Multiscale Geometric Image Processing

Justin K. Romberg, Michael B. Wakin, and Richard G. Baraniuk

Dept. of ECE, Rice University, Houston, Texas

ABSTRACT

Since their introduction a little more than 10 years ago, wavelets have revolutionized image processing. Wavelet based algorithms define the state-of-the-art for applications including image coding (JPEG2000), restoration, and segmentation. Despite their success, wavelets have significant shortcomings in their treatment of edges. Wavelets do not parsimoniously capture even the simplest geometrical structure in images, and wavelet based processing algorithms often produce images with ringing around the edges.

As a first step towards accounting for this structure, we will show how to explicitly capture the geometric regularity of contours in cartoon images using the wedgelet representation and a multiscale geometry model. The wedgelet representation builds up an image out of simple piecewise constant functions with linear discontinuities. We will show how the geometry model, by putting a joint distribution on the orientations of the linear discontinuities, allows us to weigh several factors when choosing the wedgelet representation: the error between the representation and the original image, the parsimony of the representation, and whether the wedgelets in the representation form "natural" geometrical structures. Finally, we will analyze a simple wedgelet coder based on these principles, and show that it has optimal asymptotic performance for simple cartoon images.

Keywords: Wedgelets, geometry in images, image compression

1. INTRODUCTION

At the core of image processing lies the problem of modeling *image structure*. Building an accurate, tractable mathematical characterization that distinguishes a "real-world photograph-like" image from an arbitrary set of data is fundamental to any image processing algorithm.

A vital part of this characterization is the *image representation*. Using an atomic decomposition, we can closely approximate the image $X(s)$ as a linear combination

$$X(s) = \sum_i \alpha_i b_i(s) + \epsilon(s) \quad \{b_i\} \subset B \tag{1}$$

of atoms b_i from a dictionary B. A good representation will essentially lower the dimensionality of the modeling problem — we want B such that every image of interest can be well approximated ($\epsilon(s)$ small) using relatively few b_i. We can then design an *image model* that tells us which configurations of atoms (choices of $\{\alpha_i\}$ and $\{b_i\} \subset B$) to favor; the model specifies the *relationships* between the atoms in the representation caused by structure in the image.

We are now faced with the two central questions: "What kinds of structure exist in images?" and "What is a representation+model that captures this structure"? Defining the class of N-pixel "images" as a subset of $\mathbb{R}^N$ exactly is an ill-posed problem, even on a philosophical level. But there are two prevalent types of structure present in images that any general-purpose model should account for: images contain smooth, homogeneous regions and these regions are separated by smooth contours. That is, images exhibit *grayscale regularity* in smooth regions and *geometric regularity* along edge contours.

The wavelet transform[1] has emerged as the preeminent tool for image modeling. The success of wavelets is due to the fact that they provide a *sparse* representation for smooth signals interrupted by isolated discontinuities[2] (this is a perfect model for "image slices" — 1D cross sections of a 2D image). Simple, tree-based models

