

## Wavelet domain textual coding of Ottoman script images

Ömer N. Gerek<sup>1</sup>, A. Enis Çetin<sup>1</sup>, Ahmed H. Tewfik<sup>2</sup>

<sup>1</sup> Bilkent University, Dept. of Electrical and Electronics Engineering,  
Bilkent, Ankara TR-06533, Turkey  
E-mail: gerek@ee.bilkent.edu.tr Phone: (90) 312-266 4307 Fax: (90) 312-266 4126

<sup>2</sup> Dept. of Electrical Engineering, University of Minnesota  
EE/CSci Building, Minneapolis, MN 55455, USA

### **ABSTRACT**

Image coding using Wavelet Transform, DCT and similar transform techniques is well established. On the other hand, these coding methods neither take into account the special characteristics of the images in a database nor are they suitable for fast database search. In this paper, the digital archiving of Ottoman printings is considered. Ottoman documents are printed in Arabic letters. In [1], Witten et al. describes a scheme based on finding the characters in binary document images and encoding the positions of the repeated characters. This method efficiently compresses document images and is suitable for database search, but it cannot be applied to Ottoman or Arabic documents as the concept of character is different in Ottoman or Arabic. Typically, one has to deal with compound structures consisting of a group of letters. Therefore, the matching criterion will be according to those compound structures. Furthermore, the text images are gray tone or color images for Ottoman scripts for the reasons that will be described in the paper. In our method the compound structure matching is carried out in wavelet domain which reduces the search space and increases the compression ratio. In addition to the wavelet transformation which corresponds to the linear subband decomposition, we also used nonlinear subband decomposition. The filters in the nonlinear subband decomposition have the property of preserving edges in the low resolution subband image.

**Keywords :** Textual Image Coding, Document Imaging, Image Databases, Wavelet Transforms.

### **1. TEXTUAL IMAGE ARCHIVING AND COMPRESSION**

An Ottoman document image mainly consists of printed text and some marks and drawings. There can also be gray tone and color images inside the document. Usually, the marks, shady areas and ink smears in the page are important for a historian. As a result, the scanned image should be kept in gray tone or color format for Ottoman archives.

A number of authors studied the case of binary textual image compression [1] - [6]. Most of the available methods exploit the characteristics of textual images which are different than arbitrary binary images. A text image contains repetitions of small character images, i.e. letters. Exploiting



Figure 1: Part of the original document image

the redundancy of these repetitions is the key step in most of the textual image coding algorithms. A good approach to take advantage of this redundancy is to encode the repeated character images and their locations. This method efficiently compresses the textual image and it is appropriate for fast database search. Since the character images are preserved, the keyword search is available via individual characters and their locations.

The procedure of textual image coding can be described in a sequence as:

- 1) Find and extract a mark in the image,
- 2) add it to the library constructed by these mark images,
- 3) find the locations of the marks that are similar to the extracted one inside image, and remove those repetitions from the image,
- 4) go to 1 until all marks in the image are deleted,
- 5) compress (i) the constructed library and (ii) the symbol locations.

This operation is illustrated in Fig. 1. The repetitions of letter “waw” are found. Note that the letter can be connected to other compound structures.

A further sixth step in this procedure is proposed in [1] to encode the residue image and perform a lossless compression.

In this paper, we will describe how this procedure can be applied to wavelet transformed textual images. The transformed images have smaller size, so the advantages of using smaller sized images in terms of speed will be exploited. The compression results for directly using this procedure will be compared with wavelet transformed and nonlinear subband decomposed images.

## **2. IMAGE CODING USING SUBBAND DECOMPOSITION**

Image coding using subband decomposition or wavelet transform has been widely investigated for gray-tone and color images [7] - [10]. By using the relation between subband coding and wavelet transformation [11] - [13], the wavelet transform is implemented with a perfect reconstruction filter bank in practice. The idea is to divide the image data into subbands corresponding to different frequency contents. Let us assume that  $H_0(\omega)$  and  $H_1(\omega)$  are the low-pass and high-pass filters of a perfect reconstruction filter bank, respectively. In 1-D case with one level decomposition, the input signal  $x[n]$  is filtered by  $h_0[n]$  and  $h_1[n]$  and the resultant signals are down-sampled by a factor of two. In this way two sub-signals  $x_0[n]$  and  $x_1[n]$  are obtained, i.e.,

$$x_i[n] = \sum_k h_i[k]x[2n - k], \quad i = 0, 1 \quad (1)$$

The sub-signals  $x_0[n]$  and  $x_1[n]$  contains the low-pass and the high-pass information of the signal  $x[n]$ . The synthesis part of this decomposition is performed by the complementary synthesis filters after upsampling stage [7] - [14].

