A Lifting Based System for Optimal Compression and Classification in the JPEG2000 framework

G. F. Fahmy and S. Panchanathan
Visual Computing and Communications Laboratory
Department of Computer Science and Engineering
Arizona State University
Email: {fahmy, panch@asu.edu}

Abstract
In this paper, we propose a novel design for a lifting based system that achieves the optimal trade off between compression and classification performances. In addition, it can also achieve superior compression performance compared to existing wavelet kernels. The proposed system is based on bi-orthogonal filters and can operate in the JPEG2000 framework. Typically, in our proposed system the trade off point between compression and classification is automatically determined by the system. However, the user can also fine-tune the relative performance if required using two adjustors (one for compression and one for classification). Extensive simulations have been performed to demonstrate the efficiency of our system. Our simulation result shows that a 99% classification performance can be achieved at a comparable reconstruction quality whilst a superior compression performance over the JPEG2000 standard can be achieved at a classification rate of 70%.

Introduction
The rapid growth of visual media in many applications has led to the proliferation of a variety of compression standards (MPEG-4 and JPEG2000). It is therefore likely that visual media will be increasingly stored in a compressed format. Hence there is an impending need for sophisticated compressed domain indexing techniques, where an image/video can be retrieved from a database based on the compressed domain features. Compression and indexing are typically pursued independently. Since both compression and indexing rely on the extraction of content features, they can be treated jointly.

Several joint compression and indexing techniques reported in the literature explored the possibility of using the compression parameters as indices [5] or using the indices to enhance the compression performance [6]. Compression and indexing performances are inversely related, for example a good classification performance is typically obtained at the expense of some compression performance degradations, and vice versa [12]. Classification is an important step in visual indexing. Recently, a number of techniques in the signal processing literature investigated the trade off between compression and classification performances in the encoder/decoder design, by designing a decoder that classifies at a given bit rate or decodes at a specific classification error [10], [11], [12]. Wavelets have become popular in image coding. Most of the existing wavelet image coders employ bi-orthogonal filters. This provides the advantages of selective frequency localization and Perfect Reconstruction (PR) compared to classical orthogonal filters [7]. It also obviates the need for any phase compensation in the pyramid filter structure.

Lifting has emerged as a powerful scheme in bi-orthogonal wavelet reconstruction [4]. It exploits the similarity of the filter coefficients in both low pass and high pass filters resulting in a higher speed of implementation compared to conventional convolution based coders. In addition, lifting inherits the advantages of in-place calculations and reversibility. Lifting has been adopted in many of the recent wavelet based compression standards, namely MPEG-4 and JPEG2000.

In this paper, we propose a novel lifting based joint compression/classification coder in the wavelet domain. A Lagrangian multiplier has been employed to achieve the optimal trade off between compression and classification performances. The proposed system can therefore achieve a superior compression at a given classification rate or the best possible classification for a given compression ratio. The proposed technique is presented in section 2. Details about applying it in the JPEG2000 domain are presented in section 3. Experimental results are demonstrated in section 4, followed by conclusions in section 5.

The Proposed Bi-orthogonal Systems
Bi-orthogonal wavelets are widely used in data compression [8]. It has been shown in [2], that for a bi-orthogonal system to be perfect reconstruction (PR)

\[
\int \phi(t) \overline{\phi}(t-k) dt = \delta(k) \quad \int \psi(t) \overline{\psi}(t-k) dt = 0
\]

where, \(\phi(t)\) and \(\psi(t)\) are the scaling and wavelet functions, and \(\overline{\phi}(t)\) and \(\overline{\psi}(t)\) are their dual functions defined below.

\[
\phi(t) = \sum_{k=0}^{N_1} h_o(k) \phi(2t-k) \quad \psi(t) = \sum_{k=0}^{N_2} g_o(k) \phi(2t-k)
\]

\[
\overline{\phi}(t) = \sum_{k=0}^{N_1} h_i(k) \phi(2t-k) \quad \overline{\psi}(t) = \sum_{k=0}^{N_2} g_i(k) \phi(2t-k)
\]

\(N_1, N_2\) are the orders of \((H_o(z), G_o(z))\) and \((H_i(z), G_i(z))\). \(H_o(z), H_i(z)\) are the low and high analysis filters, while
modifying the high pass filter is effective for texture compression and classification since most of the energy is concentrated in the high frequency hands. We note that the optimal filter coefficients are typically fine-tuned for each input image by minimizing eq.4(c).