E-mail: jrom@rice.edu, wakin@rice.edu, richb@rice.edu
Web: dsp.rice.edu

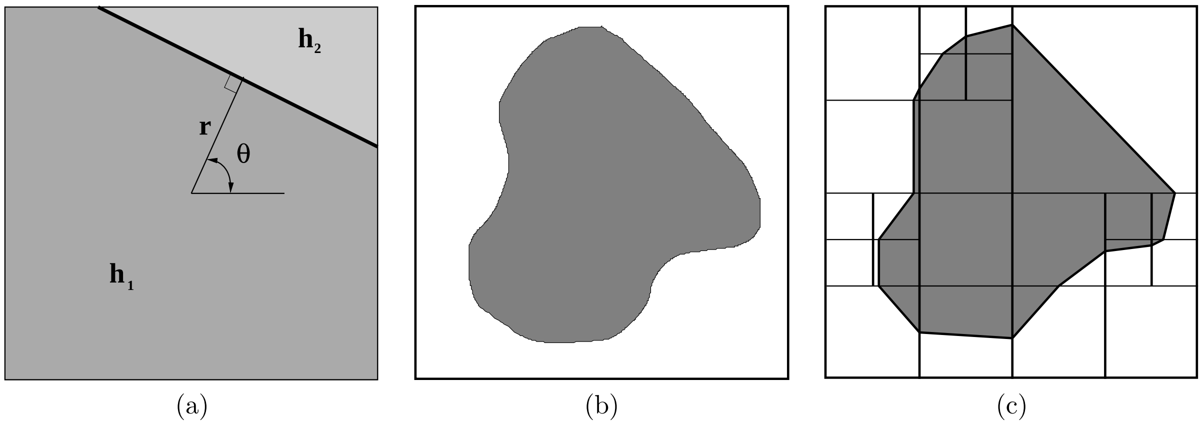


Figure 1. (a) A wedgelet in a dyadic block is parameterized by an angle θ, an offset r, and heights h_1, h_2. (b) Example of a "simple cartoon image", (c) The wedgelet representation divides the domain of the image into dyadic squares, using a piecewise constant function in each square to approximate the image.

for wavelet coefficients[3–5] have defined the state-of-the-art. These models leverage the principle that wavelet coefficients representing smooth regions in the image tend to have small magnitude, while those representing edges have relatively large magnitudes. Arranging the coefficients in a tree (binary tree in 1D, quadtree in 2D), this property allows us to *prune* the wavelet tree over smooth regions, while growing it over edges. For 1D image slices, tree-pruning model based algorithms enjoy not only practical success in denoising and compression, but are also theoretically optimal.[6,7]

The success of wavelets does not extend to 2D images. Although wavelet-based images processing algorithms define the state-of-the-art, they have significant shortcomings in their treatment of edge structure. No matter how smooth a contour is, large wavelet coefficients cluster around the contours, their number increasing at fine scales. The wavelet transform is not sparse for our simple class of images made up of smooth regions separated by smooth contours[8]; it simply takes too many wavelet basis functions to accurately build up an edge. In short, the wavelet transform makes it easy to model grayscale regularity, but not geometric regularity.

There are two paths we can take: we can either develop more complicated wavelet models, abandoning the simplicity and computational efficiency of tree pruning, or we can develop new representations for which simple models will suffice. The second approach has recently become a major focus for the field of Computational Harmonic Analysis. Several new representations have been proposed (see, for example, "curvelets" in,[8–10] "bandelets" in,[11] and "wedgelets" in[12]) with nice theoretical sparsity properties, but have yet to be used as part of a modeling framework for practical applications.

This paper will concentrate on modeling the geometric regularity in images using the simplest of these new decompositions, the *wedgelet representation*.[12] To start, we will consider *simple cartoon images*, the class of piecewise constant images with flat regions separated by smooth (twice differentiable) contours (see Figure 1). These images are purely geometrical in the sense that the only thing of interest in the image is the contour structure. Studying the simple cartoon class isolates the problem of decomposing and capturing the behavior of edge contours; a final solution will incorporate both kinds of regularity in the model (see[13] for combining the two).

The wedgelet representation, introduced in,[12] is natural for simple cartoon images. A *wedgelet* is a piecewise constant function on a dyadic square S that is discontinuous along a line through S with *orientation* (r, θ), see Figure 1(a) for an illustration. A *wedgelet representation* of an image X consists of a dyadic partition of the domain of X along with a wedgelet function in each dyadic square, see Figure 1(c). The wedgelet representation can be thought of as a *decorated quadtree*; a quadtree dictates how to divide the domain into dyadic squares, and attached to each leaf of the quadtree is a wedgelet telling us how to approximate the image in that local region.