In 2-D case, the same one level decomposition results in four subband images  $ll$ ,  $lh$ ,  $hl$ , and  $hh$ . The subband image  $ll$  is obtained by first lowpass filtering the image in horizontal direction and then lowpass filtering in the vertical direction,  $lh$  is obtained by first lowpass filtering the image in horizontal direction and then highpass filtering in the vertical direction, and so on.

Different frequency content and characteristics of the subband images enable us to treat each subband image differently. For most of the gray tone images, the corresponding subband images can be quantized without introducing any perceptual degradation. After the quantization step, the data set in each subband can be encoded separately by using the most suitable encoding strategy for the subband image.

In our experiments, we used the simplest FIR perfect reconstruction filter bank :

$$h_0[0] = h_0[1] = 1/2, \quad h_0[n] = 0 \quad \text{for } n \neq 1, 2 \quad (2)$$

$$h_1[n] = (-1)^n h_0[n] \quad (3)$$

This filter bank corresponds to Haar Wavelet Transform [14] and it has a number of advantages in our application. The time localization of this filter pair is very good, so the edges of the edge locations

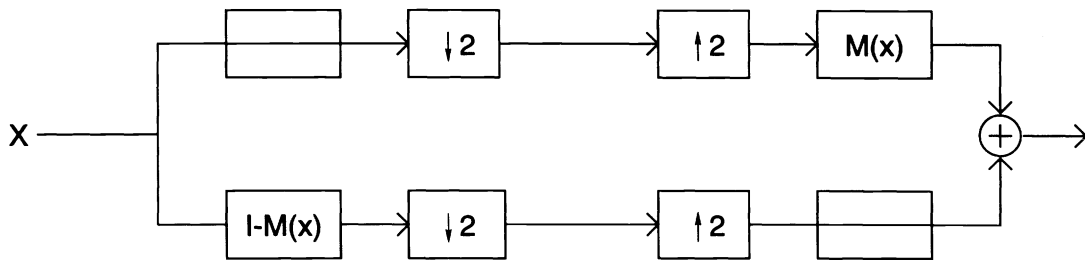


Figure 2: One stage nonlinear subband decomposition

inside the image do not get blurred and there is no ripple effect at the edges. Furthermore it is easy to implement and fast to perform.

There is another subband decomposition scheme that is being investigated. The scheme is based on the work by Egger et al [15]. In this work, the decomposition filter bank consists of non-linear filters instead of the standard perfect reconstruction linear filters. The general one stage, one dimensional analysis and synthesis operation is illustrated in Fig. 2. Specifically, the nonlinear operation that we use is the median filtering. Egger et al propose 6, 8, ... pixels for the median calculation, however, we used odd number of pixels for this purpose to come over the quantization problem. Furthermore, the subtraction operation between the interpolated pixel and the true pixel value is performed in modulo-256. In this way, the overall scheme is again perfect reconstruction, and this time there is no need for quantization. The nonlinear subband decomposition filters have the property of halfband filters and they are good at preserving the edges inside an image [15]. Since the textual images have a lot of sharp edges, the nonlinear subband decomposition yields better compression and better visual results.

### **3. TEXTUAL IMAGE COMPRESSION IN WAVELET DOMAIN**

Our approach to the textual image coding problem is to combine the wavelet transformation and the textual image compression techniques. Decomposition of the Ottoman script image into subbands brings a number of advantages in encoding and decoding. In order to acknowledge fast to a query, a low resolution image of the document can be supplied to the end user for a fast preview.

With this motivation, the steps in Section 1 are carried out in the  $ll$  subband of the original image. The decoding can start from the  $ll$  image to supply the low resolution preview image this way. The procedure starts with determining every mark in the  $ll$  image and checking for the repetitions inside the same image. After this step, the places of occurrence and the boundaries of the marks obtained from the  $ll$  image can be used for all other subbands (call Method 1). In other words, the search algorithms are not carried out in any subband image other than the  $ll$  image. We found that the described search method is not suitable for extracting the characters from  $lh$ ,  $hl$  and  $hh$  images.

