Courant Institute TR2005-875
Nonlinear Image Representation via Local Multiscale Orientation

David K. Hammond
Courant Institute of Mathematical Sciences
hammond@cims.nyu.edu

Eero P. Simoncelli
Howard Hughes Medical Institute,
Center for Neural Science, and
Courant Institute of Mathematical Sciences
eero@cns.nyu.edu

September 27, 2005

Abstract

We present a nonlinear image representation based on multiscale local orientation measurements. Specifically, an image is first decomposed using a two-orientation steerable pyramid, a tight-frame representation in which the basis functions are directional derivatives of a radially symmetric blurring operator. The pair of subbands at each scale are thus gradients of progressively blurred copies of the original image. We then discard the magnitude information and retain only the orientation of each gradient vector. We develop a method for reconstructing the original image from this orientation information using an algorithm based on projection onto convex sets, and demonstrate its robustness to quantization.

1 Introduction

Traditional linear multiscale representations describe an image using a sum of translated and scaled copies of a fixed set of basis functions. These decompositions, commonly known as "wavelets", are in widespread use throughout the field of image processing. But despite their scale invariance and other useful properties, they do not provide an explicit representation of the geometric structures (e.g., edges) found in images, because these geometric features are nonlinearly related to the pixel intensities.

The representation of local image geometry is currently a topic of active research. Some authors have proposed sparse overcomplete expansions that

approximate the local geometry well [e.g., 1]. The approach of Taubman and Zakhor [2], and the more recent bandlet approach of Pennec and Mallat construct a new local basis by resampling the image adaptively according to the local orientation [3]. Mallat and Zhong [4] showed that an image could be reconstructed with reasonable accuracy from knowledge of the locations of multi-scale zero-crossings. The wedgelet scheme [5, 6] represents image as using step edges parameterized by intensity value and orientation. Li has developed explicit representations of the local phase structure around edges [7]. A direct representation of images in terms of edges was proposed by Elder, who extracted the orientation, slope and position of edges and showed that this information was sufficient to reconstruct the original image [8].

In this paper, we develop a representation based on direct representation of the underlying image orientation. First, we decompose the image using a steerable pyramid, which is an overcomplete tight frame based on multi-scale derivative operators. This provides a direct representation of the local gradient of the image at different scales. We then discard the gradient magnitudes, and retain only their orientations. Surprisingly, this orientation information alone is sufficient to reconstruct the original image, and we develop an iterative algorithm for reconstruction using alternating projections onto convex sets. We also develop several modifications for increasing the rate of convergence.

2 Steerable Pyramid Decomposition

The steerable pyramid (SP) is linear multiscale filter bank whose filters are derivative operators [9]. The SP filter bank of order K with S scales consists of a highpass filter, K+1 oriented filters at each of the S scales and a lowpass residual filter. The SP is overcomplete by a factor of 1+4K/3. In practice these bands are computed through a cascade of filtering and subsampling operations but the filter design may be seen more clearly by disregarding this at first. The SP filters are designed to be polar separable in the Fourier domain, to prevent spatial aliasing in the subsampling of the subbands, and to form a tight frame.

The effects of these design constraints can be seen most clearly in the frequency domain. Each bandpass filter operation is equivalent to convolution with a radially symmetric bandpass operator followed by convolution with a K'th-order differentiator in one of K+1 equally spaced directions. As the signals live on a discrete lattice, we must specify precisely what we mean by differentiation. For a continuous function, differentiation in the x direction follows the following rule in the Fourier domain:

$$\widehat{\frac{d^K f(x,y)}{dx^K}} = (iw_x)^K \hat{f}\,(w_x, w_y)\,.$$

The oriented SP filters are designed by using this property to define discrete differentiation. In the Fourier domain, the filters can be written in polar coordinates as

$$\hat{B}_{s,k}(r,\theta) = \frac{g_s(r)}{r^K}(ir\cos(\theta - \frac{k\pi}{K+1}))^K,$$

with $1 \leq s \leq S$ and $0 \leq k \leq K$. The radial functions are given by

$$g_s(r) = g(2^{s-1}r)$$

for a "mother" function g(r), which may be chosen so that the filters form a tight frame. The tight frame property implies that the sum of the filters composed with their conjugate transposes yields a multiple of the identity operator. In the Fourier domain this constraint yields the tiling condition

$$\sum_{s=1}^{S} g(2^{s-1}r)^2 + g_{low}(r)^2 + g_{high}(r)^2 = 1,$$

where $g_{low}(r)$ and $g_{high}(r)$ are the radial components of the Fourier transforms of the lowpass and highpass filters. In this work all filtering is done in the frequency domain, which implicitly implies circular boundary handling.

