Wavelet-based compression of multichannel climate data

Ershad Sharifahmadian[1]*, Yoonsuk Choi[1], Shahram Latifi[1], Sergiu Dascalu[2], Frederick C. Harris[2]
[1]Dept. of Electrical and Computer Engineering, University of Nevada, Las Vegas
[2]Dept. of Computer Science and Engineering, University of Nevada, Reno

ABSTRACT

To simultaneously compress multichannel climate data, the Wavelet Subbands Arranging Technique (WSAT) is studied. The proposed technique is based on the wavelet transform, and has been designed to improve the transmission of voluminous climate data. The WSAT method significantly reduces the number of transmitted or stored bits in a bit stream, and preserves required quality. In the proposed technique, the arranged wavelet subbands of input channels provide more efficient compression for multichannel climate data due to building appropriate parent-offspring relations among wavelet coefficients. To test and evaluate the proposed technique, data from the Nevada climate change database is utilized. Based on results, the proposed technique can be an appropriate choice for the compression of multichannel climate data with significantly high compression ratio at low error.

Keywords**:** Climate data, Multichannel data compression, Wavelet subbands arranging technique, Wavelet transform

1. INTRODUCTION

Compression helps decrease data storage space or transmission capacity. The design and development of data compression methods involves trade-offs among different factors such as the amount of distortion introduced, the compression degree, and the computational resources required to compress and uncompress the data [1].

With enough information about data's features such as the size, type, and statistical features of data, it is possible to define or select an appropriate scheme for the compression of data [2]-[6].

The scientific climate data needs large bandwidth for transmission and there is a high traffic from/to climate data servers. Those climate data should be stored or transmitted in reasonable amounts of time. It is not always possible to increase bandwidth to support all users, and prepare good service for on-line clients. Therefore, efficient compression of climate data is emphasized [2].

To achieve better decompressed data quality without affecting the main features of the reconstructed data and high compression ratio, various compression schemes have been proposed. To compress climate data, there are different types of compression schemes have been reported in the literature. In [5], the effects of several compression schemes including GRIB2 encoding, GRIB2 using JPEG 2000 and LZMA, and the commercial Samplify APAX method were tested on climate datasets, and compression ratio, data quality, and processing time have been evaluated.

Steffen and his colleague at NOAA research-forecast systems laboratory, Boulder, Colorado proposed the application of compression methods such as GZIP and BZIP2 to weather data. They showed the benefits of applying a difference filter to extract the smoothness quality, and how it improves the compression ratio when GZIP and BZIP2 are used [6].

In order to compress three-dimensional meteorological data, several lossy compression methods have been reported in literature [7], [8]. Computer modeling programs such as the battle scale forecast model are able to produce three-dimensional meteorological data which is used to study the efficiency of data compression techniques for storage or transmission. In [7], the benefits of applying lossy data compression techniques to meteorological data have been presented. Due to the state of development of digital image compression methods such as the JPEG 2000, the scheme proposed in [7] used the two-dimensional, single-component JPEG 2000 method on horizontal 2-D slices of data.

Another 3-D compression method presented in [8] is based on different bit allocation techniques used to compress Karhunen-Loeve transformed data.

In general, to provide a better performing system for transmitted data, and to reach a high compression ratio in multichannel data compression, it is required to decrease redundancy between adjacent data of each channel of multichannel data acquisition system. In this paper, the technique called WSAT is utilized to simultaneously compress multichannel climate data. In the following section, the wavelet subbands arranging technique is briefly explained. In the section 3, the method is tested using data from the Nevada climate change database and results are presented. The Section 4 concludes the paper.

*sharifah@unlv.nevada.edu

2. METHOD

In the first step, the wavelet transform is utilized. During the phase of wavelet decomposition, a low pass filter and a high pass filter are applied on input data. After down-sampling of the two filtered outputs, an approximation subband and a detail subband are obtained. If the N-level wavelet decomposition is performed, $N+1$ wavelet subbands are obtained. The length of the approximation subband is named as L_a which equals to the length of the detail subband.