The wedgelet representation is sparse for the class of simple cartoon images[12] in the sense that a close approximation to X can be built out of relatively few wedgelets. In fact,[12] shows that using a sufficiently large wedgelet dictionary, the wedgelet representations $\mathcal{W}$ found by solving the optimization problem

$$\min_{\mathcal{W}} \|X - \mathcal{W}\|_2^2 + \lambda^2 |\mathcal{W}| \tag{2}$$

have the optimal rate of error decay, that in (1) as $|\{b_i\}| = |\mathcal{W}| \to \infty$ ($\lambda \to 0$) we have $\|\epsilon(s)\|_2^2 \to 0$ like $|\mathcal{W}|^{-2}$. Solving (2) can be done efficiently using a simple "tree pruning" dynamic program similar to the one used for optimal wavelet tree pruning in practical image coders.[5]

In Section 3, we present a *multiscale geometry model* that captures the joint behavior of the wedgelet orientations along smooth contours. The model assigns a probability $P(\mathcal{W})$ to every possible set of wedgelet orientations, assigning high probability to configurations that are "geometrically faithful", and low probability to those that are not. The model can be used as a *prior distribution* on the wedgelet orientations, and can be used as such in statistical image processing applications such as coding, denoising, and detection. In this paper, we will be particularly concerned with coding wedgelet representations; the algorithms developed here are applied in a novel image coder in.[13]

Instead of considering only wedgelet orientations at the leaves of the quadtree, the multiscale geometry model (MGM) associates an orientation from the wedgelet dictionary to each node in the quadtree that describes the general linear behavior of the image in the corresponding dyadic square. Transition probabilities model the progression of orientations from parent to child in the quadtree; the MGM is a Markov-1 process with the orientations as states. For smooth contours, we expect little geometric innovation from parent to child at fine scales, so child orientations that are close to the parent (in (r, θ) space) are favored (assigned higher probabilities) over those that deviate significantly.

In Section 3, we show how to use the MGM to select a wedgelet representation $\mathcal{W}$ for an image X that balances the fidelity $\|X - \mathcal{W}\|_2^2$ against the complexity $-\log_2 P(\mathcal{W})$

$$\min_{\mathcal{W}} \|X - \mathcal{W}\|_2^2 + \lambda^2[-\log_2 P(\mathcal{W})]. \tag{3}$$

The Markov structure of the MGM allows us to solve (3) with an efficient dynamic program. While simple tree pruning allows us to exploit the fact that smooth contours can be closely approximated by lines at coarse scales, the more sophisticated model-based approach will also take advantage of the wedgelet orientations "lining up" between dyadic blocks. Since $-\log_2 P(\mathcal{W})$ corresponds to the number of bits* it would take an ideal coder to code the orientations in the wedgelet representation, the representation is optimal in the *rate-distortion* sense.

Section 4 analyzes the asymptotic rate-distortion performance of a coder using a simplified version of the MGM. We measure how quickly the distortion $D(R)$ (the error between the image X and the coded wedgelet representation) decays to zero as the rate (the number of bits we spend coding the wedgelet representation) goes to infinity. In,[15] the asymptotic rate-distortion performance of the "oracle" coder (where the contour is known to the encoder) for an image consisting of a C^2 contour separating two flat regions was derived as $D(R) \sim 1/R^2$. In Section 4, we construct a coder with the same asymptotic performance, suggesting that the multiscale Markov-1 wedgelet model is sufficient to capture the geometrical structure of smooth contours.

2. WEDGELET REPRESENTATIONS

2.1. Wedgelet Basics

A *wedgelet* is a piecewise constant function on a square domain S that is discontinuous along a line $\ell_{r,\theta}$ passing through S. Four parameters specify a wedgelet on S: an angle θ and a normalized offset r for the *orientation* of $\ell_{r,\theta}$, and heights h_1, h_2 for the two values the wedgelet takes on either side of $\ell_{r,\theta}$ (see Fig. 1 for an illustration). The wedgelet function on S with parameters (r, θ, h_1, h_2) is denoted by $w(S; r, \theta, h_1, h_2)$.