3 Separation of Orientation and Magnitude

For an image of dimension MxN, the output of the first order (K=1) SP consists of the highpass, oriented bandpass and lowpass subbands

$$\begin{aligned}
H(m,n) &: 1 \leq n \leq N, 1 \leq m \leq M \\
B_{s,x}(m,n) &: 1 \leq n \leq N/2^{s-1}, 1 \leq m \leq M/2^{s-1}, 1 \leq s \leq S \\
B_{s,y}(m,n) &: 1 \leq n \leq N/2^{s-1}, 1 \leq m \leq M/2^{s-1}, 1 \leq s \leq S \\
L(m,n) &: 1 \leq n \leq N/2^{S}, 1 \leq m \leq M/2^{S}.
\end{aligned}$$

Orientation and magnitude bands can be computed from the oriented bandpass bands by transforming to polar coordinates at each location and scale

$$\begin{aligned}
M_s(m,n) &= \sqrt{B_{s,x}(m,n)^2 + B_{s,y}(m,n)^2} \\
\Theta_s(m,n) &= \arctan(B_{s,y}(m,n)/B_{s,x}(m,n)).
\end{aligned}$$

These transformations divide the information contained in the gradient bands into two parts. One may then ask whether the magnitude or orientation bands are more important for representing the image. This is similar in spirit to the classic work of Oppenheim & Lim who demonstrated that Fourier phases are much more important than Fourier magnitudes for representing image structure [10].

Wundrich et. al. have shown that images may be represented by the local magnitudes (i.e. discarding the phase) of a doubly-overcomplete Gabor representation [11]. However, in the case of multi-scale gradient measurements, and analogous to Oppenheim and Lim's result, we find that the visible image structure is essentially determined by the orientation. As one illustration of this, one may take two images and build the first order SP and extract the orientation

and magnitude bands from both images, swap them, and then resynthesize two new images. As is shown in Fig. 1, the synthesized images are more similar to the image from which the orientation information was obtained. The magnitude information of the other image is seen to have a local contrast modulation effect, but the orientation information is clearly critical for forming features like edges.

Another interesting demonstration of the relative importance of magnitude and orientation information is provided by combining valid orientation or magnitude bands with a randomly generated counterpart. This is illustrated in Fig. 2. The random orientation bands were sampled from a uniform distribution over $[0, 2\pi]$, while the random magnitude bands were formed by resampling from the histogram of the original magnitude band. The lack of features in the random orientation image demonstrates the importance of coherent orientations for representing image structure.