Now, the threshold of each wavelet subband is expressed as:

$$Th_N = 2^{\left\lfloor \log_2 \left(Max_{i \in Nth\ subband} \left|wav_i\right|\right) \right\rfloor} \tag{1}$$

where Th_N is the threshold of the N^{th} wavelet subband, wav_i is the wavelet coefficient at location i [9]. Now, the largest calculated threshold is selected as the initial operation threshold.

Wavelet coefficients with larger magnitude are placed in lower frequency subbands [10]. Arranged wavelet subbands of input channels provide more efficient multichannel compression. In fact, appropriate parent-offspring relations among coefficients of subbands are built at a hierarchical structure, and it leads to more efficient multichannel compression.

For simultaneous multichannel signal compression, the wavelet subbands of input channels are arranged appropriately. In order to rearrange the subbands, the largest threshold from every channel is determined based on formula (1). Among available input channels, a channel with the largest threshold is selected as initial channel. The second selected input channel has threshold larger than residual channels, however, its threshold is smaller than that of the initial channel. Rest of input channels will be chosen by this procedure. Thus, a new structure is built in which the approximation and detail subbands of chosen input channels are sorted from highest subband threshold to lowest subband threshold [2].

Eventually, the coefficients are rearranged in hierarchies, with roots placed in the lowest frequency subband, and branching into higher frequency subbands consecutively.

In the hierarchical structure, *Des*(*i*) is called as the set of all descendants of a node defined at location *i*, which is named as a set of type I. *Ex*(*i*) is called as the set of all descendants excluding offspring of a node defined at location *i*, which is named as a set of type II.

When the initial operation threshold was selected, the given threshold is compared with the threshold of each wavelet subband. This procedure is performed from the highest frequency subband to the lowest frequency subband, and it will be finished if the threshold of a wavelet subband becomes equal to the initial operation threshold for first time. Then, the location of the last wavelet coefficient from the subband where comparison was performed up to, is saved. The given location, Frontier, limits the evaluation by which the significance of *Ex*(*i*) or *Des*(*i*) is decided [9].

During the wavelet subbands arranging technique, the main information about the wavelet coefficients is placed in different classes:

Class1: The class of locations relating to coefficients that have magnitude smaller than a given threshold.

Class2: The class of locations relating to coefficients that have magnitude larger than/equals to a given threshold.

Class3: A two-column matrix including the class of locations corresponding to the sets of the coefficients defined by hierarchical structure; the given sets have magnitude smaller than a given threshold.

The coefficients are compressed during several scans. For every scan, the wavelet coefficients with magnitudes which are larger than a certain threshold are compressed. In the beginning of scans, the Class2 is set as empty. If the length of the approximation is even, the locations of {0 to L_a-1} will be placed into the Class1, and the locations of {$L_a/2$ to L_a-1}, will be placed into the Class3 as type I. If the approximation subband is odd, the locations of {0 to L_a -2} will be placed into the Class1, and the locations of {$(L_a - 1)/2$ to L_a -1}, will be placed into the Class3 as type I [2].

A coefficient which is relevant to the Class1 is significant if the given coefficient is larger than/equal to the current operation threshold. Therefore, a one is sent with a sign bit, and its position is transferred to the Class2. If the wavelet coefficient is not significant, a zero is sent and its location kept until the next compression level. For the negative sign, the sign bit is set as 0, otherwise the sign bit is set as 1. After evaluating the Class1, coefficients which is relevant to the Class3 are checked. For a *Des*(*i*), the following procedure is performed (the location of last wavelet coefficient in the highest frequency subband is called as *Last*) [2]:

```
if $Des(i)$ is significant
    Send $1$.
    Check two offspring of $Des(i)$ like the evaluation of wavelet coefficients relating to the Class1.
    if location $4i+3 \leq Last$
        Move $Des(i)$ to second column of Class3 and change it to $Ex(i)$.
    elseif location $4i > Last$
        Remove $Des(i)$ from first column of Class3.
    elseif location $4i < Last$
        Check wavelet coefficients at locations $4i$, $4i+1$, and $4i+2$.
        Remove $Des(i)$ from first column of Class3.
    else
        Check wavelet coefficient at location $4i$.
        Remove $Des(i)$ from first column of Class3.
    end
else
    Send $0$
    Preserve location and type of $Des(i)$ until next level of compression.
    Remove $Des(i)$ from first column of Class3.
end
```

