Image Mapping Using Local and Global Statistics

Yuanzhen Li and Edward Adelson

Department of Brain and Cognitive Sciences
and Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology, Cambridge, MA

ABSTRACT

We describe a set of techniques for mapping one image to another based on the statistics of a training set. We apply these techniques to the problems of image denoising and superresolution, but they should also be useful for many vision problems where training data are available. Given a local feature vector computed from an input image patch, we learn to estimate a subband coefficient of the output image conditioned on the patch. This entails approximating a multidimensional function, which we make tractable by nested binning and linear regression within bins. This method performs as well as nearest neighbor techniques, but is much faster. After attaining this local (patch based) estimate, we force the marginal subband histograms to match a set of target histograms, in the style of Heeger and Bergen.[1] The target histograms are themselves estimated from the training data. With the combined techniques, denoising performance is similar to state of the art techniques in terms of PSNR, and is slightly superior in subjective quality. In the case of superresolution, our techniques produce higher subjective quality than the competing methods, allowing us to attain large increases in apparent resolution. Thus, for these two tasks, our method is very fast and very effective.

Keywords: Image mapping, denoising, superresolution.

1. INTRODUCTION

Many tasks in image processing and machine vision take the form of image to image mapping. Given an image, X, the problem is to estimate another image Y, which is in register with the first. Classical image processing problems such as denoising and superresolution are examples of image mapping. Within machine vision, estimating intrinsic images[2] such as albedo and illumination, or depth, or optic flow would be other examples. Stylistic image mapping, under the names "texture transfer"[3] and "image analogies",[4] forms another category.

There has been much interest in image mapping methods based on machine learning. Given a set of training pairs, (X_1, Y_1); (X_2, Y_2); ..., one learns the statistical relationships between X and Y, so that a new image X_{new} can be mapped to its corresponding pair Y_{new}. The learning is challenging because images are very high dimensional. Markov properties are generally assumed, so that the problem can be approached by modeling local dependencies. But even when we look at relatively small local patches, e.g., 5×5 ones, there are still 25 dimensions, and most of the 25-dimensional space will be empty of observations. For this reason, it is popular to use non-parametric representations of the conditional density, and to use nearest neighbor techniques to estimate the values of the target image.[4,5] The density is not explicitly represented; instead a large number of samples are remembered and later queried. In high dimensional spaces, it is necessary to store a large number of samples, and the queries become quite slow. In addition, artifacts are introduced by the difficulties in enforcing coherence between adjoining patches.

For this reason, we have reconsidered the problem of directly representing the conditional density by learning parametric or semi-parametric models. To make the problem tractable, we work in the subband (e.g., wavelet) domain, so that we can take advantage of the kurtotic distributions found with natural images. In most cases, we represent a single number (the mean) rather than the full density, so that our problem reduces to function approximation. Our thinking is illustrated in Fig.1, where we begin by considering nearest neighbor methods.

Suppose that we wish to represent a 2D function and are given sample values at the points shown in Fig.1(a). The points are densest near zero (here represented by a Gaussian falloff, which is much less kurtotic than image subbands). We can simply store all the samples, and later do a nearest neighbor search. In effect, we place a

Human Vision and Electronic Imaging XIII, edited by Bernice E. Rogowitz, Thrasyvoulos N. Pappas,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6806, 680614, © 2008 SPIE-IS&T · 0277-786X/08/$18