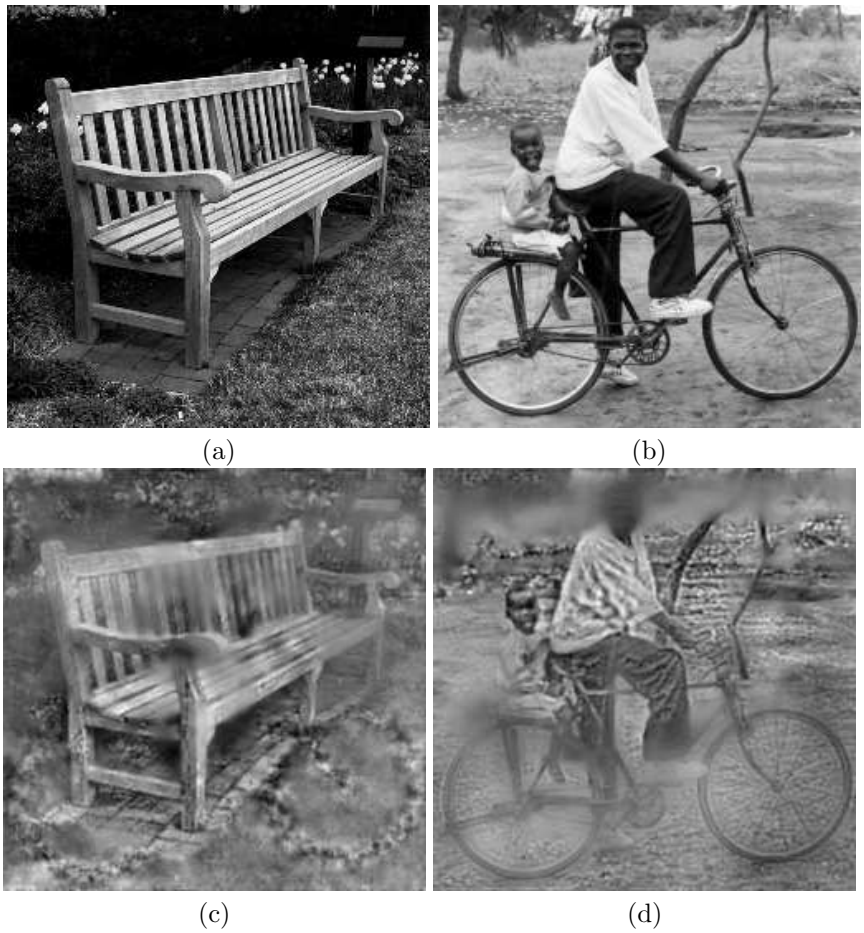
4 Reconstruction Algorithm

The image in Fig. 2(b) was synthesized by combining the original orientation bands and random magnitude bands into a pyramid and then inverting the pyramid decomposition. If this new image is re-analyzed using the SP filters, its orientation band will generally not be the same as the original orientation, because of the overcompleteness of the SP. But one may then reimpose the original orientation information and again reconstruct the image. Surprisingly, iterative application of this simple algorithm converges to the original image!

In order to explain this, it is helpful to first introduce some additional notation. Let $Im = R^{MxN}$ be the space of image pixels, and W be the space of SP coefficients. W is a linear space and can be thought of the direct product of the separate linear spaces L, $B_{x,s}$, $B_{y,s}$ and H for $1 \leq s \leq S$. Let $A : Im \rightarrow W$ denote the analysis operator, ie, the linear transformation corresponding to building the steerable pyramid representation. As the SP is a tight frame, $A^\dagger : W \rightarrow Im$ satisfies $A^\dagger \circ A = I_{Im}$ where I_{Im} is the identity operator on the space Im. Applying $A^\dagger$ is equivalent to inverting the steerable pyramid representation.

As the SP is overcomplete, i.e. $\dim(W) > \dim(Im)$, the composition $A \circ A^\dagger : W \rightarrow W$ is not equal to the identity. The tight frame property implies that $A \circ A^\dagger$ is an orthogonal projection onto the set $W_{Im} = A(Im)$, the "image under A of the image space". As W_{Im} is a linear space, it is convex. This fact, that the process of reconstructing an image followed by rebuilding the pyramid is an orthogonal projection onto a convex subset of W, is crucial to understanding the reconstruction algorithm.

The orientation representation consists of orientation bands at 5 scales, the highpass band and the lowpass band. The reconstruction algorithm consists of iteratively building the steerable pyramid, imposing the specified orientation bands, highpass and lowpass bands, and transforming back into the image domain. A single iteration takes about 6 seconds for a 512x512 size image, running in Matlab on a 2.4 GHz Pentium III machine. In the coefficient domain, denote



(a) (b)

(c) (d)

Figure 1: Swapping orientation and magnitude. Image (c), formed from orientation of (a) and magnitude of (b). Image (d), formed from orientation of (b) and magnitude of (a). Highpass and lowpass bands were set to zero for synthesized images.

(a) (b) (c)

Figure 2: (a) Original image. (b) Image synthesized with orientations from (a) and random magnitudes. (c) Image synthesized with magnitudes from (a) and random orientations.