After the evaluation of all sets in the first column of the Class3, *Ex*(*i*) in the second column of the Class3 are processed. Before each *Ex*(*i*) is checked, the two offspring, direct descendants of a wavelet coefficient at location *i*, are reconsidered. These direct children of a tree node at position *i* are used to improve the evaluation of *Ex*(*i*)s. If these direct children are smaller than the current operation threshold, those are selected as new entries of type I, and directly moved to the first column of the Class3. Thus, evaluation of *Ex*(*i*) is not done for those children. If at least one of the direct children is larger than/equal to the current operation threshold, the evaluation of *Ex*(*i*) is done. Therefore, if *Ex*(*i*) is not significant, a zero is sent. Afterwards, the *Ex*(*i*) is removed from the second column of the Class3, its type and position are saved till the next compression level. If the set of type II is significant, a one is sent. Then, its two offspring are chosen as new sets of type I and moved to the first column of the Class3. After that, the *Ex*(*i*) is removed from the second column of the Class3. When all *Ex*(*i*)s in the second column of the Class3 were processed, new *Des*(*i*)s in the first column of the Class3 are checked. Evaluating the Class3 is finished when no new *Des*(*i*) is moved to the first column of the Class3. After the process of the Class3, the process of Class2 is begun, and each old entry of the Class2 is evaluated. When an old entry of the Class2 is significant over the current operation threshold, a zero is sent, otherwise a one is sent. After processing all entries in the Class2, the magnitude of all wavelet coefficients for which their locations placed in the Class2 and are larger than the current operation threshold, are subtracted by the current operation threshold [2]. When the Class2 was processed, the current operation threshold is halved. In next, the Class1, Class3, and Class2 are processed again until the desirable bit rate is provided.

3. RESULTS

To demonstrate the efficiency of the proposed technique, the data from the Nevada climate change database [11] is used. The method is tested on two datasets.

In order to evaluate the redundancy removing capability of compression techniques, the compression ratio, CR, is generally utilized. The percent root mean square difference, PRD, is a popular criterion to evaluate the fidelity of reconstructed data. The evaluation formulas are presented as follows [12]:

$$PRD = \sqrt{\frac{\sum_{i=1}^{N}(m_i - \hat{m}_i)^2}{\sum_{i=1}^{N} m_i^2}} \times 100 \qquad (2)$$

$$CR = \frac{Sr \times N}{B_{compress}} \qquad (3)$$

where m_i is the i^{th} sample from input data, $\hat{m}_i$ is its reconstructed sample, Sr is the sample resolution (bits/sample), N is the number of samples in the input signal, and $B_{compress}$ is the total number of bits to be transmitted or stored after compression. The average percent root mean square difference, APRD, is used to assess the reconstructed data in multichannel compression [2], [9], [13].

3.1. First Dataset

The 3-month air temperature data (2-meter height) from the research sites of Sheep Range Blackbrush, Sheep Range Montane, Snake Range Pinyon-Juniper (West), Snake Range Subalpine (West), Snake Range Subalpine (East) are considered. The temperature data consists of minimum, maximum and average values. The sample resolution is set as 12 bits/sample, and samples are gathered every 10 mins. The Daubechies 4 wavelets are selected [14], and a 5-level wavelet decomposition is applied. The average PRD results at different CR's per 3 channels are shown in Figure 1.