A *wedgelet representation* of an image X is an approximation of X build out of wedgelets on dyadic squares. Choosing a wedgelet representation $\mathcal{W}$ requires (a) a choice of dyadic partition $\mathcal{P}$ of the domain of X and (b) for each dyadic square $S \in \mathcal{P}$, a choice of wedgelet (we will include the "null" wedgelet, $w(S; \cdot)$ = constant among

*$-\log P(\mathcal{W})$ is the "Shannon codelength" of $\mathcal{W}$[14]

our choices); $\mathcal{W} = \{\mathcal{P}, \ \ w(S; r, \theta, h_1, h_2) \ \ \forall S \in \mathcal{P}\}$. As an example, a wedgelet representation of a simple image is shown in Fig. 1.

We will choose the wedgelet orientation (r, θ) from a finite dictionary $\mathcal{D}_S$, possibly different for each dyadic square. The orientation of $\ell_{r,\theta}$, defines a two-dimensional linear space of wedgelet functions, a *wedgespace*, that are discontinuous along $\ell_{r,\theta}$. The (h_1, h_2) corresponding to the closest point on the wedgespace to the image $X(S)$ restricted to S is a simple projection, calculated by averaging $X(S)$ over the subregions of S on either side of $\ell_{r,\theta}$. For the majority of the paper, we will determine the heights in this manner, and denote the resulting wedgelet $w_{r,\theta}$ (there is one unique wedgelet for each orientation).

There are two steps in finding a wedgelet representation for an image. The *analysis* step tests each orientation $(r, \theta)_i \in \mathcal{D}_S$ in every dyadic square S, measuring the fit of the corresponding wedgelet with the L^2 error $||X(S) - w_{(r,\theta)_i}||_2$. The *inference* step then uses this information to choose a dyadic partition and a set of wedgelet orientations.

2.2. Fidelity vs. Complexity

The wedgelet representation is sparse for simple cartoon images[12, 16]; it takes relatively few wedgelets (a coarse partition) to form an accurate approximation. Given an image X, we can find such a representation by solving a complexity regularized optimization problem:

$$\min_{\mathcal{W}} \|X - \mathcal{W}\|_2^2 + \lambda^2 |\mathcal{W}|. \tag{4}$$

Solutions to (4) are found efficiently using the "CART" algorithm[†], a simple dynamic program for optimal tree pruning.

Equation (4) can be thought of in more general terms as a *fidelity vs. complexity* optimization problem.

$$\min_{\mathcal{W}} \ \mathrm{Error}(\mathcal{W}, X) + \lambda^2 \mathrm{Complexity}(\mathcal{W}) \tag{5}$$

where the constant λ allows us to control the importance of small error versus low complexity. The goal is to find a wedgelet representation that is accurate but is also easily described. Implicit in the choice of complexity term ($|\mathcal{W}|$ in (4)) is a *model*; certain wedgelet configurations are being favored (assigned lower complexity) over others.

