Hybrid System for Speech Coding Based on Contourlet Transformation

Khalil I. Alsaif   Hadia S. Albadrani

Hadiasalih@uomosul.edu.iq
College of Computer Sciences and Mathematics
University of Mosul, Mosul, Iraq

ABSTRACT

The main objective of speech coding is to allow the transmission of speech over digital channel of the highest speech quality and least possible bit rate, beside the security purpose.

In this paper, the speech was coded by transforming it applying (which is often single dimension) into a two dimensional array to be suitable for transferring the countourlet transformation. Applied EZC (Embedded Zero tree Contourlet) algorithm, then is applied to the Huffman coding on the results of EZC, and used RLE (Run Length Encoding).

The above idea gave the ability for coding, compression with retrieved information of high accuracy, using some measurements for quality measured of reconstructed signal, and found results show high similarity between the original and reconstructed signal.

Keywords: Contourlet Transform, Embedded Zero tree Contourlet.

نظام هجين لترميز الصوت باستخدام التحويل الكنتوري

خليل إبراهيم السيف   **هدية البدراني**

كلية علوم الحاسوب والرياضيات
جامعة الموصل، الموصل، العراق

الملخص

إن الهدف الرئيسي من ترميز الصوت هو السماح لنقله على قنوات رقمية وبجودة عالية ومساحة خزنيه اقل إضافة إلى الأغراض الأمنية.

في هذا البحث تم تشفير الكلام (الذي هو في الغالب ذو بعد واحد) حيث تم تحويله إلى مصفوفة ثنائية ليصبح أكثر ملائمة إلى التحويلات الكنتورية. تم استخدام خوارزمية الشجرة الصفرية الكنتورية المضمنة, وبعدها يتم تنفيذ الـ Huffman coding للنتائج التي تم الحصول عليها من خوارزمية الشجرة الصفرية الكنتورية المضمنة، ومن ثم تنفيذ الـ Run Length Encoding.

إن الفكرة أعلاه أعطت قدرة على التشفير وكبس واسترجاع المعلومات بكفاءة عالية. تم استخدام عدد من المقاييس لقياس جودة الإشارة المسترجعة,وأظهرت النتائج تقارباً عالياً بين الإشارة المسترجعة والإشارة الأصلية.

الكلمات المفتاحية: تحويل كونتوري، الشجرة الصفرية الكنتورية المضمنة.

1- Introduction

The digital signal is a sequence of binary bits; it is more reliable to represent these bits by a special code in a few bits. Few bits mean less memory from the point of view of storage, and less bandwidth and bit rate from the point of view of transformation. In spite of the wide bandwidth in the wired communication systems, but in the wireless, it is still limited, so the need to transform the signal by small number of bit per sample. Beside these reasons, the cost, the delay, and complexity are other reasons. If less memory is utilized, the cost becomes less with lower bit rates and the required channel capacity is also reduced [1]. For coding used contourlet transform.

In spite of the speech is one dimension, but the contourlet is applied with it, then applied EZC, then used Huffman coding on the results of EZC, and then used RLE, and the results are good given.

2- Principles of EZC

A new method has been used depending on the Contourlet Transform for compression. The method is EZC which depends on the Embedded Zeros Tree for Contourlet (ETC). The relationship between parent and child for Contourlet Transform coefficients is similar to that between Wavelet coefficients. In the Contourlet Transform, one can assume two different parent-child relationships depending on the number of directional decompositions in the contourlet subbands. The parent lay, the four children will be in two adjacent directional subbands as shown in Figure (1) [2].
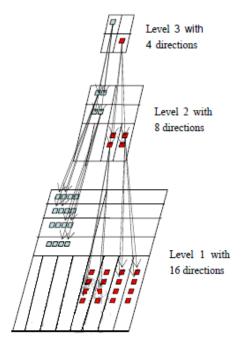

Figure (1). Parent-Child Relationships in the Contourlet Transform (CT).

3- ETC or EZC algorithm

The following steps represent all the principles of ETC

- First step: The Contourlet Transform coefficients of the entire 2-D image are created.
- Second step: Progressively ETC are made on the coefficients.
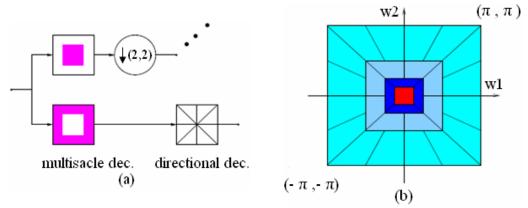 All the coefficients are scanned from top to down

– If the leaves are less than root, they will be marked to be zero.
– If the leaves are greater than root, they will stay as before without any change.
• Third step: Huffman coding is applied [3].