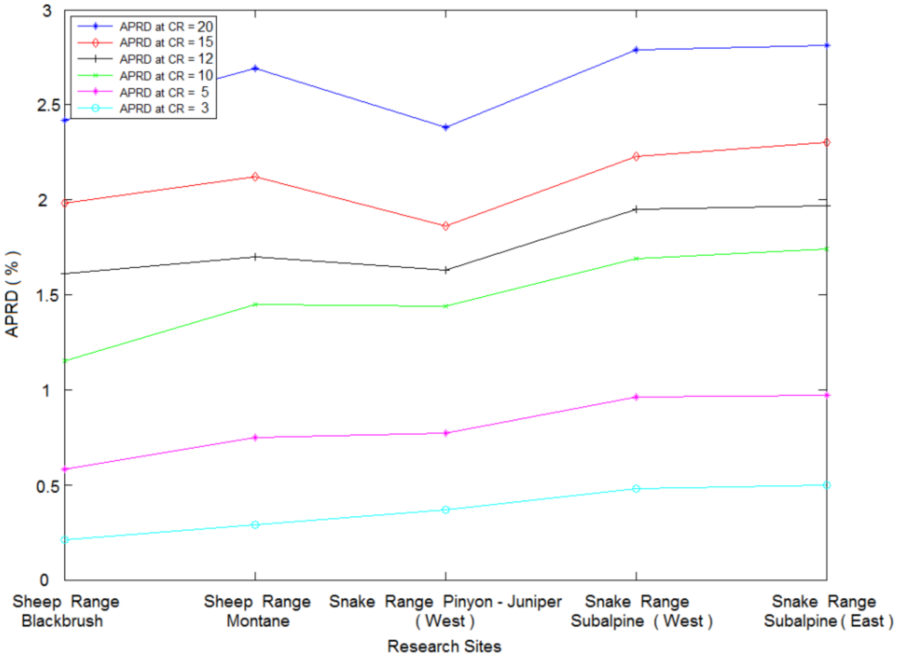

Figure 1. APRD values of air temperature from selected research sites corresponding to certain CR's

3.2. Second Dataset

The 3-month relative humidity (2-meter height), air temperature (2-meter height) and wind velocity (10-meter height) data from the research sites including Sheep Range Blackbrush, Sheep Range Montane, Snake Range Pinyon-Juniper (West), Snake Range Subalpine (West), Snake Range Subalpine (East) are considered. The sample resolution is set as 12 bits/sample, and samples are collected every 10 mins. The average PRD results at different CR's per 3 channels (i.e. relative humidity, air temperature, and wind velocity) are shown in Figure 2.

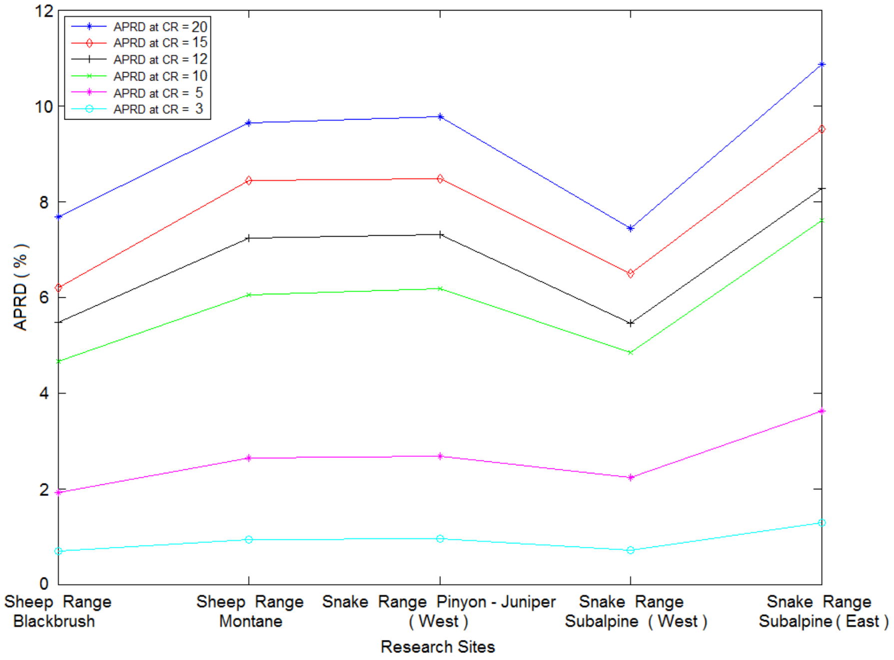

Figure 2. APRD values of relative humidity, air temperature, and wind velocity from selected research sites corresponding to certain CR's

4. CONCLUSION

In this paper, multichannel compression of climate data based on the wavelet subbands arranging technique was proposed. Among lossy compression methods for climate data presented in the literature, the proposed technique provides good performance due to the appropriate sorting of wavelet subbands of input data, and the efficient assessment of the coefficients. In fact, appropriate sorting of the subbands builds convenient parent-offspring relations among coefficients of subbands at a hierarchical structure.