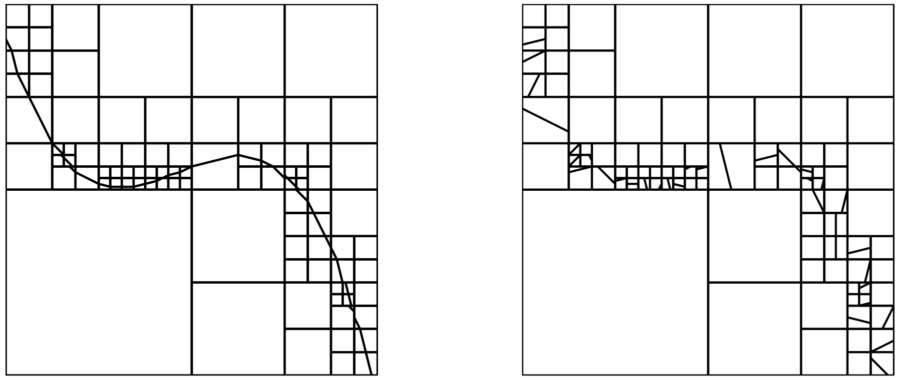
In the context of image coding, we would like the complexity of a wedgelet representation to be (at least an approximation of) the actual number of bits it will take to code the representation. Equation (5) then becomes a rate-distortion optimization problem.[18] The codelengths associated with each wedgelet representation depend on an underlying probability model $P(\mathcal{W})$ (this connection is made theoretically through the Kraft inequality[14] and practically through arithmetic coding[19]). Using the ideal Shannon codelength as the complexity measure, and the L^2 error metric, equation (5) becomes

$$\min_{\mathcal{W}} \ ||X - \mathcal{W}||_2^2 + \lambda^2[-\log P(\mathcal{W})]. \tag{6}$$

The wedgelet representation probability model $P(\mathcal{W})$ should reflect the behavior of contours found in images (be "geometrically faithful"), allow fast solutions to the optimization problem (6), and yield efficient coding algorithms.

The complexity term $|\mathcal{W}|$ in (4) depends only on the partition size; the wedgelet orientations are independent given a partition and as a result, all possible sets of wedgelet orientations in the dyadic squares are treated equally. As the example in Figure 2 shows, independence does not take full advantage of the contour structure in the images. Given a partition, there are very few joint sets of wedgelet orientations that are geometrically coherent, and these should be favored very highly over the many geometrically incoherent orientation sets. In Section 3, we present a more sophisticated geometrical modeling framework that captures the *dependencies* between wedgelet orientations chosen in each dyadic square.

[†] "CART" stands for "Classification and Regression Trees", the context in which[17] dynamic programs of the form (4) are heavily used.



(a) (b)

Figure 2. Two sets of wedgelet orientations for the same partition. We want a geometrical model that favors "geometrically coherent" joint sets of orientations like those in (a) over incoherent sets like those in (b).

3. WEDGELET MODELING AND INFERENCE

Since edge contours in images are smooth (geometricly regular), wedgelet representations approximating these contours will be structured. In this section, we show how to capture this structure with a multiscale geometry model.

As discussed in Section 2.2, using the number of wedgelets as a complexity penalty is equivalent to a model only on the size of the partition used for the wedgelet representation — given that partition, all sets of orientations on the leaves are treated equally. As the example in Figure 2 shows, independence is not a sufficient model to capture the geometrical regularity of contours in image.

Geometrical regularity, like grayscale regularity, is readily described by its multiscale behavior. In a sense, both smooth regions and smooth contours become "uninteresting" as we zoom in on them. At fine resolutions, smooth regions are essentially flat and contours are essentially straight. Wavelet models capture grayscale regularity by expecting wavelet coefficients representing smooth regions to be close to zero at fines scales. The multiscale geometry model will favor sets of wedgelet orientations that remain close to a straight line.

The MGM is based on the same multiresolution ideas as the 2D wavelet transform. When we use wavelets to represent an image, the image is projected onto a series of approximation spaces at different scales[2] (see Figure 3(a)). As the scales get finer, the projection onto the scaling space moves closer to the actual image. The wavelet coefficients specify the *difference* between the images projected onto successive scaling spaces; a model on the wavelet coefficients is a model on how the grayscale values are changing as the approximation becomes finer. At fine scales, we expect small wavelet coefficients in regions where the image is smooth.

We can also approximate geometrical structure in the image using a series of scales. At each scale j, the domain of the image is partitioned into dyadic squares with sidelength 2^{-j}. In each of these squares, we use a line (from a finite wedgelet dictionary) to approximate the contour as it passes through (see Figure 3(b)). The orientations can be organized on a quadtree, with each dyadic square corresponding to a node; each "parent" node at scale j splits into four "children" nodes at scale $j+1$. Along smooth contours at fine scales, we expect the deviations of the lines in the child squares from the line in the parent square to be small.

