On the Set of Images Modulo Viewpoint and Contrast Changes

G. Sundaramoorthi P. Petersen V. S. Varadarajan S. Soatto
University of California, Los Angeles
{ganeshs@cs, petersen@math, vsv@math, soatto@cs}.ucla.edu

Abstract

We consider regions of images that exhibit smooth statistics, and pose the question of characterizing the "essence" of these regions that matters for 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. We also show that such statistics are supported on a "thin" (zero-measure) 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 change of viewpoint and local illumination (contrast). Finally, we formalize the notion of "information" an image contains for the purpose of viewpoint- and illumination-invariant tasks, which we call "actionable information" following ideas of J. J. Gibson.

1. 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[1] and that represents a *sufficient statistic* for the task at hand.

[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".

The most common nuisances in recognition are (a) viewpoint, (b) illumination, (c) visibility artifacts such as occlusions and cast shadows, (d) quantization and noise.[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 an image occludes another cannot be determined from an image alone, but can be ascertained as part of the matching process with a training datum. What about the former two? *Can one devise image representations that are invariant to both viewpoint and illumination, at least away from visibility artifacts*[3] *such as occlusions and cast shadows?*

Viewpoint? Yes. Contrast? Yes. 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 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 [3] 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 [13]. 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 [5] that general-case (global) illumination invariants do not exist, even for Lambertian objects. However, there is a considerable body of literature 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 [1] that the geometry of the level curves (the iso-contours of the image), is contrast invariant, and therefore so is its dual, the gradient

[2]Note that we intend (a) and (b) to be absent of visibility artifacts, that are considered separately in (c).

[3]Visibility is addressed explicitly in [11].

direction.[4]

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! And 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 zero-measure subset of the image domain; finally, despite being thin, (iii) these invariants are sufficient statistics!

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.[5] *But what about regions with smooth statistics?* These would include shaded regions (Fig. 1) as well as texture gradients at scales

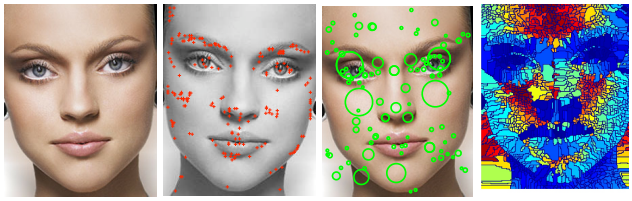

Figure 1. *Regions of an image that exhibit smooth texture gradient are not picked up by local feature detectors (Harris-affine, SIFT), and are over-segmented by most image segmentation algorithms. How do we "capture" the essence of these regions that matters for recognizing an object regardless of its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

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), a deterministic construction that is uniquely determined from each image.

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

5. We conclude that the ART is a viewpoint-illumination invariant, and that it has measure zero in the image domain.

6. Finally, we show that the ART is a sufficient statistic, in the sense that it is equivalent to the original image up to an arbitrary domain deformation and contrast change.[6]

7. We propose a notion of "actionable information" that measures the complexity *not* of the data, but of what remains of the data after the effect of the nuisances (viewpoint and illumination) is removed, i.e. the ART.

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 $\mathbb{C}^2$, 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, 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 face recognition [10] and shape coding [2] use a representation closely related to the ART.

In the next section, we introduce the mathematical tools that are necessary to characterize the set $\mathcal{S}''$ of viewpoint-illumination invariants. A summary of this section is provided in Sect. 1.2, for the reader who wishes to skip the mathematical details and proceed with the rest of the paper.

1.1. Mathematical Preliminaries (summary in 1.2)

For simplicity, we will represent a smooth portion of an image by a positive-valued Morse function on the *plane*. A *Morse function* $f : \mathbb{R}^2 \to \mathbb{R}^+; x \mapsto f(x)$ is a 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 x is where the Hessian is non-singular, $\det(\nabla^2 f(x)) \neq 0$. Morse are dense in $\mathbb{L}^2$, and therefore can approximate edges, ridges and other discontinuities in the image arbitrarily well. We introduce the following subset of Morse functions that have

[4] 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).