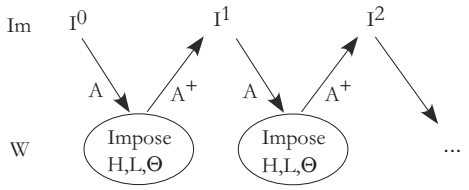


Figure 3: Schematic of reconstruction algorithm.

by W_Θ the set of all pyramid coefficients that have the correct orientation bands, highpass band and lowpass band. This set is convex. It consists of a cartesian product of rays at each location in scale and space.

Viewed from the pyramid coefficient domain, the reconstruction algorithm is performing alternating projection onto the convex sets W_Θ and W_{Im}. If we denote the operator $O_\Theta : W \rightarrow W$ the operation of imposing the orientation, highpass and lowpass bands, a single iteration of the reconstruction algorithm (starting and ending in W) is equivalent to applying $A \circ A^\dagger \circ O_\theta$. One natural implementation of O_Θ would be to simply replace the current orientations by the desired orientations, this is equivalent to 'rotating' each pair of x and y pyramid coefficients at each location and scale to lie along the ray of specified orientation. An alternative form for O_Θ would be as an orthogonal projection to W_Θ. These are illustrated in Fig. 4(a)

If the orthogonal projection form of O_Θ is used, then the reconstruction algorithm is performing alternating projection onto convex sets. Such an algorithm will always converge to an intersection point of the two sets, provided such a point exists. As the orientation bands being imposed are taken from an actual image, $W_\Theta \cap W_{Im}$ must contain at least one point, namely the steerable pyramid representation of the original image! The reconstruction algorithm is thus guaranteed to converge to something. In practice, convergence to the original image has always been observed.

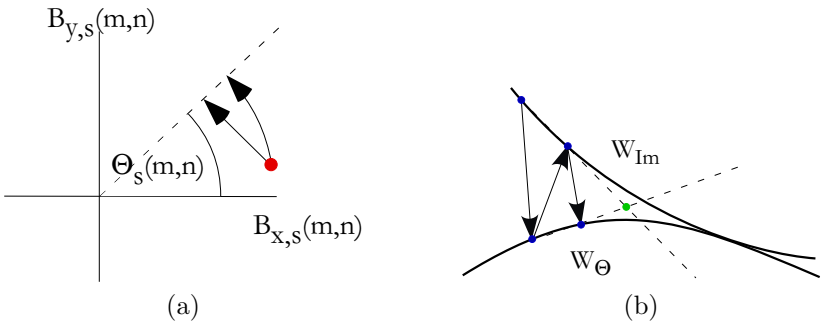


Figure 4: (a) Orientation imposition. The axes represent the x and y coordinates of a gradient vector at a particular position and scale. The dotted line represents a slice of W_Θ, the set of coefficients having the correct orientation. This orientation may be imposed by rotating or orthogonally projecting onto this line. (b) Acceleration by extrapolation. The first four points indicate successive states of the algorithm, extrapolation along the dotted lines gives the last point. See text.

5 Speeding up Convergence

Our iterative reconstruction algorithm converges, but quite slowly. For a typical 512x512 image, beginning with random noise as a starting point, the basic algorithm takes about 100 iterations before the result is perceptually indistinguishable from the original image. This is dues to the fact that the amplitudes are encoded only implicitly in the representation, and their recovery results from the interactions between orientations at different positions and scales as well as the lowpass band. We found the process can be accelerated by including and imposing the total power (sum of squared magnitudes) at each scale by rescaling each band by a single scalar.

Another technique for accelerating the convergence is illustrated in Fig. 4(b). Two successive iterations of the algorithm correspond to projecting onto four points in W, alternatively projecting onto W_Θ and W_{Im}. We extrapolate along the two lines formed by these four points (which may not intersect in the high dimensional space W!) and find the points where the two lines are closest. We take the average of these two points as the new extrapolated point, and continue the algorithm. The combination of imposing power at each scale and extrapolating yields a much faster reconstruction algorithm that requires approximately 25 iterations for a 512x512 image, starting from a random noise image, before the resulting image is perceptually indistinguishable from the original.