4- Contourlet Transform

The contourlet transform is one of the new geometrical image transforms, which can efficiently represent images containing contours and textures. This transform uses a structure similar to that of curvelets, that is, a stage of sub band decomposition followed by a directional transform. In the contourlet transform, a Laplacian pyramid (LP) is employed for the first stage, while directional filter banks (DFBs) are used in the angular decomposition stage.
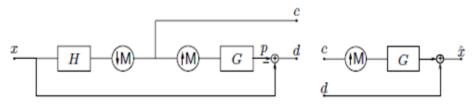
The contourlet transform is constructed as a combination of the Laplacian pyramid and the directional filter banks (DFB). Conceptually, the flow of operation as shown in Figure (2), where the Laplacian pyramid iteratively decomposes a 2-D image into low pass and high pass sub bands, and the DFBs are applied to the high pass sub-bands to further decompose the frequency spectrum [4].

Figure (2). The Original Contourlet Transform. (A) Block Diagram. (B) Resulting Frequency Division

5- Pyramid frames

One way to obtain a multiscale decomposition is to use the Laplacian pyramid. The LP decomposition at each level generates a down sampled low pass version of the original and the difference between the original and the prediction, resulting in a band pass image. As shown in Figure (3) depicts this decomposition process, where H and G are called low pass analysis and synthesis filters, respectively, and M is the sampling matrix. The process can be iterated on the coarse (sampled down) signal (Low frequencies) . Note that in multidimensional filter banks, sampling is represented by sampling matrices; for example, down sampling   x[n] by M yields xd[n] = x[Mn], where M is an integer matrix [4].

(a) One level of decomposition   **(b) The new reconstruction scheme**
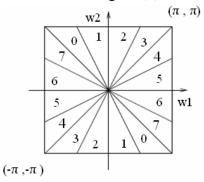**Figure (3).** Laplacian pyramid
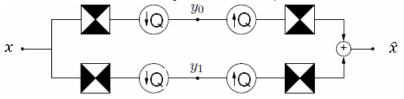
6- Iterated Directional Filter Banks

Bamberger and Smith constructed a 2-D directional filter bank (DFB) that can be maximally decimated while achieving perfect reconstruction [5]. The DFB is efficiently implemented via an l-level binary tree decomposition that leads to 2^l sub bands with wedge-shaped frequency partitioning as in Figure (4).The original construction of the DFB involves modulating the input image and using quincunx filter banks with diamond-shaped filters [6].

To simplify the analysis of the iterated DFB, propose a new formulation for the DFB that is based only on the QFB's with fan filters. The new DFB avoids the modulation of the input image and has a simpler rule for expanding the decomposition tree [7].

This simplified DFB is intuitively constructed from two building blocks. The first building block is a two-channel quincunx filter bank with fan filters, as in Figure (5), that divides a 2-D spectrum into two directions: horizontal and vertical. The second building block of the DFB is a shearing operator. Figure (6) shows an application of a shearing operator, where a −45◦ direction edge becomes a vertical edge. By adding a pair of shearing operator and its inverse ("unshearing") before and after respectively, a two channel filter bank, as in Figure (5), obtains a different directional frequency partition while maintaining perfect reconstruction. Thus, the key in the DFB is to use an appropriate combination of shearing operators together with two-direction partition of quincunx filter banks at each node in a binary tree-structured filter bank, to obtain the desired 2-D spectrum division as shown in Figure (6).

Figure (4). Directional filter bank. Frequency partitioning where l = 3 and there are $2^3 = 8$ real wedge-shaped frequency bands. Sub-bands 0–3 correspond to the *mostly horizontal* directions, while sub bands 4–7 correspond to the *mostly vertical* directions

Figure (5). Two-dimensional spectrum partition using quincunx filter banks with fan filters. The black regions represent the ideal frequency supports of each filter. Q is a quincunx sampling matrix

Quincunx matrix

$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

(a) The "cameraman" image   (b) After a shearing operation

Figure (6). Example of shearing operation that is used like a rotation operation for DFB decomposition

The following four basic unimodular matrices are used in the DFB in order to provide the equivalence of the rotation operations:

$$R_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad R_1 = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix},$$

$$R_2 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \qquad R_3 = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}.$$

Note that $R_0R_1 = R_2R_3 = I_2$ (here I_2 denotes the 2×2 identity matrix). So, for example, up sampling by R_0 is equivalent to down sampling by R1.