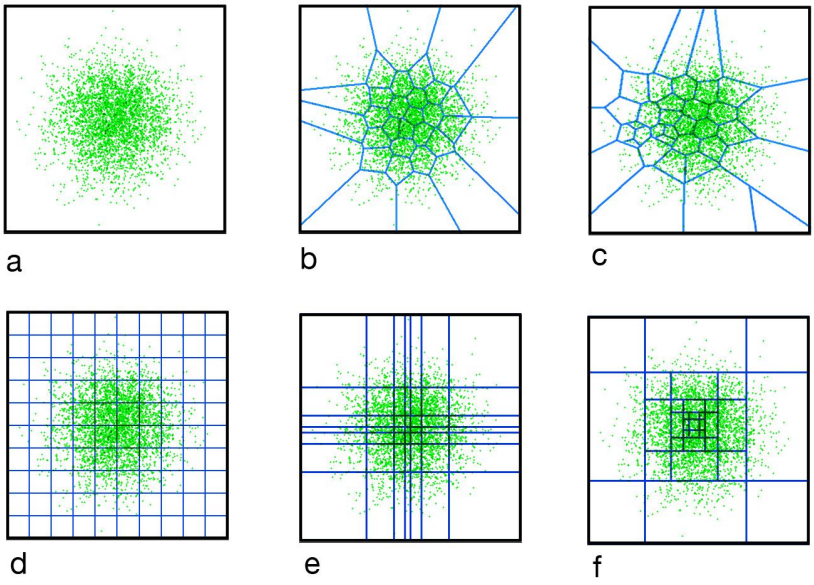


Figure 1. Illustration of different space partitioning techniques. (a) Data cloud. (b) Vector Quantization. (c) Tappen *et al.* (d) Cartesian quantization. (e) Separable binning representative of marginal densities. (f) Nested binning.

Voronoi neighborhood around every observation. When we have many samples, it becomes prohibitive to store and to query this information. We can reduce it to a smaller set of neighborhoods that are deemed representative. In vector quantization (VQ), shown in Fig.1(b), nearby points are merged to form larger neighborhoods containing a certain number of samples. Popat and Picard[6] utilize VQ. Rather than basing the neighborhoods on sample density, Tappen et al.[7] greedily select a subset that are determined to be most informative. Fig1(c) is suggestive of the sort of neighborhoods they might find. Next, consider a set of more direct representations. Fig.1(d) shows a simple case of Cartesian quantization. The space is divided into rectangular bins and the function is represented parametrically within each bin. Instead of storing samples, one represents the function explicitly, and there is no need for nearest neighbor computations at query time. There is one big disadvantage: representing a high dimensional function requires a great many bins. The number of bins required grows exponentially with dimensionality. We can improve things by moving the bin boundaries around to reflect the marginal densities, as shown in Fig.1(e). However, this separable method still leaves a great many bins.

We have devised a better approach, which we call nested binning, shown in Fig1(f). The separable bins in (e) which are representative of the marginal densities get merged at certain places, leading to partitions better matched to image statistics, and still extremely easy to query. We will give more details in section 3. Within each bin, the function is approximated parametrically, with the parameters fit by regression within the bin. We achieve state of the art performance with compact models and very fast lookup. Our method is faster than k nearest neighbor methods by factors of 1000 or higher, even when compared to fast methods like k-d trees.[8,9]

The outputs of many image mapping techniques, including our nested binning technique, tend not to look quite right. For example they may look too soft, or they may have irritating artifacts. There is something about the overall image quality that is not quite right. This subjective impression reflects a legitimate issue: these images look wrong because they are very unlikely to occur. We can force them back into a more plausible subspace by imposing textural statistics such as subband marginals, in the manner of Heeger and Bergen.[1] For applications in which a human viewer is the end user, optimizing the appearance is often the right thing to do.

In this paper, we demonstrate the performance of our system on denoising and superresolution. In both cases, we first learn to make local estimates of subband coefficients by learning from training pairs. We next force the image to have the marginal statistics that are estimated from a set of natural images used for training. The combination of local and global techniques leads to images that are competitive with state of the art methods, and allow for fast computation. In the case of superresolution, our method produces images that are noticeably

superior in appearance to any competing method we have found. We have verified this advantage with user studies.

2. ALGORITHM OVERVIEW

Let $\{(X_i, Y_i)\}_{i=1}^N$ be a set of N training pairs. Each pair contains input image X_i and output image Y_i. We decompose the images into subbands, and for each subband coefficient of Y_i compute a feature vector from X_i and/or its subbands. We train regressions to predict Y_i's subband coefficients given X_i's feature vectors. Given a new input image, feature vectors are computed, and the above learned regressions are used to predict the subband coefficients of the output image. Subband coefficients constrain local neighborhoods of an image, and we refer to this stage as the learning of local constraints. Details of the regression algorithm will be given in section 3.

The output image reconstructed from the above estimated subband coefficients tends to look too soft. We impose textural statistics such as subband histograms, to correct the "look". Subband histograms constrain the whole image, and we refer to this stage as the learning of global constraints. Details will be given in section 4.

2.1 Markov random fields

The learning and utilization of local constraints, can be understood as learning and inference on a Markov random field, with the assumption that the probability of a local neighborhood can be modeled as a product of probabilities of individual subband coefficients, and that each coefficient is Gaussian when conditioned on the input. In probability language, we wish to model the conditional probability of Y given X, and when given a new input image X_{new}, the output can be estimated as the image that maximizes the probability of Y given that X is equal to X_{new}, *i.e.*, $\hat{Y}_{new} = \arg\max p(Y|X = X_{new})$.

