AUDIO LOSSLESS CODING/DECODING METHOD USING BASIS PURSUIT ALGORITHM

Wenxin He, Tianshu Qu

Key Laboratory on Machine Perception (Ministry of Education), Speech and Hearing Research Center, Peking University, Beijing, China

ABSTRACT

Basis pursuit algorithm is one of the most popular methods of sparse coding. The goal of the algorithm is to represent signal using as few coefficients as possible, which is suitable for acoustic signal compression. This paper presents a lossless coding/decoding method using the basis pursuit algorithm. In this method, wavelet packets bases were used to compose the dictionary because of their natural sparse property. Experimental results are obtained by comparing the proposed method with the four popular lossless coding/decoding methods using various types of acoustic signals. The results show that the proposed method is competitive with the well-known methods for lossless compression, in terms of compression ratio and computational efficiency.

Index Terms— Lossless coding/decoding, Basis pursuit, Wavelet packets transform, Sparse coding, Lifting scheme

1. INTRODUCTION

The audio lossless coding/decoding technology is defined as an algorithm that can compress the digital audio data without any loss in quality due to a perfect reconstruction of the original signal. Recently, with the continuing development of technologies on massive storage devices, high-speed Internet connections, and broadband wireless networks, lossless coding/decoding technology are advancing. Till now, several lossless audio coding technologies have been developed, including MPEG4-ALS[1,2,3], Monkey[4], FLAC[5], and WavPack[6] et. al. The core technology of them is based on Linear Predictive Coding (LPC), which was originally proposed and developed as a technique for speech analyzing, and now is a most common and practical technique for lossless audio compression. The compression ratio depends on the source material, and is generally between 30% and 70%. Different from these methods, this paper proposed a lossless coding/decoding method based on a basis pursuit algorithm (BP).

Basis pursuit method is one of the most popular approaches of sparse coding. The goal of sparse coding is to represent a given signal as a linear superposition of a small number of stored bases (atoms), which come from an overcomplete set (dictionary)[9]. In traditional signal representation methods, such as the discrete cosine transform (DCT) or wavelet transforms, the dictionary is equal to the dimensionality of the signal space and the representation is unique. However, in sparse coding method, due to an overcomplete dictionary, the representation is no longer unique because the number of atoms is greater than the dimensionality of the signal space. This method enables flexibility in representation; as well it requires a criterion to select among the possible representations. The first one that comes to mind is sparsity, by which the selected representation is the one which uses the fewest atoms. Computing sparse representation is NP-hard, and several methods have been developed to solve it. Among them, matching pursuit (MP) [7, 8] and basis pursuit methods[9, 10, 11] are two effective methods. MP is a greedy algorithm: a signal representation is iteratively built up by selecting the atom that maximally improves the representation at each iteration step. MP is easy to implement, converges quickly and has good approximation properties. However, there is no guarantee that MP computes sparse representations. BP method, instead of seeking sparse representations directly, is a method that seeks representations that minimize the l_1 – norm of the coefficients. Recent theoretical studies[7, 8] show that representations computed by BP can be guaranteed to be sparse under certain conditions.

Wavelet packets bases have good properties on sparse representation. So, in this paper, we proposed a lossless coding/decoding method based on a basis pursuit algorithm using the dictionary composed with wavelet packets bases.

The following chapters provide more details of the proposed audio lossless coding/decoding algorithm. Section 2 introduces the structure overview of the coder and decoder; Section 3 introduces the detail of basis pursuit method used in this study; Section 4 provides the results of the comparison of the proposed method with other methods to assess the performances of the proposed method; At last the conclusion is given in section 5.

2. THE STRUCTURE OF THE LOSSLESS CODING/DECODING ALGORITHM

2.1. Structure of the Encoder

As shown in Fig. 1, the proposed encoder consists of four main blocks. Firstly, the input signal is divided into many frames, and the buffer is used to store the frames one by one; The sparse coding block is the key part of the proposed lossless coding scheme, it uses basis pursuit method to select the bases of wavelet packets, and then transforms the signal into sparse coefficients; In the block of entropy coding, the method proposed by Shu et. al in [12] is used to encode the sparse coefficients; and in the multiplexing block the output of entropy coding and side information (including wavelet packets base indices, and wavelet packet tree information) are combined to form the compressed bitstream.