A useful tool in analyzing multi-dimensional multirate operations is the Smith form that can diagonalize any integer matrix M into product UDV where U and V are unimodular integer matrices and D is an integer diagonal matrix. The quincunx matrix in:

$$Q_0 = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \qquad Q_1 = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

can be expressed in the Smith form as:

$$Q_0 = R_1D_0R_2 = R_2D_1R_1 \quad \text{and}$$
$$Q_1 = R_0D_0R_3 = R_3D_1R_0,$$

$$D_0 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad D_1 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

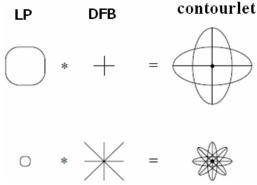Where

are two 2-D diagonal matrices that correspond to dyadic sampling in each dimension [7, 8].

7- Multiresolution Analysis

As for the wavelet filter bank, the iterated Pyramid Directional Filter Bank (PDFB) can be associated with a continuous domain system, called Contourlet. This connection will be made precise by studying the embedded grids of approximation as in the multi resolution analysis for wavelets. The new elements are of multiple directions and they combine with multiscale as in Figure (7) [7, 9].

Figure (7). Illustration of the contourlet basis images that satisfy the curve scaling relation. From the upper line to the lower line, the scale is reduced by four while the number of directions is doubled.

8- Multiscale

Suppose that the LP in the PDFB uses orthogonal filters and down sampling by two is taken in each dimension. Under certain conditions, the low pass filter G in the LP uniquely defines an orthogonal scaling function $\phi^{(t)} \in L^2(R^2)$ via the two-scale equation:

$$\phi(t) = 2 \sum_{n \in Z^2} g[n] \, \phi(2t - n).$$

Where $g[n]$ is the impulse response of the low pass synthesis filter G Let:

$$\phi_{j,n} = 2^{-j} \phi\left(\frac{t - 2^j n}{2^j} \right) , \qquad j \in Z, n \in Z^2.$$

Then the family $\{\phi_{j,n}\}_n \in Z^2$ is an orthogonal basis of V_j for all $j \in Z$. The sequence of nested subspaces $\{V_j\}_{j \in Z}$ satisfies the following invariance properties:

Shift invariance: $f(t) \in V_j \iff f(t - 2^j k) \in V_j, \quad \forall j \in \mathbb{Z}, k \in \mathbb{Z}^2$

Scale invariance: $f(t) \in V_j \iff f(2^{-1} t) \in V_{j+1}, \quad \forall j \in \mathbb{Z}.$

In other words, Vj is a subspace defined on a uniform grid with intervals $2^j \times 2^j$, which characterize the image approximation at the resolution 2^{-j}. The different image in the LP carries the details necessary to increase the resolution of an image approximation. Let W_j be the orthogonal complement of Vj in V_{j-1}. Also see Figure (8) [7]. $V_{j-1} = V_j \oplus W_j.$
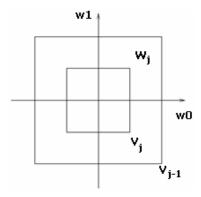

Figure (8). Multiscale subspaces generated by the Laplacian pyramid.

The LP can be considered as an oversampled filter bank where each polyphase component of the difference image $d[n]$ comes from a separate filter bank channel like
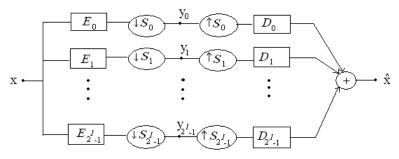
the coarse image $c[n]$. Let $Fi(z)$, $0 \leq i \leq 3$ be the synthesis filters for these polyphase components. Note that Fi are high pass filters. As in the wavelet filter bank, with each of these filters a continuous function $\psi^{(i)}(t)$ is associated where:

$$\psi^{(i)}(t) = 2 \sum_{n \in Z^2} f_i[n]\phi(2t - n).$$

where $f_i[n]$ is the impulse response of the high pass synthesis filter $F_i(z)$ [7, 10].

9- Multidirectional

By using multi rate identities, it is instructive to view an *l*-level tree-structured DFB equivalently as a 2^l parallel channel filter bank, as shown below in Figure (9), with equivalent analysis filters, synthesis filters and overall sampling matrices. In Figure (9), the equivalent directional analysis filters are denoted as $E_k^{(l)}$, $0 \leq k \leq 2^l$, and the directional synthesis filters as $D_k^{(l)}$, $0 \leq k < 2^l$, which correspond to the subbands indexed as in Figure (10).