The MGM models the interscale behavior of these lines using a transition matrix. An entry $A_{p,q}$ in this matrix specifies the probability of an orientation $\ell_c \in \mathcal{D}_{S_c}$ in the child dyadic square S_c given an orientation $\ell_\rho \in \mathcal{D}_{S_\rho}$ in parent dyadic block S_ρ

$$A_{p,q} = P(\ell_c = q | \ell_\rho = p). \tag{7}$$

The probabilities are assigned based on the Hausdorff distance between the lines defined by ℓ_ρ and ℓ_c restricted to the child square S_c.

The Markov-1 structure of the MGM means that the probability of a given set of wedgelet orientations (the leaves of the geometry quadtree above) can be calculated using the upwards-downwards algorithm.[20] We can

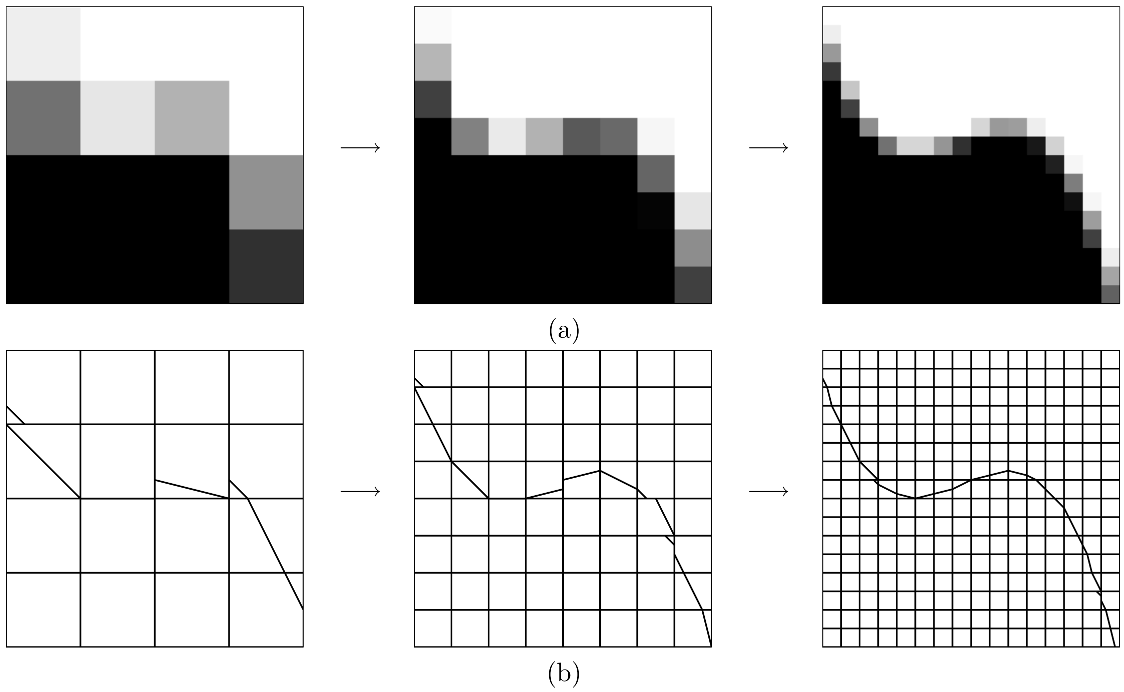


Figure 3. Just like (a) grayscale values are refined as we project onto finer scaling spaces, (b) geometry can be refined by using a denser dyadic partition.

also use a slightly modified version of the Viterbi algorithm to choose a wedgelet representation for a given image by solving the fidelity v. complexity optimization problem (3).

