

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

$$\varphi_{uuu} + \varphi_{vvv} = 0.$$  

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

**Remark 14.** 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 15.** Condition 2 in Definition 7 ensures that ART$(f)$ does not change under small perturbations of $f$, e.g., $f + \epsilon g$ for small $\epsilon$. 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:

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

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

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

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 \( \phi \) 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 \( \phi \) only associates the group of coalesced saddles of \( f_1 \) to the group of coalesced saddles of \( f_2 \).

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

Proof. Let \( T' \in T' \) 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 \( S^2 \)) inward.

- Let \( R = \{ x \in \mathbb{R}^2 : |x| \leq 1 \}. \) We define \( f \) on \( \mathbb{R}^2 \setminus R \) 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 \( \partial R. \)
  - Otherwise, let \( v_1, v_2 \) be the two vertices adjacent to \( v \) that have not been visited.
    * Let \( R' \subset R \) such that \( cl(R') \subset R \) 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.\(^{13}\) Define \( f \) on \( R' \) to be diffeomorphically equivalent to a function that has level sets

\(^{13}\)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
Note

The function above is well-defined since any representative \( H \times W \) action of contrast and domain diffeomorphisms, them indistinguishable from other blobs, regardless of their shape. Could question the loss of discriminative power of the representation of ridges as “thin blobs” that renders generating a maximum along the ridge. The ART is stable with respect to such perturbations, although one the sense that a ridge with constant height can be turned into a Morse function by slightly perturbing it, thus

Non-isolated extrema such as ridges and valleys are also commonplace in images, but they are accidental in segment for instance due to occlusions and material boundaries. Therefore, the analysis above applies only to a general they are neither smooth nor have isolated extrema. Lack of smoothness is caused by discontinuities in the context of image analysis we always deal with surfaces that are graphs (the intensity values), but in surfaces that are not graphs of Morse functions.

Remark 16. 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.

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 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, \( F \), under the action of contrast and domain diffeomorphisms, \( H \times W \), is

\[
S'' = T
\]

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

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

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

- Lemma states that \( ART : S/(H \times W) \to T \) is injective.
- Lemma states that \( ART : F/(H \times W) \to T \) is surjective.
- Therefore, \( ART : S/(H \times W) \to T \) is a bijection and therefore, \( S/(H \times W) = T \).
5 Where is the “Information” in an image?

The traditional notion of information pioneered by Wiener and Shannon, and later Kolmogorov, quantifies the information content in the data as their “complexity” regardless of the use of the data. More specifically, the underlying “task” implicit in traditional Information Theory is that of reproducing an exact replica of the data after it has been corrupted by accidents, typically additive noise, when passing through a “channel”. In other words, Information Theory was built specifically for the task of “transmitting” or “compressing” data, rather than using it for recognition or inference.

But in the context of recognition, much of the complexity in the data is due to spurious factors, such as viewpoint, illumination and clutter. Following ideas of Gibson [9], the notion of “Actionable Information” has been proposed not as the complexity of the data itself, but as the complexity of the quotient of the data with respect to nuisance factors.

In the case of smooth regions of the image undergoing changes in contrast and viewpoint, considered in this manuscript, this means that the information content of the data is the complexity, or coding length, of the ART corresponding to the given region:

\[ I(f) = 6(#\text{max} + #\text{min}) - 7. \]  \hspace{1cm} (24)

Note that the above is the coding length of the ART, which would include codes for each minimum, maximum, saddle, their values, and the edge set. The number of maxima and minima completely determine the number of saddles (by the constraints imposed by the Betti numbers [14]), and edges (since ART is a tree). For color images, or images that are locally represented by more complex statistics than raw intensity, the above will have to include the coding length of the statistics stored at the nodes of the ART, but this is beyond our scope here. The information content \( I(f) \) measures the discriminative power of a smooth portion of an image. This pertains to portions of the image with smooth radiance; one would have to separately encode discontinuities (e.g. edges) and co-dimension one structures (e.g., ridges), to quantify the overall informative content of the image for the purpose of any task that requires contrast and viewpoint invariance.

6 Conclusion

In this manuscript we have focused on analyzing portions of the image that exhibit smooth shading or smooth texture statistics. 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), no salient blobs or ridges. 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 [8]. And yet, smoothly shaded regions convey a significant amount of “information,” however one wishes to define it. We have shown that

- It is possible to compute functions of an image region that exhibit smooth statistics that are invariant to both viewpoint and a coarse illumination model (contrast transformations), called ARTs.
- Such statistics 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 statistics have support on a set of measure zero of the image domain.
- The “information content” of an image for the purpose of recognition (as opposed to transmission) is given by the coding length of its associated ART. Such actionable information grows with the discriminative power of the representation, and measures the complexity of the data after the effect of nuisance factors, specifically viewpoint and contrast changes, is factored out.
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 pale resemblance 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.

Acknowledgments

We wish to thank Andrea Mennucci and Veeravalli Varadarajan for extended discussions, suggestions and comments at various stages of this project. This research was supported by AFOSR FA9550-06-1-0138 and ONR N00014-08-1-0414. A short version of this paper appeared in the Proceedings of the IEEE Intl. Conf. on Comp. Vis. and Patt. Recog. (CVPR), June 2009.

References