\[
\begin{align*}
H^*(z) &= H(z) - S(z^{-1}) H_1(z) \\
G^*_i(z) &= G_i(z) - S(z^{-1}) G_s(z)
\end{align*}
\]

for any arbitrary values of \(S(z)\). However, due to the linear phase restriction that is required in the filter kernels used in most of the image coders, \(S(z)\) has to be symmetric. In other words any bi-orthogonal system \(\{H_1(z), H_2(z), G_1(z), G_2(z)\}\) can be alternatively represented by \(\{H^*(z), H_1(z), G_s(z), G^*_1(z)\}\) based on eq.(3), and vice versa. We propose a modification to the high pass analysis and low pass synthesis filter coefficients \(\{H_1(z), G_2(z)\}\) by changing the \(S(z)\) polynomial while fixing the coefficients of the low pass analysis and high pass synthesis \(\{H_2(z), G_1(z)\}\) filter coefficients in order to obtain a superior compression and classification. For example, modifying the high pass filter is effective for texture compression and classification since most of the energy is concentrated in the high frequency bands. We note that the optimal filter coefficients are typically obtained for either the best compression or classification. We believe that compression and classification have to be treated jointly and hence the optimal trade off between their performances can be obtained. We therefore propose a novel design where the high pass analysis and the low pass synthesis are modified while the other low pass and high pass filters are fixed. This modification can be represented in the form of two adjustors \(A\) and \(B\), which can be fine-tuned to obtain the optimal trade off.

We note here from eq.3, where both the low pass and high pass filters are used to construct the updated high pass filter, that bi-orthogonal systems matches the lifting architecture where one of the resulting filters updates the other, fig.5. The proposed bi-orthogonal system maintains the PR condition Fig.2, and it also increases the regularity of the updated filter.

**Compression and Classification in JPEG2000**

In the JPEG2000 standard [3], the image is transformed into sub-bands and each sub-band is partitioned into a number of blocks, which are coded independently and progressively through several bit-planes. The final embedded block bit-stream is packed into quality layers as shown in Fig.1. This embedded coder generates a bit-stream such that the prefixes in the stream can be decoded to reconstruct the original image. The coder can choose at which point the stream can be truncated. This is the main trade off between bit rate and performance.

Recently in [1], an indexing approach has been developed in the JPEG2000 framework that is based on using the progressive coding scheme as an index. Here a histogram is constructed for each quality layer corresponding to the blocks that contribute in this layer. Classification is based on histogram comparison of the Query and Candidate images histogram for each quality layer. It has been shown that the accuracy of classification depends on the clustering of histograms for each quality layer of all images from the same class. Our proposed system can trade off between optimal compression and classification at the expense of some extra preprocessing. This preprocessing is needed in the \(S(z)\) fine-tuning and if the system decides automatically the desired level of compression/classification performance. In order to achieve optimal compression, the coefficients of \(S(z)\) are fine-tuned until we obtain the minimum MSE of the reconstructed image eq.4(a). In order to achieve optimal classification the coefficients of \(S(z)\) are fine-tuned until the histograms of the quality layers cluster together for each layer for every image class (which will lead to faster and accurate classification) eq.4(b).

We note that compression and classification are typically inversely related Fig.3. Hence, the polynomial \(S(z)\) can be fine-tuned for each input image by minimizing eq.4(c).

\[
R = (x - \beta(x))^2 \quad 4(a)
\]

\[
K = \sum_{j=0}^{M-1} I(\delta(\beta(x)) \neq j) P(Y = j) P(X = x) \quad 4(b)
\]

\[
J = R + \lambda K \quad 4(c)
\]

Where \(\alpha(\cdot)\) is the encoder, \(\beta(\cdot)\) is the decoder, \(\delta(\cdot)\) is the classifier that classifies the processed input into one of \(M\) predefined classes, \(x\) is the input data and \(Y\) is the overall vector space, \(\lambda\) is the Lagrangian multiplier that controls the trade off between classification and compression. We note here that the value of \(\lambda\) determines the rate of classification desired from the compressed bit stream. If \(\lambda\) is set to zero, superior compression is achieved, while the classification rate is marginal. If \(\lambda\) is set to one, superior classification can be obtained, while maintaining comparable compression. A block diagram for the proposed lifting based system is shown in fig.5, where the first adjustor controls the compression performance for the coder, and the second adjustor controls the indexing performance from the resulted compressed bit-stream.

**Experimental Results**

Simulations have been performed using 16 classes belonging to VisTex database. Each class contains 10 to 12 different images with different databases in the database.
Images were compressed using the JPEG2000 standard. Histograms were formed for each quality layer in the compressed bit stream for each of the 16 classes as in [1]. In order to achieve superior compression, the value of $\lambda$ was set to zero. $\phi(z)$ is fine-tuned in eq.3, until $J$ becomes minimum in eq.4(c), by adjusting switch $A$ in fig.5. The resulting high pass and low pass filters were used in the JPEG2000 domain for optimal compression. Table 1 compares the performance of the reconstructed images obtained using our optimized lifting system when $\lambda =0$, and the widely used 9X7, 10X18, 5X3 kernels that are provided by the JPEG2000 standard, from the perspective of compressed bit stream size (with a fixed PSNR of 35db). It can be shown that our proposed system achieves superior compression performance, fig.6.

We note here that these improvements are obtained with highly textured images that contain most of their information in high frequency regions, because of the high pass coefficient modification that is performed in the optimization process. Fig.3 plots the average compression performance and classification performance (for 16 classes) for different values of $\lambda$. Changing the value of $\lambda$ from 0 to 1, by adjusting switch $B$ in fig.5, will increase the classification rate, however it will degrade the compression as shown in fig.3. Fig.4(a) shows the histograms of the first quality layer for 4 images of the same class (Bark class). Fig.4(b) shows how these histograms are clustered after our filter optimizations for a value of $\lambda =0.6$.