Both X and Y are very high-dimensional. We assume that Y obeys the Markov property when conditioned on X: $p(y_k|X, \mathbf{y}_w, w \neq k) = p(y_k|X, \mathbf{y}_w, w \in N^y_{(k)})$, where y_k is a pixel value at location k, and $N^y_{(k)}$ is a neighborhood of k in Y but not containing k. By the Hammersley-Clifford theorem of random fields, the joint distribution of $(y_1, \ldots, y_n)$ given X has the form

$$p(Y|X) \propto \exp\Big(-\sum_k M_k(\mathbf{y}_{(k)}|X)\Big), \tag{1}$$

where $\mathbf{y}_{(k)}$ is a clique, and $M_k(\mathbf{y}_{(k)}|X)$ is the potential function of that clique, given the observed image X.

We model the probability distribution $p(\mathbf{y}_{(k)}|X)$, which is related to $M_k(\mathbf{y}_{(k)}|X)$ by an exponential, by a product of T one-dimensional distributions:

$$p(\mathbf{y}_{(k)}|X) = \prod_{t=1}^{T} p(B_t^T \mathbf{y}_{(k)}|X), \tag{2}$$

where $\{B_t\}_{t=1}^T$ are a set of subband basis vectors. Here the cliques are defined on $m \times m$ neighborhoods where all the m^2 pixels are connected.

The next question is to model each of the individual conditionals, $p(B_t^T \mathbf{y}_{(k)}|X)$. We simplify $B_t^T \mathbf{y}_{(k)}$, the subband coefficient at location k in subband t, by $b_{t,k}$. The simplest choice for $p(b_{t,k}|X)$ is a conditional Gaussian: $b_{t,k}|X \sim N(\hat{b}_{t,k}, \sigma^2_{t,k})$. $\hat{b}_{t,k}$ depends on X:

$$\hat{b}_{t,k} = E\big[b_{t,k}|X\big] = \hat{b}_{t,k}(X), \tag{3}$$

and can be estimated through regression. $\sigma_{t,k}$ can also be estimated during regression, but for simplicity we make it be a constant.

Once ($\hat{b}_{t,k}$, $\sigma_{t,k}$) are in place, the clique potentials can be written as:

$$M_k(\mathbf{y}_{(k)}|X) = \sum_t \frac{1}{2\sigma^2_{t,k}} \Big(B_t^T \mathbf{y}_{(k)} - \hat{b}_{t,k}(X)\Big)^2. \tag{4}$$

The log probability of the random field is (offset by a constant):

$$\log\big(P(Y|X)\big) = -\sum_k M_k(\mathbf{y}_{(k)}|X)$$
$$= -\sum_k \sum_t \frac{1}{2\sigma_{t,k}^2}\Big(B_t^T\mathbf{y}_{(k)} - \hat{b}_{t,k}(X)\Big)^2, \tag{5}$$

which we wish to maximize by estimating an image $\hat{Y} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)$. When $\sigma_{t,k}$ is a constant, and $\{B_t\}_{t=1}^T$ is a tight frame,[10] $\hat{Y}$ can be reconstructed from $\hat{b}_{t,k}$ by applying the self-inverting subband transform to the estimated coefficients $\hat{b}_{t,k}$.

2.2 Subband histograms

The log probability in (5) is quadratic, and can easily be maximized. By constraining overlapping neighborhoods with subband coefficient estimates, we do not need to explicitly enforce spatial coherence, as Freeman *et al.* and Hertzmann *et al.* do. But we do have a problem of blurring, which is related to the Gaussian conditional model we use. Blurring is reflected in the subband marginals, in that they tend to be much less kurtotic than those of a sharp image. The global, marginal histograms of subband coefficients can provide important statistical constraints for images that are perceptually alike. Heeger and Bergen[1] demonstrated that by matching the subband histograms of a Gaussian noise image to those of a stochastic texture, new examples of the texture can be synthesized with matching appearance.

We use subband histograms as constraints to impose globally on the output image, Y. By making Y have similar subband histograms to those observed from a particular class of images, we make it a more probable instance of that class. One concern, however, is that the subband marginals of different images of the same class, while having similarities (*e.g.*, all being kurtotic), may still be different. In section 4, we show how to account for variations among images, and how to utilize the observed image X to retrieve good estimates of the target histograms.