At the end of this process, four library of symbols are generated corresponding to the outcomes

of the operations in the  $ll$  subband image. These subband library (SL) images constitute the subband decomposition of the original library (OL) of symbols which can be obtained by using the original text image and the same pattern matching method (call Method 2). These SL images have one fourth the size of the OL image.

Since the same method is used for encoding the locations of the repeated characters (step 5.ii) in Method 1 and Method 2, the amount of bits required for encoding the locations for Method 2 is larger. This is due to the size of the page to be encoded. The size of the original page is four times greater than the size of the subband images, so the location coordinates for Method one is twice larger. The amount of bytes for encoding the locations of the repeated characters (step 5.ii) in Method 1 is about 1.4 times smaller than the amount of bytes for encoding the locations for Method 2.

By the elimination of the repeated character or compound structure images from each subband image, we have four subband residue images which may be coded as a last step for achieving lossless compression [1].

After these operations, the compression efficiency is mainly determined by how successful the SL images are compressed. Experiments show that the subband library images can be more efficiently compressed.

Another good point for using the subband images is about the computation time during encoding. The experiments show that most of the computation time is spent in finding the repetitions of a character or a compound structure image inside the text image. We take the character image and slide it over the whole text image to find the sum of the absolute difference at every location. In this way, a matching error matrix  $E(i, j)$  is constructed. Our matching criterion is based on a combination of absolute error between the pixel values at each location and a constraint on the sizes of the compound structures. This  $E$  matrix thus has high and low values of the described cost function. The points corresponding to the local minima of  $E$  matrix are then tested for being detected as the searched compound structure image. This test is necessary for removing the errors that occur when the pixel-wise error is small, but the letter sizes do not match. Typically, when one compound structure is entirely included inside another compound structure, the pixel-wise error is small, but the sizes do not match. If the included compound structure is at the end or at the beginning of the other compound structure, that location should be marked as a repetition of the compound structure.

The time required for computing the  $E$  matrix decreases approximately by a factor of 10 when the size of the character image and the size of the text image are one fourth of the original image (i.e. the width and the height of the images are half of the originals).

#### **4. SIMULATION STUDIES**

The compressed image consists of bit-streams corresponding to the character library images, the locations of the characters, and their symbol sequences.

Experimentally, the compression ratio for separate coding of each of the four SL images was

inferior to the compression ratio for the lossless coding of the OL image. We used a  $1300 \times 1900$  textual Ottoman script with 8 bits per pixel which was scanned at 300 dpi. When we do not consider the coding of the residue images, the lossless compression of the OL image results in an overall compression ratio (CR) of 20.02 : 1. On the other hand, the lossless compression of four SL images generated from Haar basis results with an overall CR of 18.13 : 1. By simple quantization, this ratio can be increased to 23.77 : 1 without introducing a perceptual difference.

When we consider the nonlinear subband decomposition, the compression of four SL images results with an overall CR of 24.00 : 1.

Instead of compressing the SL and OL library images in a lossless manner, a lossy scheme like JPEG can be used. The results for JPEG encoding the library images are as follows :

- For direct OL images, total CR = 42.44 : 1
- For Haar subband decomposition, total CR = 40.78 : 1
- For Haar subband decomposition + quantization, total CR = 48.41 : 1
- For nonlinear subband decomposition, total CR = 50.19 : 1

In order to get the advantage of subband decomposition, one can exploit the correlation between the subband images. Since the document images mostly have regions of black pixels and a gray tone background, the subband images including the *ll* image are highly correlated with each other.

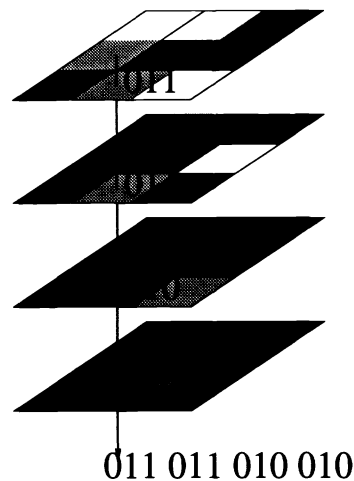


Figure 3: Visualization of appending the bit-planes of subband images

If the bit planes of subband images are to be realized at top of each other, they form 3-D data. The total number of bit planes increases this way. In order to keep the total number of bit planes reasonable, the SL images are quantized to less number of bits. By taking into account the variances of the subband images, 5 bits are assigned to *ll* subband image, 3 bits are assigned to *lh* and *hl* images and 2 bits are assigned to *hh* image. The appended image becomes a 13-bit image this way. Although