We also note that our classification performance is evaluated based on the number of images that are classified correctly to each class, to the total number of images in that particular class. Compression performance is evaluated based on the compression achieved for each class, to the best compression achieved for this particular class (with fixing the PSNR). In order to achieve the optimal compression and classification, point $G$ in fig.3, both switches $A$ and $B$ should be adjusted jointly. Table 2 shows the classification rate obtained by setting $\lambda =0.3$, 0.6 and 0.7 which corresponds to substantial improvements in classification performance. We note here that the value of $\lambda$ can be adjusted automatically by the system if the input image class is predetermined and specific classification is targeted. It can be seen from fig.3 that when compression is degraded, classification is enhanced (with increasing the value of $\lambda$) and vice versa. Table 3, shows a full comparison in terms of the overall compression and retrieval time between the three filters employed by the JPEG2000 standard, and the proposed optimized lifting coefficients for a value of $\lambda =0.5$. It can be shown that our proposed system does not require significant complexity.

**Conclusion**

In this paper, we have presented a lifting based system for joint compression and classification. The proposed system has two adjustor switches, one for compression and one for classification, that control the trade off between them. This system maintains the perfect reconstruction property of the coding system and improves the regularity of the generated filters. When the classification factor is cancelled, our proposed system generated bi-orthogonal lifting kernels that achieved superior compression results to the existing kernels defined in the JPEG2000 standard. With a classification result of over 99%, we have been able to get comparable compression results with the existing compression kernels. Our system can help the user in choosing the optimal coding filters for both compression and classification.

**References**

### Table 1

<table>
<thead>
<tr>
<th>Class</th>
<th>9x7 Kernel</th>
<th>10x18 Kernel</th>
<th>5x3 Kernel</th>
<th>Our kernel ( \lambda = 0.0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fabric</td>
<td>0.48</td>
<td>0.44</td>
<td>0.51</td>
<td>0.41</td>
</tr>
<tr>
<td>Grass</td>
<td>0.55</td>
<td>0.45</td>
<td>0.52</td>
<td>0.43</td>
</tr>
<tr>
<td>Bark</td>
<td>0.51</td>
<td>0.52</td>
<td>0.54</td>
<td>0.48</td>
</tr>
<tr>
<td>Sand</td>
<td>0.45</td>
<td>0.46</td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td>Metal</td>
<td>0.55</td>
<td>0.53</td>
<td>0.56</td>
<td>0.52</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Class</th>
<th>( \lambda = 0.3 )</th>
<th>( \lambda = 0.5 )</th>
<th>( \lambda = 0.7 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fabric</td>
<td>98.5</td>
<td>98.9</td>
<td>99.3</td>
</tr>
<tr>
<td>Clouds</td>
<td>98.4</td>
<td>98.6</td>
<td>98.9</td>
</tr>
<tr>
<td>Bark</td>
<td>98.2</td>
<td>98.6</td>
<td>98.2</td>
</tr>
<tr>
<td>Sand</td>
<td>98.8</td>
<td>98.5</td>
<td>99.1</td>
</tr>
<tr>
<td>Water</td>
<td>98.2</td>
<td>98.3</td>
<td>97.9</td>
</tr>
</tbody>
</table>

### Table 3

<table>
<thead>
<tr>
<th>Class</th>
<th>9x7 Kernel</th>
<th>10x18 Kernel</th>
<th>5x3 Kernel</th>
<th>Our kernel ( \lambda = 0.5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fabric</td>
<td>0.40 S</td>
<td>0.38 S</td>
<td>0.42 S</td>
<td>0.56 S</td>
</tr>
<tr>
<td>Grass</td>
<td>0.37 S</td>
<td>0.37 S</td>
<td>0.41 S</td>
<td>0.55 S</td>
</tr>
<tr>
<td>Bark</td>
<td>0.43 S</td>
<td>0.42 S</td>
<td>0.44 S</td>
<td>0.37 S</td>
</tr>
<tr>
<td>Sand</td>
<td>0.59 S</td>
<td>0.37 S</td>
<td>0.40 S</td>
<td>0.53 S</td>
</tr>
<tr>
<td>Metal</td>
<td>0.43 S</td>
<td>0.41 S</td>
<td>0.45 S</td>
<td>0.55 S</td>
</tr>
</tbody>
</table>

### Fig. 2
The loss in the sub-band system, Group delay, using the JPEG2000 9x7 (red) and our kernel (blue).

### Fig. 3
A flow graph for the compression and classification performance on one graph. Point G represents the optimal point between them.

### Fig. 4
The histogram for the first quality layer in the JPEG2000 format for 4 images that belong to the same class. (a) With the regular JPEG2000 coder (b) With our proposed system.

### Fig. 5
The proposed lifting based system with two switches A (for compression) and B (for classification).

### Fig. 6
(a) Original Grass Image (b) Grass Image compressed with our optimized kernel \( \lambda = 0.0 \) (PSNR = 38) (c) Grass Image compressed with the JPEG2000 9x7 kernel (PSNR 35)

IV - 156