Figure (9). The multichannel view of an *l*-level tree-structured DFB
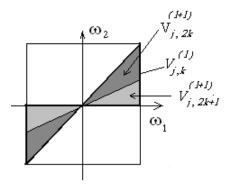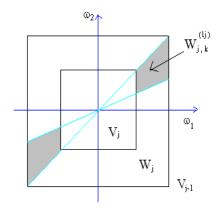
Figure (10).: Multiresolution Analysis: direction

From the multiscale analysis of the previous LP stage, the DFB in the contourlet filter bank utilizes orthogonal filters and when such DFB is applied to the difference image (detail) subspaces, the resulting detail directional subspaces $W_{j,k}^{(l)}$ in the frequency domain will be as illustrated in Figure (11). The indexes *j, k,* and *n* specify the scale, direction, and location, respectively. Note that the number of DFB decomposition levels *l* is different at different scales *j*, and is denoted by l_j. Recall that W_j is not a shift-invariant subspace. However, its subspaces $W_{j,k}^{(l)}$ are regenerative, since they are generated by a single function and its translations [7 and 10].

Figure (11). Multiscale and Multidirectional Subspaces Generated by the Transform
this is Illustrated on 2-D Spectrum Decomposition

10- Huffman Coding

Huffman encoding is an important part of many data compression algorithms; Huffman encoding is effective to compress numerical data [11], the advantage from applying Huffman for numerical data to convert it to binary. The following Huffman coding algorithm given L symbols for coding:

1. Arrange the symbol probability P**i** in descending order and consider them as leaf nodes of a tree.
2. While there is more than one node
   - ✓ Merge the two nodes with smallest probability to form a new node whose probability = sum of the two merged nodes.
   - ✓ Arbitrarily assign 1 or 0 to each pair of branches merging into a node (e.g. 0 to left branch, 1 to right branch).
3. Read sequentially from root node to leaf where symbol is located.

11- Run-Length Encoding

Run Length Encoding or RLE is a simple method from lossless compression [12 and 13]. The idea is that if item d occurs n consecutive times in the input stream, the n occurrences are replaced with the single pair d n. The n consecutive occurrences of a data item are called a run length of n, and this approach to data compression is called Run Length Encoding or RLE [14].

12- Performance Measures

A number of quantitative parameters can be used to evaluate the performance of the wavelet based speech coder, in both reconstructed signal quality after decoding and compression. The following parameters are compared

- Signal to Noise Ratio (SNR) the following formula

$$SNR = 10 log 10 \left[\frac{\sigma_x^2}{\sigma_z^2}\right]$$

  σ_x^2 is the mean square of the speech signal σ_z^2 is the mean square difference between the original and reconstructed signal.
- Normalized Root Mean Square Error

$$NRMSE = \sqrt{\frac{(x(n) - r(n))^2}{(x(n) - \mu_{x(n)})^2}}$$

x(n) is *the speech signal,* r(n) is the *reconstructed signal ,and* $\mu_{x(n)}$ *the mean of the speech signal* [1, 15] .

- Correlation between original signal and reconstructed signal shows how much the returned signal is close to the original one, and is represented in equation [16]:

$$corr = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2)(\sum_m \sum_n (B_{mn} - \bar{B})^2)}}$$

Where:

$\bar{A}$ = mean2 (A), and

$\bar{B}$ = mean2(B).

- **Compression ratio:** It is the ratio of the original signal to the compressed signal.

$$CR = \frac{\text{original signal size}}{\text{compressed signal size}}$$

13- The Proposed Algorithm

The CT applied on the image, but in this research is modified to be used on the speech compression, the speech is one dimension, so it is to be convert to 2-D before using CT. The results applied the EZC algorithm. The results feed to the Huffman coding, and then applied RLE. The coded result treated to reconstruct the expected speech signal by applying inverse process. Figure (12) represents the proposed algorithm.
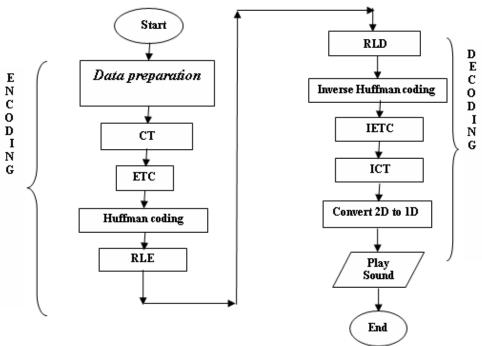