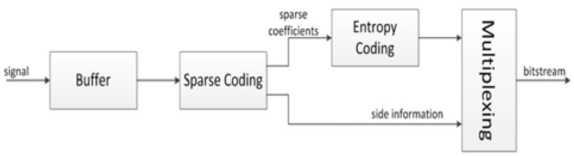

Fig. 1. The structure of the encoder

2.2 Structure of the Decoder

As is shown in Fig. 2, the structure of the decoder is the reverse progress of that of the coder, and consists of three blocks. The function of demulitiplexing block is to separate the side information (including wavelet packets base indices, and wavelet packet tree information) from the bitstream; In entropy decoding block, the sparse coefficients are generated from the remainder of the bitstream; In the sparse decoding block, the original signal is reconstructed using the sparse coefficients and the side information.

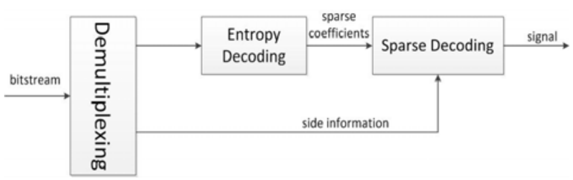

Fig. 2. The structure of the Decoder

3. DETAIL OF THE SPARSE CODING BLOCK

Sparse coding is a powerful trend in recent years. The key point of sparse coding is that the atoms in the dictionary which is used to represent signal are redundant, rather than just bases. In such a setting, we consider a linear equation,

$$s = D\alpha \tag{1}$$

where s is a given signal, D is the representation dictionary, and α is the signal's sparse representation. The matrix D is a general full rank $N\times L$ matrix, where $L > N$, and is assumed to have l_2 normalized columns. The number of non-zero elements in the coefficient vector α is measured by the l_0-norm, $\|.\|_0$, on R^L. The goal is to find, within the $(L-N)$-dimensional affine space of the solutions for this equation, the sparsest representation for s, which has the least number of non-zero entries. This can be formalized by the following optimization problem.

$$(\mathrm{P}_0)\ \ Arg \min_{\alpha\in R^L} \|\alpha\|_0 \ \text{ s.t. } D\alpha = s \tag{2}$$

The problem (P_0) is NP-hard, demanding an exhaustive search over all the subsets of columns of D. One of the most effective techniques to approximate its solution is the convex relaxation of the l_0-norm. It uses the l_1-norm, the closest convex norm on R^L:

$$(\mathrm{P}_1)\ \ Arg \min_{\alpha\in R^L} \|\alpha\|_1 \ \text{ s.t. } D\alpha = s \tag{3}$$

The idea of using (P_1) to find the sparsest solution is called basis pursuit (BP) [10,11].

As shown in Fig. 3, the detail structure of the sparse coding block is composed of three stages.

(1) Dictionary creating. In this stage, several types of wavelet basis, such as Cohen-Daubechies-Feauveau wavelets (cdf), Daubechies wavelets (db), and Symlets (sym) were merged together to create an overcomplete dictionary.
(2) Wavelet packets transforming. In this stage, the signal was represented by wavelet packets transform using all the bases in the dictionary, respectively. Because of the overcomplete property of the dictionary, the representation is not unique. The number of representations is decided by the volume of dictionary.
(3) Base Choosing. In this stage, basis pursuit method was used to choose one best presentation among all the representations. The coefficients of the best presentation have minimum l_1-norm.

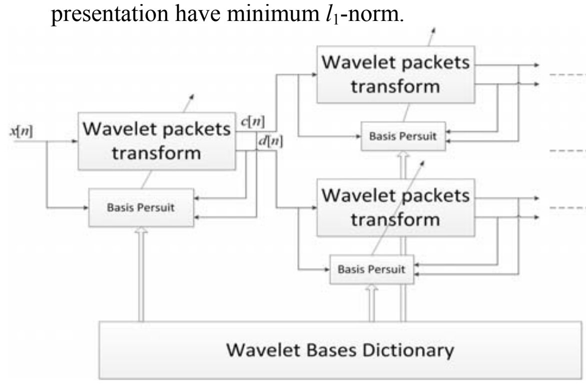

Fig. 3. The detail structure of sparse coding block

The wavelet packets transform block in Fig. 3 is realized by lifting scheme. It is named lifting wavelet packets transform. The lifting scheme was developed by Ingrid W. Sweldens [13]. It has many advantages over the classical implementation of the wavelet transform. For example, this method has a faster implementation, and it can yield reversible integer-to-integer wavelet transforms. Integer wavelet transforms when implemented via lifting scheme have better computational efficiency and lower memory requirements. The integer-to-integer property along with the reversible property makes the lifting scheme ideal for lossless compression.

