HARMONIC WAVELET TRANSFORM FOR MINIMIZING RELATIVE ERRORS IN SENSOR DATA APPROXIMATION

Seonggoo Kang[1], Seunghoon Yang[1], Sukho Lee[1], Sanghyun Park[2]

[1] School of Electrical Engineering and Computer Science
Seoul National University, Seoul 151-742, Korea
exodus@db.snu.ac.kr, ysh0089@db.snu.ac.kr, shlee@cse.snu.ac.kr
[2] Department of Computer Science,
Yonsei University, Seoul 120-749, Korea
sanghyun@cs.younsei.ac.kr

ABSTRACT:

As the Ubiquitous generation approaches, the importance of the sensor data processing is growing. The data approximation scheme, one of the data processing methods, can be the key of sensor data processing, for it is related not only to the lifetime of sensors but also to the size of the storage. In this paper, we propose the Harmonic Wavelet transform which can minimize the relative error for given sensor data. Harmonic Wavelets use the harmonic mean as a representative which is the minimum point of the maximum relative error between two data values. In addition, Harmonic Wavelets retain the relative errors as wavelet coefficients so we can select proper wavelet coefficients that reduce the relative error more easily. We also adapt the greedy algorithm for local optimization to reduce the time complexity. Experimental results show the performance and the scalability of Harmonic Wavelets for sensor data.

1. INTRODUCTION

1.1 Motivation

As the Ubiquitous generation approaches, the uses of sensors are increasing. It means how to manipulate a large number of sensor data can be an important issue. One of these researches on this issue, the data approximation may suggest the effective way of sensor data processing. It is because by reducing the size of the transmission packet data, the lifetime of sensors can be extended and the efficiency of the storage also can be improved.

The definition of the approximation is to get the representation that is close enough to be useful for given memory bound. And historically, sampling, histogram and wavelets have been studied as approximation schemes. Among these schemes, are wavelets well-known for the effectiveness of their decomposition in reducing large amounts of data to compact data synopsis. And recent work has demonstrated the applicability and scalability of wavelets for sensor data.

Wavelets use the greedy mechanism by approximating the several data values to their representative value. Given that the original data is the highest resolution data, we can get the lower resolution data by replacing the original data values with their representative value. So there can be an error in approximated data, and moreover, which representative value we select instead of original data values determines the performance of the approximation scheme.

In this paper, we propose Harmonic wavelet transform which can minimize the relative error for given sensor data. Harmonic wavelets replace two data values with their harmonic mean at each decomposition step and retain the relative error as a wavelet coefficient. As we will explain later, the harmonic mean is the minimum point of the maximum relative error between two data values. So Harmonic wavelets can minimize the relative errors generated at each decomposition step and we can select proper wavelet coefficients that reduce the relative error more easily.

1.2 Our Contributions

Previous works are based on Haar wavelet transform which focuses on the absolute error minimization problem. However, previous works have attempted to minimize the relative error with the Haar wavelet transform. So, it is natural that a new wavelet transform focusing on the relative error is needed. Harmonic wavelets can be the solution for this problem.

Harmonic wavelets improved the performance and the time complexity. By using the harmonic mean instead of the arithmetic mean, we could reduce the maximum relative error about 10~30% compared to the previous works. And by adapting greedy algorithm for local optimization, we could reduce the time complexity down to $O(NB \log B)$ where N is the length of data and B is the given memory bound.

The remainder of this paper is organized as follows. Section 2 provides related works on approximation schemes. And then, we will introduce the main body of this paper, Harmonic wavelets, in Section 3. Section 4 shows the performance of Harmonic wavelets with experimental results. Finally, several concluding remarks will be given in Section 5.

2. RELATED WORKS

Various methods have been proposed to approximate large amounts of data. However, the majority of the proposed methods can be classified into sampling, histograms and wavelets [1].

Sampling-based techniques use random samples as synopsis for large data sets, but have the limitation in the combination of uniform random samples. Histogram-based techniques use statistical information and provide higher-quality approximations compared to sampling. But histogram-based approaches are faced with difficulties when dealing with the high-dimensional data sets. Wavelet-based techniques transform original data into a small number of compact data synopsis and recent studies have demonstrated the applicability of wavelets to the approximation of sensor data avoiding the high construction costs and storage overheads [1, 2, 3].