3. LEARNING LOCAL CONSTRAINTS

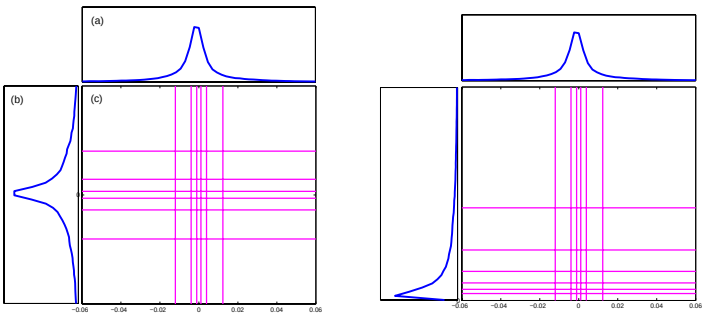
For each subband of Y, we train a regressor to predict its coefficients using features extracted from local neighborhoods of the input image X. In general, three types of features are considered: linear subband coefficients, nonlinear local energy values, and pixel values. Local energy is computed by rectifying the subband coefficients and then blurring them across space or orientations.

We can use powerful nonparametric methods to perform the regression. A problem is that they tend to be very costly when the training set is large and the dimensionality is high. For the applications considered in this paper, the size of the training set can be quite large, because each subband coefficient (together with its feature vector) constitutes a sample. For example, if we have 20 500 by 500 images for training, and use a spatially oversampled subband decomposition, then for each subband there will be 5 million training samples. When the training set is so large, even with a moderate number of dimensions, the computation involved in nonparametric methods can be very expensive.

We notice that with the subband coefficient based features we are considering, images demonstrate interesting statistics in the feature space. Based on the feature statistics we propose a partitioning procedure to divide the space into easily retrievable bins. In each bin we perform parametric regression, leading to a compact representation of the learned model.

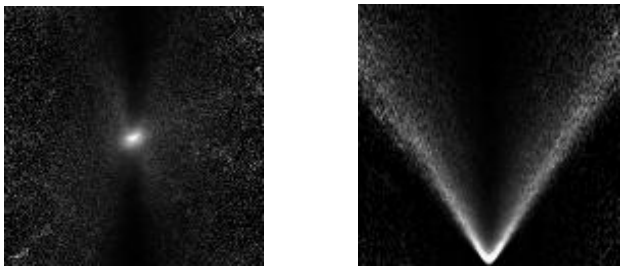
3.1 Space partitioning

A conceptually easy way to obtain an easily retrievable partition, is to divide the space into hyper-rectangular bins. Marginal equiquantization,[11,12] used in the estimation of information theoretic measures, is a method to acquire separable hyper-rectangular bins. Each feature dimension is examined separately, and the partition is defined by marginal bins which are not equidistant but equiprobable intervals, meaning each marginal bin encapsulates approximately the same number of observed data samples. The partition is separable, so is extremely easy to obtain (via order statistics) and to retrieve. Fig.2 shows two such partitions on two dimensions, with different feature pairs. In (a), the features are subband coefficients at the same location and orientation, but at adjacent scales. In (b), one feature is subband coefficient, and the other is local energy computed by rectifying and



(a) (b)

Figure 2. Marginal equiquantization using two features.(a)Features are subband coefficients at adjacent scales.(b)One feature (horizontal) is subband coefficient, and the other (vertical) is local energy.



(a) (b)

Figure 3. Conditional histograms of: (a) one coefficient on another at an adjacent scale; (b) local energy on subband coefficient. Dark means low, and bright means high density.

blurring the coefficients. All the features here have kurtotic distributions, as shown by the marginal histograms on the top and the left of the partitions, and the data are highly clustered around zero. The bins are small near zero, and big away from zero.

If the features were statistically independent of each other, the joint bins obtained through marginal equiquantization would also be statistically equivalent.But the features are usually not independent of each other. Subband coefficients at adjacent locations, scales, or orientations, tend to be independent when amplitudes are small, but correlated when amplitudes are large,[13] as illustrated by a conditional histogram in Fig.3(a). Coefficient amplitudes and local energy values, also tend to be correlated, as shown by a conditional histogram in Fig.3(b).

We find such pattern to be quite consistent when more features, such as coefficients and energy values at neighboring locations and orientations, are added. The pattern is that the data tend to be highly clustered around zero, and get predictably sparse away from zero, except when amplitudes of features show correlations. The joint bins acquired via marginal equiquantization, therefore, tend to be empty when some feature amplitude is large but not correlated with the amplitudes of others. We use a "nesting" strategy to merge the bins that tend to be sparsely populated. The nesting is performed in a hierarchical fashion, as illustrated in Fig.4. The boundaries of the nested bins are still parallel to the feature axes, therefore the retrieval is efficient. Nesting slows down the rate of total bin number growing with dimensionality. After nesting, most bins will be filled with enough data samples, *i.e.*, no fewer than the number of unknown parameters, to perform regression. Some bins will still be empty, and for them we progressively include points from neighboring bins until we can reliably estimate the regression parameters.