[5] 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.

[6] Note that this does not necessarily mean that a viewpoint-illumination invariant is a unique signature for an object. As [13] 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 [13].

distinct critical values and where all the "structure" is concentrated, to avoid having to deal with critical points that escape outside the domain of the image.

Definition 1 ($\mathcal{F}$) *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*

If $f \in \mathcal{F}$, then we may identify f with a Morse function on the sphere $\tilde{f} : \mathbb{S}^2 \to \mathbb{R}^+$ via the inverse stereographic projection from the north 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, 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 more precise this notion of stability later. Now consider the set of surfaces that are the graph of a function in $\mathcal{F}$:

$$\mathcal{S} \doteq \{\{(x, f(x)) | x \in \mathbb{S}^2\} \mid f \in \mathcal{F}\}. \quad (1)$$

The set of monotonically increasing continuous functions, also called *contrast functions* in [4], is indicated by

$$\mathcal{H} \doteq \{h : \mathbb{R}^+ \to \mathbb{R}^+ \mid 0 < \frac{dh}{dt} < \infty,\ t \in \mathbb{R}^+\}. \quad (2)$$

Contrast functions form a group, 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 [4] that the quotient of the surfaces $\mathcal{S}$ modulo $\mathcal{H}$ is given by their topographic map,

$$\boxed{\mathcal{S}' = \mathcal{S}/\mathcal{H}.} \quad (3)$$

In other words, the topographic map is a sufficient statistic of the surface that is invariant to contrast changes; 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*.

Remark 1 *In the context of image analysis, where the domain of the image is rectangular (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 used 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 deformations in $\mathcal{F}$:

$$\mathcal{W} \doteq \{w : \mathbb{S}^2 \to \mathbb{S}^2 \text{ a diffeomorphism} : w(\sigma) = \sigma\} \quad (4)$$

where σ denotes the south pole. $\mathcal{W}$ is also 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 equivalence classes $[f] = \{h \circ f \circ w \mid h \in \mathcal{H}, w \in \mathcal{W}\}$. **The goal of this manuscript is to characterize** these equivalence classes, i.e., the orbit space

$$\boxed{\mathcal{S}'' \doteq \mathcal{S}'/\mathcal{W} = \mathcal{S}/\{\mathcal{H} \times \mathcal{W}\}} \quad (5)$$

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

Remark 2 *In the context of image analysis, domain diffeomorphisms model changes of viewpoint [13] away from visibility artifacts such as occlusions.[3] Therefore, the quotient above – if it is found to be non-trivial – can be considered to be a sufficient statistic of the image that is invariant to viewpoint and illumination.*

We now give a series of definitions that are introduced to elucidate the structure of the orbit space (5).

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

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

where $(y, f(y)) \in [(x, f(x))]$ iff $f(x) = f(y)$ and there is a continuous path from x to y 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 encoding of the function value of the level set).

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 2. Therefore, by definition we have a surjective continuous map $\pi : \mathbb{S}^2 \to Reeb(f)$, and connectedness is preserved under a continuous map. □

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. To do so, 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. 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 $\pi \circ \hat{\gamma} = \gamma$. This path is constructed by solving the gradient flow $\dot{y} = \nabla f(y)$ between critical points. 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, 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. □

Definition 3 (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 4 (Attributed Reeb Tree (ART)) *Let $f \in \mathcal{F}$. Let V be the set of critical points of f. Define E to be*

$$E = \{(v_i, v_j) : i \neq j, \exists\, \gamma : [0,1] \to Reeb(f)$$

continuous such that $\gamma(0) = v_i$, $\gamma(1) = v_j$ and $\gamma(t) \neq [(v,\ f(v))]\ \forall t \in (0,1), v \in V\}$. 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

