A SPARSITY-DISTORTION-OPTIMIZED MULTISCALE REPRESENTATION OF GEOMETRY

Osman G. Sezer, Yucel Altunbasak and Onur G. Guleryuz

Center for Signal and Image Processing, Georgia Institute of Technology, Atlanta, GA 30308, USA
 Wayne State University, Detroit, MI 48202, USA

ABSTRACT

This paper describes the construction of a new multiresolutional decomposition with applications to image compression. The proposed method designs sparsity-distortion-optimized orthonormal transforms applied in wavelet domain to arrive at a multiresolutional representation which we term the Sparse Multiresolutional Transform (SMT). Our optimization operates over sub-bands of given orientation and exploits the inter-scale and intra-scale dependencies of wavelet coefficients over image singularities. The resulting SMT is substantially sparser than the wavelet transform and leads to compaction that can be exploited by well-known coefficient coders. Our construction deviates from the literature, which mainly focuses on model-based methods, by offering a data-driven optimization of wavelet representations. Simulation experiments show that the proposed method consistently offers better performance compared to the original wavelet-representation and can reach up to 1dB improvements within state-of-the-art coefficient coders.

Index Terms— sparse multiscale representations, sparse orthonormal transforms, wavelet transform, sparse dictionary.

1. INTRODUCTION

Wavelet-based multiresolutional approximations can be shown to achieve the asymptotically optimal approximation rates over uniformly smooth signals with point singularities. However, the performance of wavelet approximations decreases significantly when one considers more realistic image models such as those that have singularities along curves [1]. State-of-the-art wavelet coders [2, 3, 4] are good at exploiting the coherence among wavelet coefficients but they nevertheless fall short over the realistic models since the underlying transform does not lead to sufficient sparsity.

Available literature trying to improve the shortcomings of the wavelet transform can be decomposed into two mainstream approaches. The first category of work formulates 2D geometrical singularities in signal domain and develops representations that provide sparse decompositions over them. Curvelets [5], contourlets [6], and first generation bandelets [7] can be grouped into this class. The directional lifting-based methods such as [8, 9, 10] can also be included into this group since they modify the prediction step of the original filter-bank into a directional prediction step. The second category on the other hand, formulates the problem in transform domain and proposes representations in terms of wavelet coefficients. Wavelet-footprints [11], wedge-prints[12], and second generation bandelets [13] can be counted in this group. Foot and wedge-prints exploit inter-scale dependencies over singularities by introducing a vector dictionary, assuming step edges. Second generation bandelets adaptively reorder coefficients and apply a secondary wavelet transform.

The technique we introduce in this paper is also a second category approach. Unlike other techniques however, our method is agnostic to the type of singularity as it is instead built around core sparsity ideas. Generalizing our early work on block and lapped transforms [14], we design sparsity-distortion-optimized orthonormal transforms that exploit the inter-scale and intra-scale dependencies of wavelet coefficients. These transforms, when applied on the top of a given wavelet transform, map wavelet coefficients over signal singularities to a new, sparser decomposition. We term the resulting overall multiresolutional representation as the sparse multiresolutional transform (SMT).

The paper is organized as follows. In the next section, we formulate the tree data-structure using which the new transforms are optimized and review basic approximation results. Section 3 details the construction of the proposed transforms and Section 4 outlines integration inside a codec. Simulation experiments are discussed in Section 5, followed by Section 6 of concluding remarks.

2. WAVELET DECOMPOSITION AND SUB-TREE FORMATION

Consider a multiscale decomposition of an image f with the 2D separable wavelet transform. Beyond the lowest frequency sub-band which our method leaves unaltered, such a decomposition represents an image in terms of sub-bands having three different orientations given by vertical, diagonal, and horizontal subbands (refer to Fig.1.) Let \( \omega_{j,n} \) be the
wavelet coefficient of the image at scale $j$, subband orientation $s$, and location $n = (n_1, n_2)$ [1]. Due to the coherence of wavelet coefficients, one can build “sub-trees” of wavelet coefficients across scales to analyze the local image structure. For a discrete image $f$ of size $N \times N$, the children of the coefficient $\omega^s_{j,n}$ can be defined as

$$C^s_{j,n} = \{\omega^s_{j+1,n'}, n' = 2n + (x, y), (x, y) \in \{0, 1\}\}.$$

Using this relation between wavelet coefficients, a sub-tree rooted at $(j, n)$ (as in Fig.1) can be expressed as

$$Q^s_{j,n} = \omega^s_{j,n} \cup \{\omega^s_{j',n'} | j < j' \leq \log_2(N), n'_1 \in [2^{j'-j}n_1, 2^{j'-j}(n_1 + 1/2)], n'_2 \in [2^{j'-j}n_2, 2^{j'-j}(n_2 + 1/2)]\}.$$

The sparse representation achieved by the wavelet decomposition ensures that the wavelet coefficients of a smooth region will be rapidly decaying. Hence, many of the coefficients within a sub-tree as defined above are likely to have small (or insignificant) magnitudes. On the other hand, if the region contains a singularity, the wavelet coefficients within corresponding sub-trees are expected to be significant. The wavelet-based compression algorithms such as [3, 2] utilize these observations and obtain significant gains whenever the wavelet-based compression algorithms such as [3, 2] utilize these observations and obtain significant gains whenever the number of significant sub-trees is small.