3.2 Piecewise linear regression

After the feature space is partitioned into bins, parametric regression is performed within each bin. If we choose the simplest parametric model, *i.e.*, linear model with Gaussian noise, the regression parameters can be estimated using the standard least squares method.

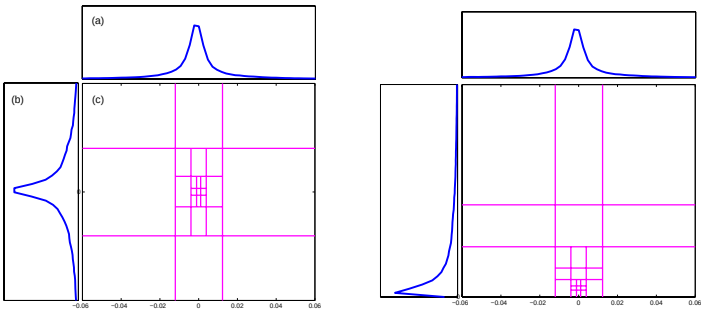


Figure 4. Nested binning using the same feature pairs as the ones used in Fig.2 and Fig.3. The bins in Fig.2 are merged at places where one feature but not the other has high amplitude.

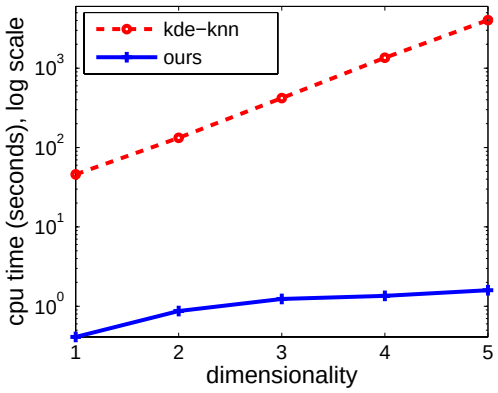


Figure 5. CPU time comparison, on log scale, between our method and kd-tree accelerated k nearest neighbor (knn).[9]

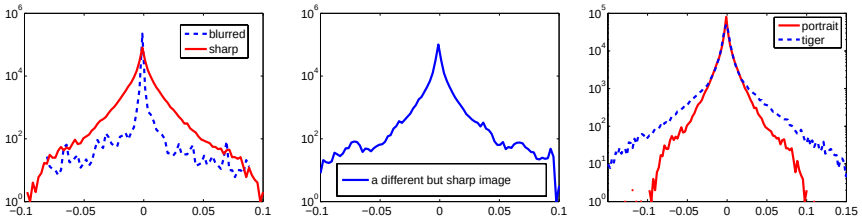
When there are many training images, it is often desirable to start learning without having to wait for all data to be observed and stored. We make the regression process online, using recursive least squares.[14] The space partitioning requires marginal order statistics (percentiles), which can be estimated from partial training data.

Compared to nonparametric methods, our method leads to a much more compact model, which can be efficiently queried. For each subband, only the bin boundaries and a small number of parameters per bin need to be stored. We empirically compare the running time of our method to that of kd-tree accelerated k nearest neighbor,[9] in Fig.5. Both methods are tuned so that they produce matching accuracy (difference in estimation PSNR < 0.1dB). There are 305,133 training samples and 262,144 testing samples. The plot is on log scale.

4. LEARNING GLOBAL CONSTRAINTS

In the stage of learning local constraints, we typically learn a conditional mean of each subband coefficient, given a feature vector representing the surrounding patch. If the conditional probability of the coefficient is truly Gaussian, as has been assumed, then the conditional mean is also the maximum probability solution. But in some cases, the conditional cannot be well approximated with a Gaussian, and by doing regression we end up blurring the estimated signal. This phenomenon can be observed in the application of super-resolution. A superresolution result using the regression estimates is shown in Fig.9(e), which looks blurred.

We think that "blur" is a textural quality, and can be captured by textural statistics, for example subband histograms. Fig.6(a) shows a subband histogram of an image, against that of a blurred version of the image. The two histograms are vastly different. A different sharp image, on the other hand, has subband histogram much more alike, as shown in (b). We propose to enforce "sharpness" as global textural constraints, matching an image's subband histograms to ones that are observed from the training images.