Approximation schemes based on wavelets are extended to minimize the L^p error. Several studies have shown that Haar wavelet is useful for L^2 error metric [4]. However, recently, it is pointed out that Haar wavelet synopsis with the L^2 error minimization cannot guarantee the error bound of individual data value. To solve this problem, a recent research suggested the probabilistic wavelet coefficient selection, and after that, the deterministic selection [2, 3].

However, the time complexity of this algorithm is not low enough to manipulate the large amount of data. Another limitation of this algorithm is in the nature of Haar wavelets. As Haar wavelets focus on the minimization of the absolute error, the performance of this algorithm on the side of relative error is not good as we expect.

3. HARMONIC WAVELET TRANSFORM

In this section, we propose the Harmonic wavelet transform for the relative error minimization. In section 3.1, we will deal with the overview of the Haar wavelet transform and its limitations. Section 3.2 provides the decomposition and reconstruction of the Harmonic wavelet transform. And then we introduce Greedy algorithm in section 3.3.

3.1 Overview of the Haar Wavelet Transform

The basics of the Haar wavelet decomposition are in the use of the pairwise decomposition. Given data $\vec{x}$ = (x_1, x_2) whose length is 2, the Haar wavelet transform uses the arithmetic mean that is defined by $(x_1 + x_2)/2$ and the difference between x_1 and the arithmetic mean for the data decomposition that is also defined by $(x_1 - x_2)/2$. Figure 1 shows the Haar wavelet transform for the data of length 2.

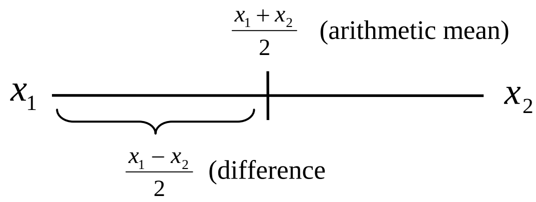

Figure 1. Basics of the Haar wavelet transform

Using this Haar wavelet basics, we can get the Haar wavelet transform for the data A = [1 3 10 6] as shown in table 1. The Haar wavelet decomposition can be regarded as a series of multi-level operations for the arithmetic mean and the difference.

Table 1. Example of Haar wavelet decomposition

resolution	arithmetic means	detail coefficients
2	[1 3 10 6]	-
1	[2 8]	[-1 2]
0	[5]	[-3]

Given that the resolution 2 is the original data, we can get the low resolution data by replacing pairwise data values with their arithmetic mean in each resolution step. For example, data [2 8] in resolution 1 are obtained by taking the arithmetic means of [1 3] and [10 6] in resolution 2. Also, detail coefficients [-1 2] are the differences made between the first value of each pair in resolution 2, which is 1 and 10, and the arithmetic means of each pair, which is [2 8]. This procedure is continued until the overall arithmetic mean is obtained. And then, Haar wavelets retain the overall arithmetic value followed by the detail coefficients in the order of increasing resolution. At this example, the Haar wavelet transform W_A of the data A can be [5 -3 -1 2].

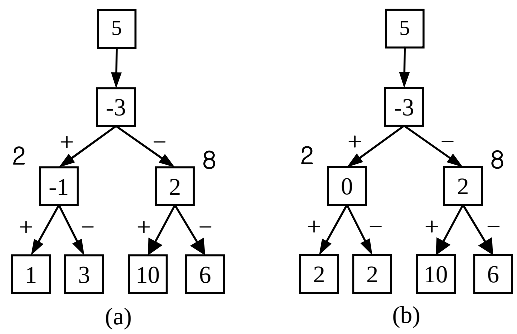

Figure 2. Reconstruction of data A with an error tree

An error tree is helpful to reconstruct the original data A from the Haar wavelet transform W_A as shown in Figure 2-(a). The root node of an error tree is the overall arithmetic mean and internal nodes are detailed coefficients in the Haar wavelet transform W_A. Given an error Tree T, there exists a path from root node to a leaf node, which is the original data value. If the subpath is to go to the left from the node u, then add the value of the node u. On the contrary, if the subpath is to go to the right, then subtract the value of the node. Figure 2-(b) shows the reconstruction of data A without the coefficient -1.

Haar wavelets, however, suffer from limitations on the side of the relative error minimization. For example, the detailed coefficient -1 yields absolute errors from original data values as shown in Figure 2-(b). It means Haar wavelet coefficients are related to the absolute error. Another limitation of the Haar wavelet transform is in the selection of the representative. As we will explain later, the arithmetic mean is the minimum point of the maximum absolute error at each data value but not the maximum relative error. It means the approximation using Haar wavelets will be problematic on the side of relative error metrics.