6 Quantization of Orientations

The success of the alternating projection algorithm as described above relies on the existence of a point of intersection of W_{Im} and W_Θ. While this is

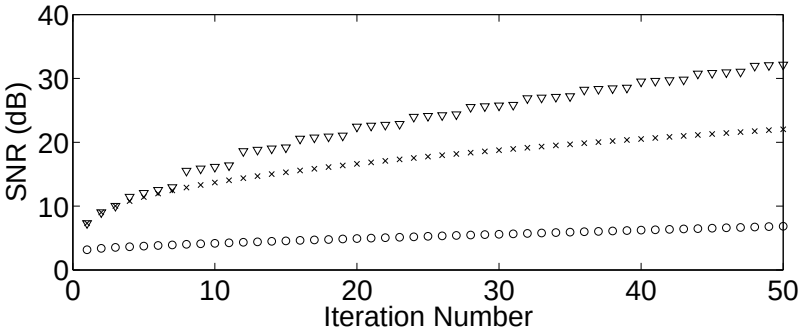
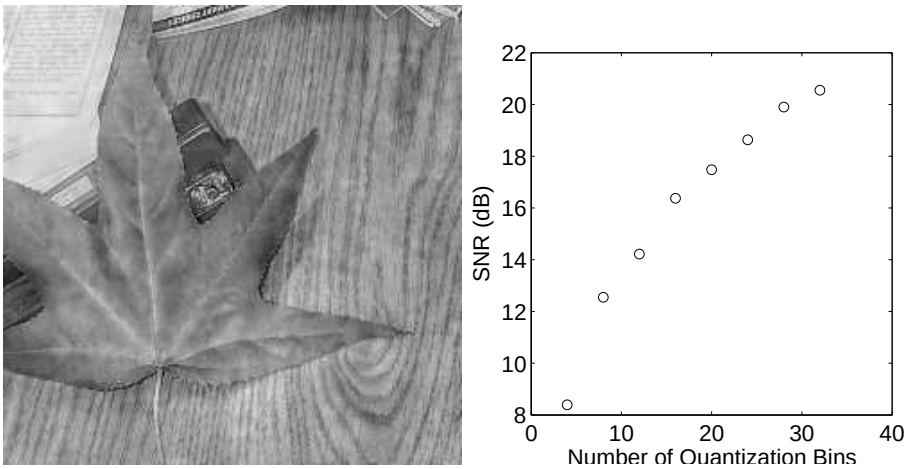


Figure 5: Convergence. The SNR of the n'th iterate relative to the original image is plotted against iteration number for the basic algorithm (∘), algorithm with energy imposed at each scale (x), and algorithm with both energy imposition and extrapolation (▽).

ensured for sets of orientations that arise from actual images, it may not be true for sets of orientations that have been altered in some way. This could be troubling for using this representation for image processing tasks in which one would manipulate the orientations and then recover the processed image. If the representation were not stable to perturbations in the orientations, this would undermine its potential utility. As a method of exploring this, we have studied the effects of quantization of the orientations, and found the representation to be well behaved.

For quantization to Q values, the unit circle was divided into Q equal bins of width $\frac{2\pi}{Q}$ with bin centers at $\frac{2\pi q}{Q}$ for $q = 0...Q-1$. The quantized orientation bands $\Theta_s^Q(m,n)$ were formed by replacing the value of $\Theta_s(m,n)$ by the angle of the center of its corresponding bin. When reconstructing from quantized orientations, the orientation imposition step was modified to impose the quantization bin, rather than the bin center value. At each step, orientations that lie within $\frac{2\pi}{Q}$ of the quantized value are left unchanged, those that are outside are pushed to the closest edge of the bin, either $\Theta_s^Q(m,n) + \frac{\pi}{Q}$ or $\Theta_s^Q(m,n) - \frac{\pi}{Q}$. This may still be interpreted as projection onto a convex subset $W_{\Theta^Q} \subset W$, but where W_{Θ^Q} is a cartesian product of "wedges" rather than rays in W. With this modification to the reconstruction algorithm, as the orientations are quantized more and more coarsely image quality degrades gracefully, as illustrated in Fig. 6(a). Even at extremely course quantization visually pleasing results are produced (Fig. 6(b)). Direct imposition of the quantization bin centers was also attempted, but gave poor results at coarse quantization.