(a) (b) (c)

Figure 6. Subband histograms, on log scale, (a) of a sharp image (solid), and of its blurred version (dash); (b) of a different but sharp image; (c) of a portrait image (solid), and a tiger image (dash).

Another way to think about this, is that the "local" result that maximizes the log probability in (5), has been found in the entire continuous space of $\Re^n$. Images are special signals, and are commonly hypothesized to lie within a subspace of $\Re^n$ instead. One possibility to take this into account is to find the subspace, and search within it for an image that maximizes (5). But it is difficult to do so. Instead we employ a two-step strategy, first finding a solution in $\Re^n$ that maximizes (5), and then projecting it to the closest point that satisfies the target subband marginal constraints. This point is found using Heeger-Bergen style histogram matching.

One thing to keep in mind, is that the subband marginals of natural images, while all being kurtotic, may still vary a lot from image to image. Fig.6(c) shows two histograms of a portrait image and a tiger image, respectively. The overall "texture" of the two images is different, and the histograms are different. When we have a varied set of training images, we want to pick one or a few that are texturally similar to the target result. The problem is, with a blurred image whose real high frequencies are unknown, how do we pick such images? Note that the low frequency subbands of a blurred image are basically undegraded, and their marginals can be used as global textural features indexing into the training set. More specifically, the histograms of low frequency subbands of each training image are compared with the low frequency subband histograms of the blurred image, and the k nearest neighbors are found. The histograms of the higher frequency subbands of the k nearest neighbors are then averaged to give the target histograms. Distances between subband marginals are calculated using K-L divergence.

5. APPLICATIONS AND EXPERIMENTS

5.1 Image denoising

To test our system on denoising, we added Gaussian white noise. We generated image training pairs by adding synthetic noise to known clean images. The mapping from "noisy" to "clean" was learned, and then applied to new noisy images.

First we will present denoising results using only the local constraints, *i.e.*, the regression estimates of subband coefficients. Training such a regression model, can be understood as learning a multi-dimensional "coring" function from actual image data. In the classical coring technique,[15,16] a subband coefficient is subjected to a 1D nonlinear function which suppresses small amplitude values and preserves high amplitude values. For the multidimensional case we use multiple features as input and learn a multidimensional function as the output. We experimented with various feature vectors, and settled on a 6D vector with the following features: the observed subband value; the observed value of its parent (same location, same orientation, but lower frequency); the observed value of its grandparent; and the local energies corresponding to these three subbands. Local energy is calculated within a subband by taking the absolute value and convolving with a blur kernel that is about the same size as the subband kernel. All operations were performed on an oversampled pyramid. For the results reported here, 9 bins per dimension were used for marginal equiquantization, which after nesting gave 2913 bins in total. The lowpass was kept unchanged.

We tested the algorithm using two different subband representations, and two different training sets. The first subband representation was a variant of the steerable pyramid,[17] using 8 orientations and 4 scales as in.[18] The second subband representation is an oversampled QMF pyramid.[19] The first training set contains 18 images (5.7 million pixels altogether) which all have people in them. The second training set contains 40 images (6.2

Figure 7. Denoising results. (a) Noisy image, with Gaussian additive noise of standard deviation 20. Noisy PSNR is 22.11dB. (b) Denoised image, using local constraints only. PSNR = 32.47dB. (c) Denoised image, using both local and global constraints. It looks pleasantly sharper, but the PSNR is lower than (b), being 31.13dB.

million pixels) from the Berkeley segmentation data set.[20] They are the same 40 used by Roth and Black for training image priors.

The combination of subband representation and training set gives four different settings. Performances under the four settings are compared in Table 5.1, in terms of Peak Signal-to-Noise Ratios (PSNRs), on standard test images for denoising. When the four results are averaged, the average tends to have higher PSNR than the individual results, as shown in column 7. We vary the noise levels, and list more results in Table 2. In general, the performance of our algorithm matches those of the state-of-art denoising algorithms, such as Portilla *et al.*,[18] Roth and Black,[21] etc.