4. WEDGELET CODING ASYMPTOTICS

In this section, we will analyze the asymptotic rate-distortion performance of a simple, multiscale wedgelet coder. We will consider binary images on $[0,1]^2$ with one contour, parameterized by a unit-speed C^2 curve $c(t)$ with bounded curvature, passing through the domain. We will examine how fast the distortion function $D(R)$ (the squared error between the wedgelet approximation and the actual image) goes to zero as $R \to \infty$.

The performance for an "oracle coder" (i. e. the encoder has knowledge of the contour $c(t)$) is $D(R) \sim 1/R^2$.[15] In,[15] a simple coding algorithm that treats the wedgelet orientations independently given a partition is shown to have asymptotic performance $D(R) \sim (\log R)^2/R^2$. In this section, we show that by incorporating multiscale dependencies between the wedgelet orientations, we can remove the $(\log R)^2$ term and get the oracle performance.

We will only sketch the proof here, a full exposition can be found in.[21] The crucial idea is that because the curve is smooth (C^2 with bounded curvature), given a parent orientation, only certain child orientations are possible.

1. Because the curve $c(t)$ is C^2 with bounded curvature, it can be locally approximated using straight lines. Specifically, let S be a dyadic square with sidelength 2^{-j}. There exists a line ℓ through S such that $c(t)$ is contained in a strip of width $C_\omega 2^{-2j}$ inside of S, where C_ω is a constant depending only on the maximum curvature of $c(t)$ (see Figure 4).

2. At each scale j, we will use an orientation from a wedgelet dictionary $\mathcal{D}_j$ with r and θ values spaced at $\sim 2^{-j}$ in each direction. We have $|\mathcal{D}_j| \sim 2^{2j}$.

3. With this dictionary, we can revise 1 to use a quantized set of lines. There exists a line ℓ_q with orientation $(r,\theta)_q \in \mathcal{D}_j$ such that $c(t)$ is contained in a strip of width $C_{q\omega} 2^{-2j}$.

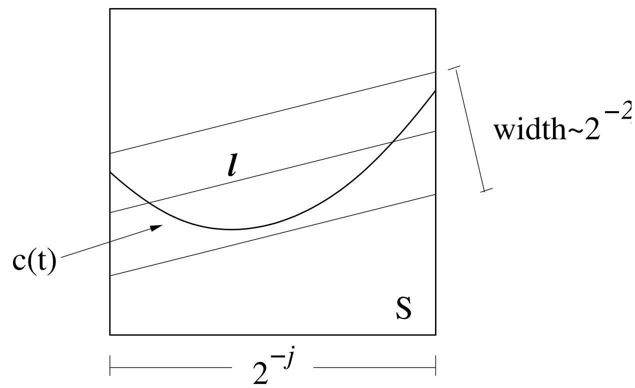


Figure 4. Since the contour $c(t)$ is smooth, inside of dyadic square S with sidelength 2^{-j} it can be bounded by a strip of width $\sim 2^{-2j}$ around a line ℓ.

4. By 3, using a wedgelet with the same orientation as ℓ_q to approximate the image in S gives us a local error of $\sim 2^{-3j}$.

5. For a given maximum scale J, the curve passes through $\sim 2^J$ dyadic squares. As such, using wedgelets at scale J to approximate the image results in a global error of $D_J \sim 2^{-2J}$.

6. We will code the orientations of all the lines chosen as in 3 in each dyadic square that the curve passes through from scales $j = 0 \ldots J$.

7. Given a parent orientation ℓ_ρ at scale j, we know that the curve doesn't leave a strip of width $C_{q\omega}2^{-2j}$. This means that the child orientation has to be within a slightly larger strip of width $C_{c\rho}2^{-2j}$. Because of this, we can put bounds on the deviation of the child orientation parameters from the parent parameters: $|r_c - r_\rho| \leq C_r 2^{-j}$ and $|\theta_c - \theta_\rho| \leq C_\theta 2^{-j}$. Since the size of the dictionary is growing like 2^j in each direction as j increases, the number of orientations in this region of possibilities can be bounded by a constant. Thus, even though the dictionary is growing as the scale becomes finer, it takes a constant number of bits to code each child given its parent.