(a) (b)

Figure 6: Effects of quantization. (a) Image reconstructed from orientations quantized to 4 values. (b) SNR relative to the original image vs. number of quantization bins.

7 Conclusions

We have developed a novel image representation based on the orientation of gradients measured at multiple scales. We have proved that the reconstruction algorithm converges and have developed several techniques for improving its convergence rate.

Much recent work in image processing has focused on explicitly modelling or otherwise exploiting the statistical properties of wavelet coefficient magnitudes, often by modelling local variance [12, 13, 14, 15]. This current paper raises the intriguing possibility of building image probability models using the local orientation, a purely geometric quantity, and provides a concrete framework for processing in the orientation domain.

By eliminating the magnitude information, we have reduced the redundancy of the steerable pyramid by nearly a factor of two. Even so, it is clear that the remaining orientation information is redundant. In particular, the prevalence of extended edges and contours in images means that the local orientation measures are likely to be aligned, both within a scale and across scales. As such, we are currently working to develop probability models of this redundancy, as a natural step toward using this representation for traditional image processing applications such as compression or denoising. The representation also seems promising for graphics applications, such as image editing and manipulation [e.g. 16], resolution enhancement, and texture synthesis.

References

[1] P. L. Dragotti, M. Vetterli, and V. Velisavljevic, "Directional wavelets and wavelet footprints for compression and denoising," in *Proc. Int'l Wkshp Adv. Methods for Multimedia Sig. Proc.*, Capri, Italy, September 2002.

[2] D. Taubman and A. Zakhor, "Orientation adaptive suband coding of images," *IEEE Trans. Image Processing*, vol. 3, pp. 404–420, July 1994.

[3] E. Le Pennec and Stéphane Mallat, "Sparse geometric image representation with bandelets," *IEEE Trans. Image Processing*, 2005, To appear.

[4] S Mallat and S Zhong, "Characterization of signals from multiscale edges," *IEEE Trans. PAMI*, vol. 14, no. 7, pp. 710–732, July 1992.

[5] D. Donoho, "Wedgelets: Nearly-minimax estimation of edges," *Annals of Statistics*, vol. 27, pp. 859–897, 1999.

[6] J. Romberg, M. Wakin, and R. Baraniuk, "Multiscale Geometric Image Processing," in *SPIE Visual Communications and Image Processing*, Lugano, Switzerland, July 2003.

[7] X. Li, "On exploiting geometrical constrants of image wavelet coefficients," *IEEE Trans. Image Processing*, vol. 12, no. 11, pp. 1378–1387, Nov 2003.

[8] J. H. Elder, "Are edges incomplete?," *Int'l Journal of Computer Vision*, vol. 34, no. 2/3, pp. 97–122, 1999.

[9] E P Simoncelli and W T Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *2nd Int'l Conf on Image Proc*, Washington, DC, October 1995, vol. III, pp. 444–447.

[10] A.V. Oppenheim and J.S. Lim, "The importance of phase in signals," *Proc of the IEEE*, vol. 69, pp. 529–541, 1981.

[11] Ingo J Wundrich, Christoph von der Malsburg, and Rolf P Würtz, "Image representation by complex cell responses," *Neural Computation*, vol. 16, pp. 2563–2575, 2004.

[12] Martin J Wainwright, Eero P Simoncelli, and Alan S Willsky, "Random cascadees on wavelet trees and their use in analyzing and modeling natural images," *Applied and Computational Harmonic Analysis*, vol. 11, pp. 89–123, 2001.

[13] M K Mihcak, I Kozinstev, K Ramchandran, and P Moulin, "Low complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters*, vol. 6, no. 12, pp. 300–303, 1999.

[14] Javier Portilla, Vasily Strela, Martin J Wainwright, and Eero P Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, 2003.

[15] Xin Li and Michael T Orchard, "Spatially adaptive image denoising under overcomplete expansion," in *IEEE International Conference on Image Processing*, Vancouver, September 2000.

[16] J. H. Elder and R M Goldberg, "Image editing in the contour domain," *IEEE Trans. PAMI*, vol. 23, no. 3, pp. 291–296, 2001.