Upon the local results we then impose global statistics, *i.e.*, subband histograms. For global statistics we use the steerable subband representation with 3 scales and 4 orientations, and the first training set. Fig.7 shows the local and global results on Lena, and Fig.8 shows a blow-up, compared against Portilla et al., and Roth and Black. Note the fine details we are able to recover. One interesting thing to note, is that although we feel our global result looks the most pleasing, it has the lowest PSNR among the four. The reason is that, by imposing global statistics, we are forcing the overall "texture" (such as the appearance of sharp edges), to resemble that we have seen in other clean images, thus pushing some individual pixel values away from the minimum square error (MSE) estimates. PSNR is a square error based criterion, and does not capture all aspects of distortions relevant to perceptual quality. Here we argue that textural similarity is often a valid additional criterion for perceptual similarity. In order to test this argument, we conducted psychophysical experiments where human subjects were asked to judge the perceptual qualities of results produced by our algorithm and the BLS-GSM algorithm by Portilla *et al.*. Rather than using a rating scale, which can be hard to interpret, we asked our subjects to match the subjective quality of images from one technique with those of the other technique at a different noise level. In this way, we could determine that one technique tolerated a certain amount more noise than the other. We express the advantage in terms of dB of the additive noise signal.

We ran 6 subjects, used 7 images and 3 levels of noise (σ=15,25,50). The results depended on the individual subjects and images, but when averaged our method outperformed BLS-GSM. When averaged across subjects, the advantage ranged from 0 dB to 6.16 dB, with mean at 3.15 dB.

5.2 Super-resolution

For super-resolution, the task is to estimate a sharp looking high- resolution image given a low-resolution image as input. The low res input was generated by downsampling the image by a factor of c, and then upsampling it by a factor of 1/c, both via bicubic interpolation, leading to a blurry image. We used high and low resolution pairs for training. Our super-resolution algorithm is almost exactly the same as that for denoising, except that the training pairs are different.

(a) (b) (c) (d)

Figure 8. Denoising results. (a) Ours, using local constraints only. (b) Ours, using local and global constraints. (c) Roth and Black,.[21] (d) Portilla *et al.*[18] Note the fine eyelash and feather boa details we recovered in (b).

	St-TS1	Qmf-TS1	St-TS2	Qmf-TS2	mean of PSNRs	σ of PSNRs	PSNR of mean result
Lena	31.47	31.44	31.09	31.12	31.28	0.205	31.52
Barbara	25.66	25.96	27.62	27.08	26.58	0.926	26.80
Boats	29.05	29.17	29.15	29.15	29.13	0.056	29.32
House	31.08	31.05	30.56	30.75	30.86	0.246	31.14
Peppers	27.94	28.71	28.43	28.89	28.49	0.416	28.80

Table 1. Denoising performance as PSNRs in dB, with Gaussian noise $\sigma = 25$ (20.17dB), using two different subband decompositions and two different training sets. The columns are: St-TS1 (steerable subbands, training set 1); Qmf-TS1 (QMF subbands, training set 1); St-TS2 (steerable, training set 2); Qmf-TS2 (QMF, training set 2); mean of PSNRs (average of the PSNRs in the first four columns); σ of PSNRs (standard deviation of the PSNRs in the first four columns); PSNR of mean result (PSNR of the average result using all the four settings).

In Fig.9 (f) we show our super-resolution result on an image downsampled by a factor of 4 in each direction, and compare it to competing techniques, including Freeman *et al.*,[22] Hertzmann *et al.*,[4] and Genuine Fractals, which is a leading commercial product. Our result with both local and global constraints looks sharp and is free of disturbing artifacts. The training set here was the set of 18 people images. Steerable subband representation with 3 scales and 4 orientations were used. In the local constraints learning stage we used the same set of 6 features as those used for denoising. The same subband representation was used for global constraints estimation and imposition.

For super-resolution, to the best of our knowledge, there has not been a satisfying measure quantifying the quality of results. Here the goal is to hallucinate missing high frequencies, especially those belonging to sharp edges. If the hallucinated edge position is slightly wrong, the square error will be large, although the image may look good. Therefore PSNR is not a good way of measuring subjective quality. Again we argue that textural similarity, reflected by the differences in global subband statistics, can be a valid criterion. We conducted psychophysics experiments to compare out method with the others. In this case, the image from one method at a given magnification was matched in quality to another technique at another magnification. Our method performed very well in this comparison. The best competitor was Genuine Fractals. Averaged across subjects and images, we are able to tolerate between 1.16 and 1.52 more magnification than Genuine Fractals, with 1.30 as the average.

ACKNOWLEDGMENTS

Supported by Shell Labs and National Science Foundation.