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

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

Definition 5 (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.*

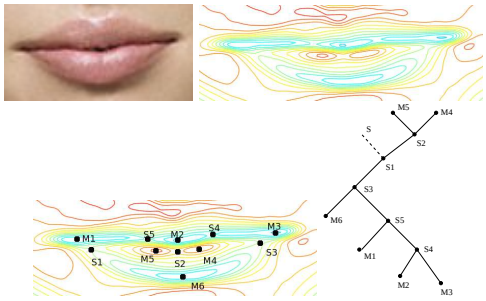

Figure 2. *The lip region of Fig. 1, its level lines, the level lines marked with extrema, and a graphical depiction of ART (note that the height of the vertex is proportional to the attribute value).*

Definition 6 (Equivalence Class 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 7 (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 8 ($\mathcal{T}$, 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 connected 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. *If $v_1, v_2 \in V$ and $ind(v_1) \neq 1$ and $ind(v_2) \neq 1$ and $\{v_1, v_2\} \neq V$ then $(v_1, v_2) \notin E$*
5. *$n_0 - n_1 + n_2 = 2$ 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 6.

1.2. Synopsis of the previous section

We summarize the relevant concepts introduced thus far that are necessary to proceed with the rest of the paper. We have started by assuming that a smooth portion of the image can be approximated with a Morse function extended to the plane and then mapped to the sphere. This can always be

done up to an arbitrarily small error. Then we have introduced the Reeb Graph for a general surface, and shown that in the case of the intensity surface of an image it reduces to a tree. The construction of the Attributed Reeb Tree (ART) is illustrated in Fig. 2: The extrema are detected, and their label (maximum, minimum, saddle) retained together with the ordering of their values, but not the values themselves. Then extrema that correspond to nested level sets are linked by an edge. Each point on the edge represents a level set, but – unlike the Reeb Graph used in [10] – its value is not stored, and is instead discarded. This construction is conceptual, and in practice one would want to devise a detector that analyzes the image at multiple scales to locate extrema in a manner that is robust to noise and quantization artifacts. Shinagawa has proposed such a procedure in [10].

In order to support the establishment of correspondence between two attributed trees, we have also introduced the notion of "equivalence" between two attributed trees if the nodes of one map to the nodes of the other, and they have corresponding labels. A subset of attributed trees with specific properties and under this equivalence relation has been called $\mathcal{T}$.

2. ART Is a Viewpoint-Illumination Invariant Sufficient Statistic

The set of attributed trees modulo the equivalence relation in Def. 6, which we called $\mathcal{T}$, is the object we have been looking for. In the rest of this section we will show that $\mathcal{S}'' = \mathcal{T}$. It follows immediately from the definitions given in the previous section that $ART(f)$ 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. It is far less immediate to see whether the Attributed Reeb Tree is a sufficient statistic, or that it is equivalent to the surface that generated it up to a domain diffeomorphism. We start by stating a fact from Morse theory that we exploit in our argument:

Lemma 3 (Morse) *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 $\psi_i : \bar{U}_i \subset \mathbb{R}^2 \to U_i \subset \mathbb{S}^2$ so that*

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

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

The image around any extremum can be locally warped into one of the three canonical forms of Fig. 3. We now move to the core part of our argument:

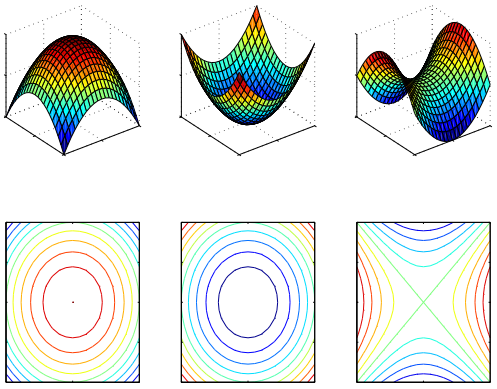

Figure 3. *The Morse Lemma states that a neighborhood of a critical point of a Morse function looks like one of the three forms (left to right: maximum, minimum, and saddle).*

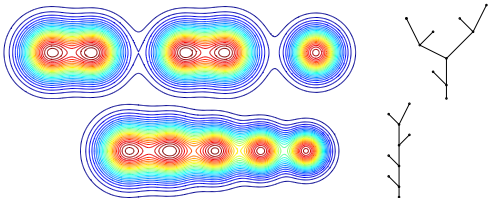

Figure 4. *This figure shows the importance of the structure of the Reeb tree in determining whether two functions are in the same equivalence class. The figure shows the level sets of two functions and their corresponding Reeb trees. In this case, each function has the same number of min/max/saddles, and values, but the Reeb trees are different and the functions are not equivalent via a viewpoint/illumination change.*