The basic lifting scheme consists of three stages [14]: split, predict and update, as shown in Fig. 4. Firstly, the input signal $x[n]$ is divided into two disjoint subsets with even indexed samples $x_e[n]$ and odd indexed samples $x_o[n]$, respectively. This stage is also named lazy wavelet transform. If the signal has a local correlation, then these two sets are highly correlated. In the predict stage, a predictor filter uses the even indexed samples $x_e[n]$ to predict the odd indexed ones $x_o[n]$. The error of prediction represents the high-pass coefficients. In the update stage, the error of prediction is used to update the current phase of the even indexed samples, producing the low-pass coefficients. This stage is accomplished by applying an update filter to the wavelet coefficients and adding the result to the even indexed samples. Iteration of the predict stage and update stage creates the scaling coefficients $c[n]$ and wavelet coefficients $d[n]$. The predict filter and update filter can be obtained using the method introduced in[13].

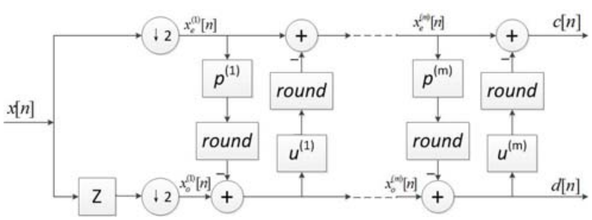

Fig. 4. The structure of analysis lifting wavelet transforms

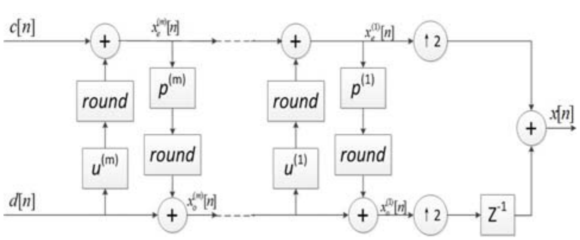

Fig. 5. The structure of reconstruction lifting wavelet transforms

At the reconstruction stage, prediction and update are done in reverse order, the two disjoint sets $c[n]$ and $d[n]$ merge into one to achieve the perfect reconstructed signal $x[n]$. The procedure is shown in Fig. 5.

For lossless coding, the output of the predictor and update filters should be rounded before the addition to the corresponding components. The integer method is described as follows [15], and shown in Fig.4.

$$\begin{aligned} x_o^{(i+1)} &= x_o^{(i)} - \left\lfloor \sum_k p_k^{(i)} x_e^{(i)} + 0.5 \right\rfloor \\ x_e^{(i+1)} &= x_e^{(i)} - \left\lfloor \sum_k u_k^{(i)} x_o^{(i+1)} + 0.5 \right\rfloor \end{aligned} \tag{4}$$

where $\lfloor x \rfloor$ is the largest integer not exceeding x.

The inverse version in the decoder can be obtained by flipping the signs and reversing the operations in (4). The process is described as follows[15], and shown in Fig.5.

$$\begin{aligned} x_e^{(i)} &= x_e^{(i+1)} + \left\lfloor \sum_k u_k^{(i)} x_o^{(i+1)} + 0.5 \right\rfloor \\ x_o^{(i)} &= x_o^{(i+1)} + \left\lfloor \sum_k p_k^{(i)} x_e^{(i)} + 0.5 \right\rfloor \end{aligned} \tag{5}$$

4. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the presented scheme by coding various types of acoustic signals in a lossless manner. Four popular lossless coding/decoding methods are chosen for comparison. They are Flac (flac -8) [5], WavePack (v. 4.60.1)[6], Monkeys's Audio (mac –c4000) [4], and MPEG4 ALS (with BGMC) [3]. Thirteen acoustic signals are used as the test signals. And they are varied in style including wind music, flute, symphony, Jazz, piano, rock, string music, electro music, pop music, country music, and speech. The evaluation is obtained in terms of compression rate and computational efficiency.

4.1. Compression rate

In order to evaluate the compression rate, three sets of audio content at different sampling rates (32kHz/16bits, 44.1kHz/16bits, and 48kHz/16bits) are used to evaluate the performance of lossless audio coding schemes. The compression ratio is defined as follows:

$$CR = \frac{Compressed\ file\ size}{Original\ file\ size} \times 100\% \tag{6}$$

The calculation results are given in Table1.

Table 1. Average compression ratios (%) for different audio formats

Format	Proposed	Flac	Wavpac	Monkey's	ALS
32kHz/ 16bit	52.19	52.95	52.37	51.25	51.87
44.1kHz/ 16bit	48.02	48.86	48.67	47.21	47.77
48kHz/ 16bit	46.38	47.80	47.93	45.38	46.40
Average	48.86	49.87	49.66	47.94	48.68

4.2. Computational efficiency

The decompression time is an important issue in real-time applications. The test of it is conducted at the same time when fulfilling the Table1 on a 2.1GHz AMD x2 dual core QL-64 laptop. The results are shown in Table. 2.