Figure 4: Detail images to show the pixel-wise correlation between subbands

this corresponds to  $2^{13}$  gray levels, the total number of unique gray tones in this image is only 230. The reason for this is the flatness of the regions inside the textual image. Actually there are not many gray tones inside the textual image. The appending scheme is illustrated in Fig. 3 and the correlation between subband library images is illustrated in Fig. 4.

When the appended library image is Lempel-Ziv compressed, the overall CR for the document is 59.74 : 1 for Haar subband decomposition and 60.10 : 1 for nonlinear subband decomposition. The residue images are again not coded.

The residue images are rather difficult to compress. The lossless encoding of the residue images decreases the compression ratio drastically. Two approaches for lossy encoding the residue images are JPEG compression and Lempel-Ziv encoding the quantized version of the residue image. The result for JPEG was visually better than the result for Lempel-Ziv encoding the quantized version of the residue image. In this way, the compression ratio is decreased because of the extra bits required for the residue images. This is the case for all the methods described above.

The final compression ratio results are given in Table 1 and 2. The portion of the original and decoded image after using Haar subband decomposition and bit-plane appending the SL images are given in Fig. 7. The residue images are JPEG encoded in this reconstructed image.

## 5. CONCLUSIONS

We presented a method to compress Ottoman document images which are composed of Arabic letters. In this method, the first step is subband decomposition. The next step is adaptively building the compound structure library by extracting the patterns corresponding to a compound structure and searching for a match in the image. The low resolution image obtained via subband decomposition is used for the search and construct algorithm. Naturally, the repeated compound structures cannot be identical throughout the image because of the printing and scan noise. In order to obtain the image in its original form, the irregularities and the natural gray tones must be retained, so the patterns with their locations inside the image and the residue image are encoded as well.

The matching criteria for finding the repetitions of the compound structures is based on a modified version of "sum of the absolute differences" at every location. The weights of absolute differences are adjusted according to the distance from the centroid of the compound structure. After this step, the matching points are tested for being entirely included in another compound structure (Fig. 5). If the inclusion is from the left or from the right, the position is considered to be a match. If the inclusion is in the middle of the other compound structure, the removal of the small compound

structure causes the larger one to be broken into two, which in turn decreases the coding efficiency and disturbs the possibility of keyword search.

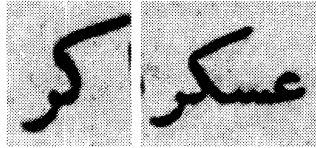


Figure 5: Two compound structures.

While extracting the library pattern from the text image, the background level of the text image is preserved. As a result mostly the background gray tone remains in the residue image and the dynamic range decreases. Compression of this residue image is more efficient than when the background levels are ignored and substituted with white pixels.

Pattern matching for constructing the pattern library is carried out over the low resolution images. These pattern images and their locations supply the necessary information to extract the *lh*, *hl* and *hh* image counterparts. Basically, the exact places and dimensions of these patterns are used for constructing the high-band pattern libraries.

The decoding procedure starts from the subband images. The library sub-images are inserted to the locations in the pointer list and the decoded residue images corresponding to each subband image are added to obtain a good approximation to the subband decomposition of the original image. These subband images are still approximations, because the residue images as well as the library images are encoded in a lossy manner.

The last step is synthesizing the sub-band images by the synthesis filter bank of the wavelet transformation or the nonlinear subband decomposition.

This method is suitable for fast database search if the properties of each extracted compound structure are supplied. The possibility of a compound structure to be concatenated with another compound structure is a grammatic property, and this possibility for each extracted compound structure must be stored inside the coded bit-stream in order to enable keyword searches.

## **6. ACKNOWLEDGMENTS**

This research is supported by NSF and TUBITAK EUREKA-625.