Lemma 4 *Let f_1, $f_2 \in \mathcal{F}$ be functions that generate two (image) surfaces. Then*

$$ART(f_1) \cong ART(f_2) \Leftrightarrow \exists\, h \in \mathcal{H}, w \in \mathcal{W} \mid f_2 = h \circ f_1 \circ w. \tag{6}$$

So two ARTs are equivalent if, and only if, the images that generated them are related by a domain diffeomorphism, which is equivalent to a change of viewpoint per [13], and by a contrast transformation, which is a local approximation of an illumination change per [1]. Note that the diffeomorphism w and contrast function h are not necessarily unique. See Appendix A for a sketch and [12] for the complete proof.

Remark 3 *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 (Fig. 4).*

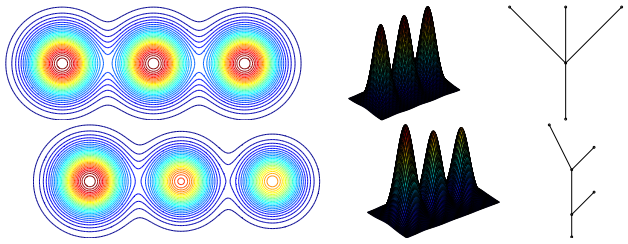

Figure 5. *Top: A Morse function (its level sets, surface, and attributed Reeb tree, respectively) of a function with multiple saddles on the same connected component of a level set. Bottom: a slightly perturbed version of the above Morse function. The attributed Reeb tree of the function on the top is not stable under small perturbations; while the one on the bottom is stable.*

Remark 4 *Condition 2 in Definition 1 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 digital images where quantization artifacts and sensor noise can introduce spurious extrema. To demonstrate this point, consider the following function with two saddle points that have the same 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 5. 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 5. 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.*

Furthermore, 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 the 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 attributed Reeb tree. The graph isomorphism ϕ in the proof of Lemma 4 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.

Lemma 5 *For each $T \in \mathcal{T}$, there exists a Morse function $f \in \mathcal{F}$ so that $ART(f) = T$.*

Proof. Let T' be an embedding of the tree T in $\mathbb{R}^3$ such that T' lies in the $x-z$ plane and that also respects the ordering of T. Thicken T' in $\mathbb{R}^3$: $S = \partial\left(\bigcup_{p\in T'} B_\epsilon(p)\right)$ where $B_\epsilon(p)$ is the ball of radius ϵ centered at p. Note that S is diffeomorphic to $\mathbb{S}^2$. Choose f to be the height function (i.e., the z coordinate of the surface), which is a Morse function so that $ART(f) = T$. □

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

Theorem 1 *The attributed Reeb tree of a surface uniquely determines it up to a contrast change and domain diffeomorphism. Equivalently, the quotient of surfaces that are graphs of Morse functions modulo contrast and domain deformations is*

$$\boxed{\mathcal{S}'' = \mathcal{T}} \tag{7}$$

3. 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 [7], we propose to quantify "actionable information" in an image *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 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:*

$$\mathcal{I}(f) = 6(\#max + \#min) - 7. \tag{8}$$

Note that the above is the coding length of the ART, which would include codes for each minima, maxima, saddle, their values, and the edge set. The number of maxima and minima completely determines the number of saddles (by the constraints imposed by the Betti numbers [8]), and edges (since ART is a tree). The case of occlusion is addressed in [11].