In a more analytical setting, using continuous time analysis, one can show that the distortion due to the compression of uniformly smooth signals with point singularities asymptotically complies with

$$D(R) \lesssim R^{-\alpha} \quad (1)$$

where $R$ is the number of bits spent to represent the signal and $\alpha$ quantifies the local smoothness of the signal, with larger $\alpha$ corresponding to smoother signals [15, 1]. Note that regardless of the point singularities, the operational distortion-rate function tracks the smoothness of the signal and obtains the asymptotically optimal performance (see [15] for conditions). In comparison, when the signal contains singularities along curves one obtains

$$D(R) \lesssim R^{-1} \quad (2)$$

regardless of how large $\alpha$ is. Hence, isolating singularities and using better decorrelation techniques has become a focal point of recent image compression related research.

### 3. CONSTRUCTION OF SMT

Rate-distortion optimized transform design aims to find the best orthonormal transform(s) that will minimize the distortion level for a given bit budget ($B$). Let $H$ be an orthonormal transform and assume image $f \in F$ where $F$ is the set of images that are of interest. As a general transform optimization we can write

$$\min_H E[|D(H; f)|] \quad \text{s.t.} \quad E[|R(H; f)|] \leq B \quad (3)$$

where the expectations $E[|D(H; f)|]$ and $E[|R(H; f)|]$ are obtained over $F$. Using a Lagrange multiplier $\lambda$, the above problem can be reformulated into an unconstrained minimization as

$$\min_H E[|D(H; f)| + \lambda R(H; f)] . \quad (4)$$

Our construction specializes this most general formulation by considering images in wavelet domain, and by defining a set of transforms that operate over classes of sub-trees. Since the wavelet transform is localized, one in general expects sub-trees to cluster depending on the regions of the image they are obtained over. One basic cluster is formed by the “smooth sub-trees” class, i.e., the sub-trees obtained over the smooth regions of the image over which the wavelet decomposition works well. The complement of this class, i.e., “non-smooth sub-trees”, is difficult to capture within one cluster as images typically contain many types of singularities. We hence sub-classify “non-smooth sub-trees” in the manner described below so that transforms can be better adapted to different types of singularities. In this setting the above minimization becomes

$$\min_{G_k} E[|D(G_k; \omega_k)| + \lambda R(G_k; \omega_k)] \quad \text{s.t.} \quad G_k^T G_k = I \quad (5)$$

where $\omega_k$ ranges over sub-trees of type $k$, $K$ is the total number of sub-tree clusters, and $G_k, k = 1, ..., K$, denotes the transform for the $k^{th}$ cluster.

Various techniques can be used to generate sub-tree classes. The one employed in this paper groups sub-trees with respect to the direction of geometric flow within the corresponding image region [7]. As wavelet basis functions have different orientations, it is clear that sub-trees emerging from sub-bands having different orientations will be capturing different aspects of the underlying geometry. We thus further refine the design process to carry out the optimization at each orientation separately to obtain

$$\min_{G_k} E[|D(G_k^s; \omega_k^s)| + \lambda R(G_k^s; \omega_k^s)] \quad (6)$$

$$\text{s.t.} \quad G_k^s^T G_k^s = I$$
where $\omega_s^k$ denotes sub-trees of type $k$ at sub-band orientation $s \in \{V, D, H\}$. With this approach, each sub-tree class has three orthonormal transforms that are separately optimized based on sub-band orientation.

Two aspects of (6), the rate term and the expectation, are further specialized for SMT. Since the analytical expression for rate at a given distortion level is difficult, we approximate rate to arrive at a more tractable cost function. We resolve expectations by using a training set. Let $S_{(k,s)}$ be a set of training sub-trees of orientation $s$ and type $k$. The simplified form of (6) becomes

$$k \in \{1, ..., K\}, s \in \{V, D, H\} :$$

$$\min_{G_k^s} \left\{ \sum_{q_i \in S_{(k,s)}} \min \|q_i - G_k^s c_i\|^2_2 + \lambda \|c_i\|_0 \right\} \quad \text{s.t.} \quad G_k^s T G_k^s = I \quad (7)$$

where $q_i$ is the $i$’th sub-tree in $S_{(k,s)}$ lexicographically ordered into a vector and $c_i$ denotes the transform coefficients of $q_i$ after application of $G_k^s$. Here, rate is estimated through nonlinear approximation ($L_0$ norm,$\|\|_0$) by counting the number of nonzero coefficients and Euclidean norm is used to estimate distortion (refer to [16], [15] for justifications.) Using iterative conditional minimization over coefficients and transforms as suggested in [14], the solution to Eq. 7 can be divided into two steps;

I. **Optimal coefficients for a given transform:** The sparsest representation for a given transform can be found by solving

$$c_i = \arg \min_d (\|q_i - G_k^s d\|^2_2 + \lambda \|d\|_0). \quad (8)$$

Note that since $G_k^s$ is orthonormal, the optimal solution is obtained by hard-thresholding the components of $d = G_k^s q_i$ with $\sqrt{\lambda}$.