8. We will need to code $\sim 2^J$ orientations overall. As such, we will spend $R_J \sim 2^J$ bits coding the wedgelet orientations.

9. Having established that $D_J \sim 2^{-2J}$ and $R_J \sim 2^J$, we can conclude that $D(R) \sim 1/R^2$.

These arguments show that for a given image, there exists a set of wedgelet orientations with the oracle rate-distortion performance. To find these orientations, we can use the dynamic program discussed in Section 3.

REFERENCES

1. I. Daubechies, *Ten Lectures on Wavelets*, SIAM, New York, 1992.
2. S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, 2nd ed., 1998.
3. J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing* **41**, pp. 3445–3462, Dec. 1993.
4. M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Processing* **46**, pp. 886–902, Apr. 1998.
5. Z. Xiong, K. Ramchandran, and M. T. Orchard, "Space-frequency quantization for wavelet image coding," *IEEE Trans. Image Processing* **6**(5), pp. 677–693, 1997.
6. A. Cohen, W. Dahmen, I. Daubechies, and R. DeVore, "Tree approximation and optimal encoding," *J. Appl. Comp. Harm. Anal.* **11**, pp. 192–226, 2001.

7. A. Cohen, I. Daubechies, O. G. Guleryuz, and M. T. Orchard, "On the importance of combining wavelet-based nonlinear approximation with coding strategies," *IEEE Trans. Inform. Theory* **48**, pp. 1895–1921, July 2002.
8. E. Candés and D. Donoho, "Curvelets – a surpringly effective nonadaptive representation for objects with edges," in *Curve and Surface Fitting*, A. C. et. al, ed., Vanderbilt University Press, 1999.
9. E. Candés and D. Donoho, "New tight frames of curvelets and optimal representations of objects with C^2 singularities," *Technical Report, Stanford University* , 2002.
10. M. N. Do, *Directional Multiresolution Image Representations.* PhD thesis, École Polytechnique Fédéral de Lausanne, October 2001.
11. E. L. Pennec and S. Mallat, "Image compression with geometrical wavelets," in *IEEE Int. Conf. on Image Proc. – ICIP '01*, (Thessaloniki, Greece), Oct. 2001.
12. D. Donoho, "Wedgelets: Nearly-minimax estimation of edges," *Annals of Stat.* **27**, pp. 859–897, 1999.
13. M. B. Wakin, J. K. Romberg, H. Choi, and R. G. Baraniuk, "Rate-distortion optimized image compression using wedgelets," in *IEEE Int. Conf. on Image Proc. – ICIP '02*, (Rochester, NY), September 2002.
14. T. Cover and J. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.
15. M. N. Do, P. L. Dragotti, R. Shukla, and M. Vetterli, "On the compression of two-dimensional piecewise smooth functions," in *IEEE Int. Conf. on Image Proc. – ICIP '01*, (Thessaloniki, Greece), Oct. 2001.
16. X. Huo, *Sparse Image Representation via Combined Transforms.* PhD thesis, Stanford University, 1999.
17. L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*, Wadsworth, Belmont, CA, 1984.
18. A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Processing Mag.* **15**, pp. 23–50, Nov. 1998.
19. I. H. Witten, R. M. Neal, and J. G. Cleary, "Arithmetic coding for data compression," *Comm. of the ACM* **30**, pp. 520–540, June 1987.
20. L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE* **77**, pp. 257–285, Feb. 1989.
21. J. K. Romberg, *Multiscale Geometric Image Processing.* PhD thesis, Rice University, 2003.