Using the proposed technique, a user has exact control on bit rate, and the compression process can be stopped as soon as the desired bit rate or quality requirement is reached. Moreover, with the progressive coding ability of the WSAT, the quality of input data can be gradually enhanced as the bit rate increases.

In addition, the WSAT decreases the number of stored or transmitted bits. Whereas the WSAT gradually enhances the quality of reconstructed data, if redundant message bits are reduced, much important information will be transmitted at a certain bit rate. To evaluate the proposed method, data from the Nevada climate change database was used. It is concluded that the proposed technique can be an appropriate option for multichannel climate data compression, and provides significantly high compression ratio at low error.

ACKNOWLEDGEMENT

This work is supported (in part) by the Defense Threat Reduction Agency, Basic Research Award # HDTRA1-12-1-0033, and the National Science Foundation (NSF) award #EPS-IIA-1301726. Any findings, conclusions, or recommendations expressed in the material are those of the author(s) and do not necessarily reflect the views of NSF.

REFERENCES

[1] Tank, M. K., “Implementation of Limpel-Ziv algorithm for lossless compression using VHDL”, Think quest 2010: Proceedings of the First International Conference on Contours of Computing Technology, Springer, 275–283, (2011).

[2] Sharifahmadian, E., Choi, Y., and Latifi, S., “Multichannel data compression using wavelet subbands arranging technique,” International Journal of Computer Application, 91(4), 17-22, (2014).

[3] Sharifahmadian, E., “Compression of vital biosignals of patient by enhanced set partitioning in hierarchical trees algorithm,” World Congress on Medical Physics and Biomedical Engineering (WC2009) & IFMBE, 86-87, (2009).

[4] Sharifahmadian, E., and Pooyan, M., “Wavelet-based compression of patient’s medical image and ECG signals using the enhanced set partitioning in hierarchical trees algorithm,” BMES Annual Fall Meeting, (2008).

[5] Hubbe, N., Wegener, A., Kunkel, J. M., Ling, Y., Ludwig, T., “Evaluating lossy compression on climate data”, Supercomputing, Lecture Notes in Computer Science, Springer, Vol. 7905, 343-356, (2013).

[6] Steffen, C. E., and Wang, N., “Weather data compression,” 19th International Conference on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, Amer. Meteor. Soc., CD-ROM, 4.9, (2003).

[7] Cabrera, S. D., “Three-dimensional compression of mesoscale meteorological data based on JPEG2000”, Proc. of Battle Space Digitization and Network-Centric Warfare II, vol. 4741, 239-250, (2002).

[8] Lucero, A., Cabrera, S., Aguirre, A., Argaez, M., Vidal, E. J., “Comparison of bit allocation methods for compressing three-dimensional meteorological data after applying KLT”, IEEE International Symposium on Geoscience and Remote Sensing, Vol.4, 2514-2517, (2004).

[9] Sharifahmadian, E. “Wavelet compression of multichannel ECG data by enhanced set partitioning in hierarchical trees algorithm,” 28th IEEE EMBS Annual International Conference, 5238–5243, (2006).

[10] Mallat S. G., “A theory for multiresolution signal decomposition: The wavelet representation,” IEEE Trans. Pattern analysis and machine intelligence, vol. 11, 674-693, (1989).

[11] http://sensor.nevada.edu/NCCP/Default.aspx

[12] Miaou, S. G., and Yen, H. L., “Quality driven gold washing adaptive vector quantization and its application to ECG data compression”, IEEE Transaction Biomedical Engineering, vol.47, 209-218 (2000).

[13] Jalaleddine, S. M. S., Hutchens, C., et al., “ECG data compression techniques-A unified approach”, IEEE Transaction Biomedical Engineering, vol.BME-37, 329-343 (1990).

[14] Abdul Karim, S. A., Abdul Karim, B., et al., “Compression of temperature data by using Daubechies wavelets”, Proc. 2nd International Conference on Mathematical Sciences, 726-734 (2010).