II. **Optimal transforms for given coefficients:** The optimal orthonormal transform for given coefficient vectors can be found by solving

$$G_k^s = \arg \min_H \left\{ \sum_{q_i \in S_{(k,s)}} \|q_i - H c_i\|^2_2 \right\} \quad \text{s.t.} \quad H^T H = I. \quad (9)$$

Let $Y = \sum_{q_i \in S_{(k,s)}} c_i q_i^T$, and its SVD be $Y = U \Lambda^{1/2} V^T$. The solution for the optimal orthonormal transform can be found by $G_k^s = V U^T$. For details of the optimization please refer to [14].

At each step, the sparsity-distortion cost is reduced until the iterations I and II converge to a steady state solution. Figure 2 shows how the proposed optimization approach couples transform ($T$) design with quantization ($Q$). Note the first step of the optimization enforces sparsity of the coefficients, thus mimicking the quantizer. In the next step, transform is updated to have better reconstruction with the new sparser representation.

### 4. IMAGE CODEC

SMTs can be used in many applications from denoising to image reconstruction. In this paper, we concentrate on a prototype image codec to evaluate the performance of the proposed method. Figure 3 shows a simple block diagram of a section of the proposed codec. First a standard (wavelet) transform ($T_H$) is applied to decorrelate the data, which is followed by sparsity-distortion-optimum transform adaptation ($T_{G_1}$ to $T_{G_5}$). Note identity transform is used when standard decorrelation suffices.

In the experiments a 5-level CDF-9/7 wavelet transform is used as the standard decorrelating transform (similar to SPIHT and JPEG2000.) First three levels of the wavelet coefficients with the same sub-band orientation are grouped into sub-trees as shown in Fig. 1. Such a sub-tree corresponds to a region in the original signal domain of size $8 \times 8$. For a given image, the proposed method selects the sparsity-distortion optimal classification similar to a CART-like algorithm [7]. The wavelet coefficients are replaced with the coefficients of new transforms in decreasing order of coefficient energies from coarse to fine scales [14],[17]. Next, the coefficients of the SMT are quantized with a scalar dead-zone quantizer which is followed by entropy coding with a SPIHT-like encoder.

![Fig. 2. Proposed iterative optimization method couples transform design with quantization.](image)

![Fig. 3. Transform and quantization module of the proposed image codec.](image)
5. EXPERIMENTS & DISCUSSIONS

We generated the optimal SMTs off-line, using a fixed training set of 20 images. These images do not contain the test images used for compression performance evaluation in this section. Depending on the local gradient of the image, eight "non-smooth" sub-tree classes are defined for angles ranging from 0 to 157.5 degrees with 22.5 degree increments. In addition to these classes, an identity transform is also added for the "smooth" sub-tree class. During optimization, initial values for the transforms were obtained based on the Karhunen-Loeve transform of the respective class.

In Figure 4, two sample rate-distortion curves are shown for barbara and chair test images. Table 1 shows the PSNR comparison between the two methods for various test images at 0.5 bits per pixel (bpp). Note that SMT-coder consistently produces better PSNR results and obtains up to 1dB improvements.

Table 1. Comparison of the CDF-9/7 and SMT at 0.5 bpp in terms of PSNR(dB).

<table>
<thead>
<tr>
<th>Image</th>
<th>CDF-9/7</th>
<th>SMT</th>
</tr>
</thead>
<tbody>
<tr>
<td>lena</td>
<td>37.22</td>
<td>37.39</td>
</tr>
<tr>
<td>barbara</td>
<td>31.75</td>
<td>32.25</td>
</tr>
<tr>
<td>mandrill</td>
<td>25.66</td>
<td>25.88</td>
</tr>
<tr>
<td>chair</td>
<td>39.68</td>
<td>40.60</td>
</tr>
<tr>
<td>goldhill</td>
<td>33.14</td>
<td>33.36</td>
</tr>
<tr>
<td>museum</td>
<td>37.95</td>
<td>38.59</td>
</tr>
</tbody>
</table>

6. CONCLUSION

This paper presents a sparsity-distortion-optimized multiresolution representation of image geometry. The proposed method uses the wavelet transform followed by a set of orthonormal transforms that are optimized for geometry. The designed orthonormal transforms locally adapt to the signal singularities in wavelet domain and provide a sparser representation. Rather than having a model-based approach, a data-driven training method is used to improve the performance of multiresolution wavelet representation. A new image codec is designed as an application of the proposed method which produces competitive image compression results with the state-of-the-art methods.

7. REFERENCES