3.2 Harmonic Wavelet Transform

As a solution to these limitations on the side of relative error, we propose a new wavelet transform.

3.2.1 Problem Definition: In this paper, we will minimize the maximum relative error from the subset of the wavelet transform for given memory bound as shown in Figure 3.

(Input) For given sensor data A whose length is N ,
(Output) get the subset W_A' from the wavelet transform W_A ,
where the maximum relative error is minimized.

$$\text{Maximum Relative Error} = \max_i \left\{ \frac{| d_i' - d_i |}{| d_i |} \right\}$$

(d_i is the i^{th} value of the original data and d_i' is the i^{th} value of the approximated data)

Figure 3. Problem Definition

3.2.2 Harmonic Mean: Harmonic wavelets minimize the relative error by replacing two data values with their harmonic mean. We will explain how the harmonic mean can be the minimum point of the maximum relative error between two data values in this subsection.

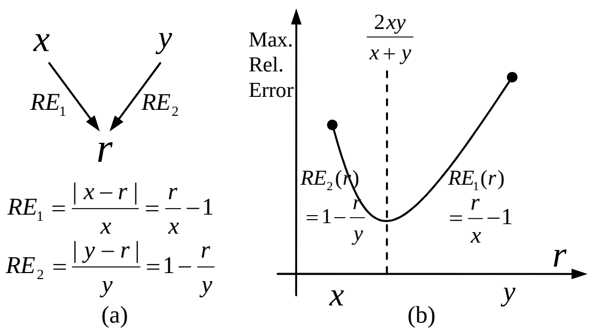

Figure 4. Relative errors for the selection of r

Given two values x and y , Figure 4-(a) shows the relative errors generated at x and y each denoted by RE_1 and RE_2 when they are replaced with the value of r . (For brevity, we assume that the value of x is less than that of y .) Then, we can get the maximum relative error between RE_1 and RE_2 denoted by MRE .

If we define these relative errors as the function of r , the maximum relative error function $MRE(r)$ can be obtained as the combination of two error functions, denoted by $\max\{RE_1(r), RE_2(r)\}$. And the graph in Figure 4-(b) illustrates the function of $MRE(r)$ as the value of r is varied from x to y . As describe in Figure 4-(b), the function $MRE(r)$ has the minimum value when the functions, $RE_1(r)$ and $RE_2(r)$, have the same value. In other words, the maximum relative error is minimized when the value of r is $2xy/(x+y)$ which is the harmonic mean of x and y. So the harmonic mean is the minimum point of the maximum relative error between two data values.

3.2.3 Harmonic wavelet decomposition: As mentioned previously, Harmonic wavelets transform two original data values to their harmonic mean and the relative error at each decomposition step. The basics of the Harmonic wavelet transform are similar to the Haar wavelet transform.

The Harmonic wavelet decomposition also can be a series of multi-level operation for the harmonic mean and the relative error. This procedure is continued until the overall harmonic mean is obtained. Table 2 shows the Harmonic wavelet transform for the data A = [1 3 10 6].

Table 2. Example of Harmonic wavelet decomposition

resolution	harmonic means	detail coefficients
2	[1 3 10 6]	-
1	[1.5 7.5]	[-0.5 0.25]
0	[2.5]	[-0.66]

Like the Haar wavelet decomposition, we can get the low resolution data by replacing pairwise data values with their harmonic mean. For example data [1.5 7.5] in resolution 1 are obtained by taking the harmonic mean of pairs [1 3] and [10 6] in resolution 2. And detailed coefficients [-0.5 0.25] are the relative errors made from the first value of each pair. This Harmonic wavelet decomposition is finished when the overall harmonic mean, the value of 2.5, is obtained. Finally, we can get the Harmonic wavelet transform W_A = [2.5 -0.66 -0.5 0.25] for the data A .

3.3 Greedy Algorithm: DP algorithm for selecting optimal synopsis needs $O(N^2B)$ space and $O(N^2B\log B)$ time[3]. So the efficiency of DP algorithm will be reduced as the size of data is increased. In this subsection, we propose the greedy algorithm reducing the time complexity to $O(NB\log B)$. This greedy algorithm is based on the error tree consisting of root node which is the overall harmonic mean and internal nodes which are the detailed coefficients as shown in Figure 5. The difference from previous error trees is that the value of leaf node indicates the current relative error. So, the maximum relative error can be obtained by sending up the maximum relative error at each subtree from leaf node. The value of 1.34 is the maximum relative error for example, in Figure 5.