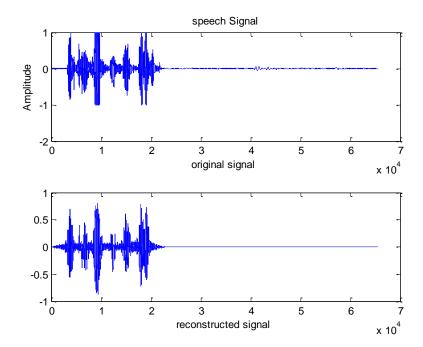**Figure (12).** The Flowchart the Proposed Algorithm

The SNR, NRMSE, Correlation (Corr.) and Compression Ratio (CR) on the speech are applied in this paper. Testing some speech signals and the results of measurement between original signal and reconstructed signal are shown in Table (1),

from the results on this table good results are shown specially the CR yielded high CR by S4 is equal to **(9.5855)**, in addition, Figure (13) shows high similarity between original signal and reconstructed signal.

Table (1). Performance of testing signals

	SNR	NRMSE	Corr.	CR
S1	3.6269	0.6325	0.7748	5.0346
S2	3.2061	0.6429	0.7672	5.8499
S3	5.8024	0.5132	0.8585	8.4880
S4	4.1031	0.6256	0.7801	9.5855

Figure(13). Represent Original Signal and Reconstruct Signal

14- Conclusion

From these are results conclusion can applied Contourlet Transform with signals and give good results, and yielded high Compression as applied with images. Then using the EZC on contourlet transform with Huffman coding and RLE give good results, and yielded high Compression.

REFERENCES

[1]   Mohammed Z., 2007, "Wavelet based Compression System Using Linear Prediction", Msc. Thesis, College of Engineering, University of Basrah.

[2]   Sudhakar R., Karthiga R., and Jayaraman S., 2006, "Fingerprint Compression Using Contourlet Transform with Modified SPIHT Algorithm", Iranian Journal of Electrical and computer Engineering, Vol.5, No. 1.

[3]   Selin A., 2010, "Two-Dimensional Wavelets ", Lecture: T, Th 8:30-9:50 a.m. 1225 Engineering Building, Electrical & Computer Engineering, ECE 802.

[4]   Zanartu M., 2005," Audio Compression using Wavelet Techniques", University Purdue, Electrical and computer engineering.

[5]   Dipeng C., and Qi L., 2005, "The Use of Complex Contourlet Transform Fusion Scheme", Proceedings of World Academy of on Science, Engineering and Technology, Vol. 7.

[6]   Bamberger R. H., and Smith M. J. T., 1992, "A filter bank for the directional decomposition of images: Theory and design," IEEE Trans. Signal Proc., vol. 40, no. 4, pp. 882–893.

[7]   Do M. N., 2001, "Directional Multiresolution Image Representations", PhD. Thesis, Swiss Federal Institute of Technology, Lausanne, Switzerland.

[8]   Ramin E., and Hayder R., 2005," New Image Transforms Using Hybrid Wavelets and Directional Filter Banks: Analysis and Design", Electrical and Computer Engineering Department, Michigan State University, IEEE.

[9]   Esakkirajan S., Veerakumar T., Senthil M. V., and Sudhakar R., 2006, "Image Compression Using Contourlet Transform and Multistage Vector Quantization", GVIP Journal, Vol. 6, Issue 1.

[10]   Abdul-Jabbar. J. M., and Fathee H. N., 2010, "Design and Realization of Circular Contourlet Transform", Dept. of Computer Eng., College of Eng., University of Mosul, Iraq ,Al rafidain Engineering journal , Vol. 18,No.4.

[11]   Pushpa R. Suri and Madhu Goel, 2010, "Ternary Tree and Clustering Based Huffman Coding Algorithm", IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 5.

[12]   Ahmed K. H. & Fadel S. H., 2007, "Hybrid Differential Pulse Code Modulation in Wavelet Domain and Lossless Methods for Image Compression", IJCCCE, Vol.7, No.1.

[13]   Balakrishnan S., 2008, "Compression of 2-Tone Manuscript for Multimedia Application", Msc. Thesis, Science.

[14]   Ganesh D. B., Raghuram J. K., and Vikram M. G., 2009, "Modified EZW Coding for the M-band Wavelet Transform and its Application to Image Compression", IEEE Xplore.

[15]   Rao N., 2001, "Speech Compression Using Wavelets", thesis of Engineering (Elec Hons), School of Information Technology and Electrical Engineering, University of Queensland.

[16]   Matlab 7.10, R2010, "Wavelet tool box".