The information content $\mathcal{I}(f)$ measures the discriminative power of a portion of an image. To see this, consider a recognition problem where a test image is given that either contains a specific object ($\omega = 1$) or not ($\omega = 0$). Assume that $P(\omega)$, the probability of the event ω, is given, for instance equal to $1/2$. Let $f \in \mathcal{F}$ be a test image, and consider the decision function (classifier) $\alpha : \mathcal{F} \to \{0,1\}$ and

a loss function $\lambda : \{0,1\}^2 \to \mathbb{R}^+$, for instance the standard 0-1 loss $\lambda(\alpha_i, \omega_j) = \delta_{ij}$. Ideally, we want to find the function α that minimizes the conditional risk

$$R(\alpha|f) \doteq \sum_j \lambda(\alpha|\omega_j) P(\omega_j|f) \tag{9}$$

for any choice of f. The conditional risk can be used as a discriminant function, and it can be shown that this choice minimizes the expected risk $R(\alpha) \doteq \int R(\alpha|f) dP(f)$. We say that a statistic $\phi : \mathcal{F} \to F$ is *sufficient* for the particular decision represented by the expected risk $R(\cdot)$ if

$$R(\alpha) = R(\alpha \circ \phi). \tag{10}$$

Note that, in general, $R(\alpha) \leq R(\alpha \circ \phi)$, that is, we cannot "create information by manipulating the data." If we wish to compute the optimal decision function using a training set $\mathcal{D} = \{(\omega_i, f_i)\}_{i=1,\dots N}$, using Bayes' rule we can express the discriminant $R(\alpha|f)$ in terms of the likelihood $p(f|\omega, \mathcal{D})$. If we isolate the role of the nuisance factors h (contrast) and w (viewpoint), we have that

$$p(f|\omega, \mathcal{D}) = \int p(f|\omega, h, w, \mathcal{D}) dP(h, w) \tag{11}$$

where the measure $dP(\cdot)$ is degenerate (uninformative) and therefore it does not depend on the training set. Nevertheless, the training set is necessary in order to perform the above marginalization and "learn away" the nuisance variables.

If, on the other hand, we consider the modified decision problem where the data f is "pre-processed" to obtain $ART = \phi(f)$, then to minimize $\tilde{R}(\tilde{\alpha}|f) \doteq R(\alpha|\phi \circ f)$ we must compute

$$\begin{aligned} p(\phi \circ f|\omega, \mathcal{D}) &= \int p(ART|\omega, h, w, \mathcal{D}) dP(h, w) = \\ &= \int p(ART|\omega, \mathcal{D}) dP(h, w) = p(ART|\omega). \end{aligned} \tag{12}$$

In other words, by using ART instead of the raw data f we can significantly reduce the complexity of the classifier, including reducing the size of the training set to one sample,[7] while at the same time keeping the conditional risk unchanged. The classifier $\alpha \circ \phi$, following the invariance properties of ϕ, is also called *equivariant*, and it can be shown to achieve the optimal (Bayesian) risk [9].

Now, if we restrict the classifier to only use a subset of the ART of a given complexity K, we have a nested chain of classifiers $\tilde{R}_K(\tilde{\alpha}|f) \doteq R(\alpha|\phi \circ f; \mathcal{I}(f) \leq K)$,

$$\tilde{R}_{K+1} \leq \tilde{R}_K \tag{13}$$

and therefore the discriminative power of the statistic $\phi \circ f$ increases monotonically with the actionable information content $\mathcal{I}(f)$ of the ART.

[7] If one considers a categorization problem, where the object of interest exhibits intrinsic variability, the training set is still necessary in the right hand-side of (12), but it is no longer needed to "learn away" the extrinsic variability.

4. Discussion

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 [6]. And yet, smoothly shaded regions convey a significant amount of "information," however one wishes to define it. But how do we define information, and how can we quantify it? We have shown that

- It is possible to compute functions of an image region that exhibits 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, see footnote 6) 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 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; they can be turned into a Morse function by an infinitesimal perturbation. The Reeb graph 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. Finally, 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.