Table 2: Decompression time (seconds) measured for different audio formats

Format	Proposed	Flac	Wavpac	Monkey's	ALS
32kHz/ 16bit	0.96	0.30	0.41	0.81	0.41
44.1kHz/ 16bit	1.34	0.30	0.51	1.00	0.50
48kHz/ 16bit	1.64	0.34	0.55	1.05	0.54
Average	1.31	0.31	0.49	0.95	0.48

5. CONCLUSION

Linear Predictive Coding (LPC) is a common and practical technique for lossless audio compression. Based on this method, the compression ratio depends on the source material, but generally is between 30% and 70%. In this paper, instead of using LPC method, we present a lossless coding/decoding method using basis pursuit algorithm. Basis pursuit algorithm is an effective method to represent signal using few coefficients, which is suitable for acoustic signal compression. For evaluating the performance of the presented scheme, four popular lossless coding/decoding methods are chosen for comparison. The results show that the proposed method is competitive with the well-known methods for lossless compression, in terms of compression ratio and computational efficiency.

6. RELATION TO PRIOR WORK

The work presented here focuses on the lossless coding/decoding technology. To our knowledge, MPEG4-ALS, Monkey, FLAC, and WavPack are popular used among the lossless technologies. The core technology of them is based on Linear Predictive Coding [3, 4, 5, 6]. The compression ratio generally is between 30% and 70%. This ratio is not satisfactory under some situations. In this paper, instead of using LPC method, we present a lossless coding/decoding method using basis pursuit algorithm. The evaluation results show that the proposed method is effective compare to the four popular lossless technologies. In this study, wavelet bases are chosen to construct the overcomplete dictionary. The bases are good, but not the best for acoustic signal. So, the future work of this study will focus on how to construct a class of bases for acoustic signal.

7. ACKNOWLEDGEMENTS

The work was supported by the National Natural Science Foundation of China (No.61175043, No.91120001, and No.61121002); "Twelfth Five-Year" National Science & Tech. Support Program of China (No.2012BAI12B01); National Health Public Welfare Industry Research Project of China (No. 201202001)

8. REFERENCES

[1] T. Liebchen, "an Introduction to MPEG-4 Audio Lossless Coding," *International Conference on Acoustics, Speech, and Signal Processing* (*ICASSP '04*), vol. 3, pp.1012-1015, 2004

[2] ISO/IEC JTC1/SC29/WG11 N5040, Call for Proposals on MPEG-4 Lossless Audio Coding (Klagenfurt, AT, July 2002)

[3] MPEG4 ALS Source Code, *ftp://ftlabsrv.nue.tu-berlin.de/mp4lossless/*.

[4] Monkey's Audio Compression Program, *http://www.monkeysaudio.com/*.

[5] FLAC Open-source Audio Compression Program, *http://flac.sourceforge.net/*.

[6] Wavpac Open-source Audio Compression Mode, *http://wavpack.com/*.

[7] S. G. Mallat, Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Trans. on Signal Process*, vol. 41(12), pp. 3397–3415, 1993.

[8] N. Cho, C.-C. J. Kuo, "Sparse Music Representation with Source-specific Dictionaries and its application to signal Separation," *IEEE Trans. On Audio, Speech, and Language Processing*, vol. 19(2), pp. 337-348, 2011

[9] P. S. Huggins, S. W. Zucker, "Greedy Basis Pursuit," *IEEE Trans. on Signal Processing*, vol. 55(7), pp. 3760-3772, 2007

[10] S. S. Chen, *Basis Pursuit*, Ph.D. Thesis, Stanford University, 1995

[11] S. S. Chen, D. L. Donoho, M. A. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM Review*, vol. 43(1), pp. 129 -159, 2001

[12] H.Y. Shu, H.B. Huang, T.E. Chan, R.S. Yu, and S. Rahardja, "An Introduction to AVS Lossless Audio Coding," *the 129th Convention of Audio Engineering Society*, Sam Francisco, CA, USA., Convention Paper 8285, 2010

[13] W. Sweldens, "The lifting scheme: A Construction of second generation wavelet," *SIAM Journal on Mathematical Analysis*, vol. 29(2), pp. 511–546, 1998

[14] I. Daubechies, W. Sweldens, "Factoring Wavelet Transforms into Lifting Steps," *The Journal of Fourier Analysis and Applications*, vol. 4(3), pp. 247-269, 1998

[15] A. R. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, Wavelet transforms that map integers to integers," *Applied and computational harmonic analysis*, vol. 5(3), pp.332-369, 1998.