افراد آئینور و بعض یرلردن همان  
وره د هم نظامسرلنی وهم زراعت  
عب اولدینی مثللو عسکر لکه کالانلرک  
یری وقطع تناسلی مستلزم اولمقده  
یت جلب ومدت استخدا منه دائر  
نک وضعی لایب ایدی .  
فریضه ذمتی ایسهده وقت حضرده

Figure 6: Part of the original document image

افراد آئینور و بعض یرلردن همان  
وره د هم نظامسرلنی وهم زراعت  
عب اولدینی مثللو عسکر لکه کالانلرک  
یری وقطع تناسلی مستلزم اولمقده  
یت جلب ومدت استخدا منه دائر  
نک وضعی لایب ایدی .  
فریضه ذمتی ایسهده وقت حضرده

Figure 7: Reconstructed image

	original image		Haar subband decomposition				
	lossless OL	JPEG OL	lossless SL	quantized SL	JPEG SL	JPEG of quantized SL	Bit-plane appended SL
Number of bits for library image	970216	451000	1077904	819304	472552	396184	318768
CR w/o residue	20.02	42.44	18.13	23.77	40.78	48.41	59.74
CR with JPEG residue	16.65	29.59	15.08	18.80	28.05	31.46	35.88

Table 1: Compression results - part 1.

	nonlinear subband decomposition				
	lossless SL	quantized SL	JPEG SL	JPEG of quantized SL	Bit-plane appended SL
Number of bits for library image	1033194	811336	445200	381704	316784
CR w/o residue	18.96	24.00	43.22	50.19	60.10
CR with JPEG residue	15.62	18.94	29.18	32.20	36.01

Table 2: Compression results - part 2.

## References

- [1] Ian H. Witten, Timothy C. Bell, Hugh Emberson, Stuart Inglis, and Alistair Moffat, "Textual Image Compression: Two-Stage Lossy/Lossless Encoding of Textual Images," *Proceedings of the IEEE*, Vol. 82, No.6, June 1994.
- [2] W. K. Pratt, P. J. Capitant, W. H. Chen, E. R. Hamilton, and R. H. Wallis, "Combined symbol matching facsimile data compression system," *Proc. IEEE* vol. 68, no. 7, pp. 786-796, July 1980.
- [3] M. J. J. Holt, "A fast binary template matching algorithm for document image data compression," in *Pattern Recognition*, J. Kittler Ed., Berlin, Germany, Springer Verlag, 1988.

- [4] O. Johnsen, J. Segen, and G. L. Cash, "Coding of two-level pictures by pattern matching and substitution," *Bell Syst. Tech. J.*, vol. 62, no. 8, pp. 2513-2545, May 1983.
- [5] A. Moffat, "Two level context based compression of binary images," *Proc. IEEE Data Compression Conf.*, J. A. Storer and J. H. Reif Eds. Los Alamitos, CA; IEEE Computer Society Press, pp.382-391, 1991.
- [6] R. N. Ascher and G. Nagy, "A means for achieving a high degree of compaction on scan-digitized printed text," *IEEE Trans. Comput.*, Vol. C-23, No. 11, pp. 1174-1179, Nov. 1974.
- [7] E. H. Adelson, E. Simoncelli, and Hingorani, "Orthogonal pyramid transforms for image coding," *Proc. SPIE Conf. VCIP*, pp. 50-58, Cambridge, MA, 1987.
- [8] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transforms," *IEEE Trans. ASSP*, 1991.
- [9] J. C. Feauveau, P. Mathieu, M. Barlaud, and M. Antinoni, "Recursive biorthogonal wavelet transform for image coding," *Proc. ICASSP'91*, pp.2649-2652, Toronto, Canada, 1991.
- [10] M. Vetterli, J. Kovačević and Le Gall, " Perfect reconstruction filter banks for HDTV representation and coding," *Image Communication 2*, pp.349-364, 1990.
- [11] Y. Meyer, *Ondelettes et Opérateurs*, Hermann, 1988.
- [12] I. Daubechies, *Orthogonal bases of compactly supported wavelets*, *Commun. Pure and Applied Math*, Vol. XLI, pp. 909-996, 1988.
- [13] S. G. Mallat, *A Theory for Multiresolution Signal Decomposition: The Wavelet Representation*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol II, no 7, pp. 674-693, July 1989.
- [14] J. W. Woods, Ed., *Subband Image Coding*, Kluwer, 1991.
- [15] O. Egger, W. Li, and M. Kunt, "High Compression Image Coding Using an Adaptive Morphological Subband Decomposition," *Proc IEEE*, vol. 83, no. 2, pp.272-287, February 1995.