	Lena		Barbara		Boats		House		Peppers	
$\sigma_n/PSNR_n$	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
15/24.61	33.45	33.65	30.02	30.40	31.33	31.52	33.07	33.32	31.20	31.51
20/22.11	32.25	32.47	28.00	28.30	30.16	30.33	31.88	32.13	29.76	30.05
25/20.17	31.28	31.52	26.58	26.80	29.13	29.32	30.86	31.14	28.49	28.80
50/14.15	28.23	28.47	23.69	23.80	26.17	26.33	27.55	27.87	24.83	25.06

Table 2. Denoising performance as PSNR, in dB, on a few standard test images, with varying levels of noise. Two PSNRs are reported for each image, the first one (C1) being the average of four result PSNRs (St-TS1,Qmf-TS1,St-TS2,Qmf-TS2, as in Table 1), the second one (C2) being the PSNR of the average of the four results.

(a) (b) (c)

(d) (e) (f)

Figure 9. Super-resolution results. 311x258 image is downsampled by 4 in each direction, and then super resolved. The downsampling is done by bicubic interpolation. (a) Bicubic upsampling. (b) Commercial software Genuine Fractals v4.1. (c) Hertzmann *et al.*, image analogies.[4] (d) Freeman *et al.*, example-based.[22] (e) Our result, without global constraints. (f) Our result, with global constraints.

REFERENCES

1. D. J. Heeger and J. R. Bergen, "Pyramid-based texture analysis/synthesis," in *SIGGRAPH*, 1995.
2. H. Barrow and J. Tenenbaum, "Recovering intrinsic scene characteristics from images," *A. Hanson and E. Riseman, editors, Computer Vision Systems* , 1978.
3. A. A. Efros and W. T. Freeman, "Image quilting for texture synthesis and transfer," in *SIGGRAPH*, 2001.
4. A. Hertzmann, C. Jacobs, N. Oliver, B. Curless, and D. Salesin, "Image analogies," in *SIGGRAPH*, pp. 327–340, 2001.
5. W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," *International Journal of Computer Vision* **40**(1), pp. 25–47, 2000.
6. K. Popat and R. W. Picard, "Cluster based probability model and its application to image and texture processing.," *IEEE Trans. Image Processing* **6**(2), pp. 268–284, 1997.

7. M. F. Tappen, E. H. Adelson, and W. T. Freeman, "Estimating intrinsic component images using non-linear regression," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006.
8. A. Gray and A. Moore, "Very fast multivariate kernel density estimation via computational geometry," *Proc. Joint Stat. Meeting* , 2003.
9. A. Ihler, "Kernel density estimation toolbox for matlab." http://www.ics.uci.edu/ ihler/code/, 2003.
10. I. Daubechies, *Ten lectures on wavelets*, Society for Industrial and Applied Mathematics, 1992.
11. M. Menendez, D. M. D, and L. Pardo, "Maximum entropy principle and statistical inference on condensed ordered data," *Statistics and Probability Letters* **34**(1), pp. 85–93, 1997.
12. J. Hudson, "Signal processing using mutual information," *IEEE Signal Processing Magazine* **23**(6), pp. 50–58, 2006.
13. E. P. Simoncelli, *Handbook of Image and Video Processing*, ch. 4.7. Statistical modeling of photographic images. Academic Press, 2nd ed., 2005.
14. G. Strang, *Introduction to Applied Mathematics*, ch. 2.5 Least squares estimation and the Kalman filter, pp. 146–148. Wellesley-Cambridge Press, 1986.
15. D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," in *Biometrika*, **81(3)**, pp. 425–455, 1994.
16. E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *Third International Conference on Image Processing*, **I**, pp. 379–382, 1996.
17. E. Simoncelli, W. Freeman, E. Adelson, and D. J. Heeger, "Shiftable multi-scale transforms," *IEEE Trans Information Theory, Special Issue on Wavelets* **38**(2), pp. 587–607, 1992.
18. J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain." *IEEE Trans. Image Processing*, 2003.
19. E. H. Adelson and E. P. Simoncelli, "Hexagonal qmf pyramids," in *Proceedings of the Optical Society of America Topical Meeting on Applied Vision*, pp. 5–8, (San Fransisco, CA), July 1989.
20. D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th Int'l Conf. Computer Vision*, **2**, pp. 416–423, July 2001.
21. S. Roth and M. Black, "Fields of experts: A framework for learning image priors." *in Proc. of CVPR, vol. 2, pp. 860–867*, 2005.
22. W. Freeman, T. Jones, and E. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics Appllications* **22**(2), pp. 56–65, 2002.