A. Proof of Lemma 4

Proof. We give an outline of the proof, details are in [12]. 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 prove the forward direction in steps:

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 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
$$f_2|U_i = h_i \circ f_1 \circ w_i|U_i$$
for some contrast change $h_i : \mathbb{R}^+ \to \mathbb{R}^+$. We may assume that $\{U_i\}$ are disjoint as are $\{\tilde{U}_i\}$.
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. We extend each $w_i : U_i \to \tilde{U}_i$ to $\hat{w}_i : W_i \to \tilde{W}_i$ where
$$\tilde{W}_i = \bigcup_{q \in \tilde{U}_i \setminus \{\tilde{p}_i\}} \pi_1^{-1}([q, f_1(q)])$$
$$W_i = \bigcup_{q \in U_i \setminus \{p_i\}} \pi_2^{-1}([q, f_2(q)])$$
as follows: $\hat{w}_i(\pi_2^{-1}([q, f_2(q)])) = \pi_1^{-1}([w_i(q), f_1(w_i(q))])$ where $q \in U_i \setminus \{\tilde{p}_i\}$ and $\hat{w}_i|\pi_2^{-1}([q, f_2(q)])$ extends $w_i|\pi_2^{-1}([q, f_2(q)]) \cap \tilde{U}_i$ via a diffeomorphism of the circle.
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 $\gamma_{ij}(1) = [(p_j, f_2(p_j))]$. 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
$$X_{ij} = \pi_1^{-1}(\gamma_{ij}([0,1])) \setminus (W_i \cup W_j)$$
$$\tilde{X}_{ij} = \pi_2^{-1}(\tilde{\gamma}_{ij}([0,1])) \setminus (\tilde{W}_i \cup \tilde{W}_j).$$
   We define $w_{ij} : X_{ij} \to \tilde{X}_{ij}$ so that the following hold:
   - Let $h_{ij} : f_1(X_{ij}) \to f_2(\tilde{X}_{ij})$ where $f_1(X_{ij}), f_2(\tilde{X}_{ij}) \subset \mathbb{R}$ be a diffeomorphism.
   - $w_{ij}(f_1^{-1}(\alpha) \cap X_{ij}) = f_2^{-1}(h_{ij}(\alpha)) \cap \tilde{X}_{ij}$ where $\alpha \in f_1(X_{ij})$
   - For each $\alpha \in f_1(X_{ij})$, $w_{ij}|f_1^{-1}(\alpha) \cap X_{ij}$ is a diffeomorphism of the circle so that $w_{ij} : X_{ij} \to \tilde{X}_{ij}$ is a diffeomorphism.
   - $w_{ij}|\text{cl}(X_{ij}) \cap \text{cl}(W_i) = \hat{w}_i|\text{cl}(X_{ij}) \cap \text{cl}(W_i)$ and $w_{ij}|\text{cl}(X_{ij}) \cap \text{cl}(W_j) = \hat{w}_j|\text{cl}(X_{ij}) \cap \text{cl}(W_j)$ where cl denotes closure. Further $Dw_{ij}(x) = D\hat{w}_i(x)$ for $x \in \text{cl}(X_{ij}) \cap \text{cl}(W_i)$.

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

□

Acknowledgments

We wish to thank Andrea Mennucci and Andrea Vedaldi for extended discussions, suggestions and comments at various stages of this project. We are grateful to an anonymous reviewer (R3) for useful and detailed comments, to William McEneany for pointing out necessary corrections, and Pierre Deligne for discussions through correspondence. This research was supported by AFOSR FA9550-06-1-0138 and ONR N00014-08-1-0414. This paper is dedicated to Donald L. Snyder's mother, who taught him to "never throw away *information*." *Data*, on the other hand ...

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] S. H. Baloch, H. Krim, I. Kogan, and D. Zenkov. Rotation invariant topology coding of 2D and 3D objects using Morse theory. In *Proc. of the IEEE ICIP*, 2005.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

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

[8] J. Milnor. *Morse Theory*. Princeton University Press, 1969.

[9] J. Shao. *Mathematical Statistics*. Springer Verlag, 1998.

[10] 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.

[11] S. Soatto. Actionable Information in Vision. *Technical Report UCLA CSD090007*, March 10, 2009.

[12] G. Sundaramoorthi, P. Petersen, and S. Soatto. On the set of images modulo viewpoint and contrast changes. Technical Report *UCLA-CSD090005*, February 2009.

[13] 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.