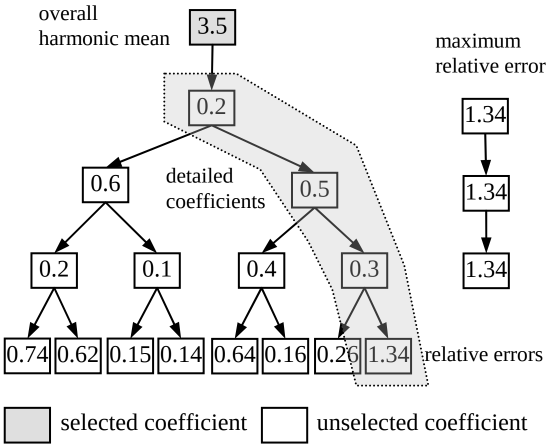

Figure 5. Greedy Algorithm

Given the maximum relative error, we can retrace the path generating the maximum relative error as illustrated with the polygon in Figure 5. By selecting the most reducible coefficient in the path, we can minimize the maximum relative error. At this example, 0.5 is the most reducible coefficient in the path [3.5, 0.2, 0.5, 0.3]. The details of calculations are omitted for the limitation of paper.

4. EXPERIMENTAL RESULTS

As an experimental data, we use the Cover Type data supported by National Forest Service. This data is composed of 581,102 numbers of values and each value is in the range of 0 and 255. Figure 6 shows the experimental results of harmonic wavelets with other algorithms.

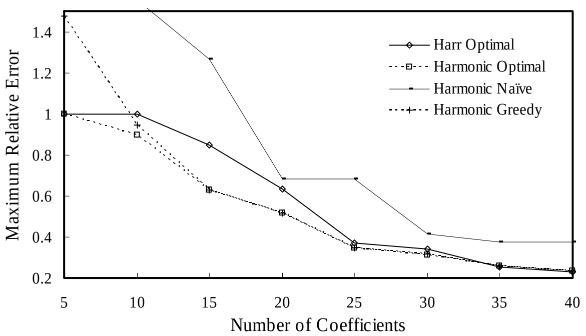

Figure 6. Experimental results for given memory bound

Harmonic optimal algorithm shows the best performance including this result, which outperforms Haar optimal algorithm by 30%. However, these two algorithms are not adequate to manipulate a large amount of sensor data for their high time complexity. Surprisingly, Harmonic greedy algorithm obtains the almost same results with Harmonic optimal algorithm as shown in Figure 6. It means that Harmonic greedy algorithm is more useful in a circumstance where the response time is more important than the accuracy like sensor network. So we can conclude that Harmonic greedy algorithm is suitable to handle a large amount of sensor data. In addition, we can also make certain that the Harmonic wavelet-based algorithms approximate more compactly than Haar wavelet-based algorithms on the side of the relative error.

5. CONCLUSION

As the importance of the sensor data processing is increased, the interest of approximation schemes is also raised. Wavelets, one of approximation schemes, are well known for the effectiveness of the data compaction. And it is pointed out that wavelet synopsis with the L^2 error minimization cannot guarantee the individual error bound. As a solution to this problem, we proposed the Harmonic wavelet transform on the side of the relative error. By using the harmonic mean, we could reduce the maximum relative error, and by adapting greedy algorithm, we could improve the time complexity. Experimental results confirmed the performance of the Harmonic wavelet transform compared to previous works.

6. REFERENCES

[1] K.Chakrabarti, M.Garofalakis, R.Rastogi, and K. Shim. “Approximate Query Processing Using Wavelets” In Proc. Of the 26th Intl. Conf. on Very Large Data Bases, pages 111-122, September, 2000.

[2] Minos Garofalakis and Phillip B. Gibbons. “Wavelet Synopses with Error Guarantees”. In Proc. Of the 2002 ACM SIGMOD International Conference on Management of Data, pages 476-487, June 2002.

[3] Minos Garofalakis and Amit Kumar. “Deterministic Wavelet Thresholding for Maximum-Error Metics”, In Proc. Of the 2004 ACM PODS, Paris, France, June, 2004.

[4] Eric J. Stollnitz, Tony D. DeRose, and David H. Salesin. “Wavelets for Computer Graphics – Theory and Applications”. Morgan Kaufmann Publishers, San Francisco, CA, 1996.

7. ACKNOWLEDGEMENT

This work was supported in part by the Brain Korea 21 Project and in part by the Ministry of Information & Communications, Korea, under the Information Technology Research Center (ITRC) Support Program in 2005.