Automatic UWB Clusters Identification

Michael Corrigan, Asia Walton, Weihong Niu and Jia Li

Oakland University, Rochester Hills, Michigan 48326, USA

Timothy Talty

General Motors Corporation, Warren, Michigan 48092, USA

Abstract—**Clustering phenomenon exists in the Ultra-Wideband (UWB) impulse responses. Although it is feasible to manually identify clusters via visual inspection, this task becomes very difficult and time consuming when a large amount of data needs to be processed. Furthermore, visual inspection highly depends on the person who performs the cluster identification task, which may cause inconsistent or unrepeatable results. In this paper, we propose an automatic procedure to identify clusters in UWB impulse responses. The algorithm takes both the amplitude and time delay into account when finding the clusters. The algorithm performs well. If the threshold is set correctly, there will not be any error.**

Index Terms — **UWB, cluster identification, intra-vehicle, multipath**

I. Introduction

Due to multiple paths or the transceiver motions, overlapping and fading exist in the UWB propagation. The Saleh-Valenzuela (S-V) model is widely referenced in the UWB literature to characterize channel impulse response [1]-[2]. One of the main characteristics described by the model is that paths arrive in clusters. A cluster is defined as a group of multipath components with similar arrival times and exponentially decaying amplitudes. The first step to derive the S-V model parameters which characterize the UWB propagation is to identify these clusters from the UWB impulse responses. In the UWB literature, a lot of papers stated that the identification of clusters was performed manually via visual inspection [3]-[4]. This method can be easily performed on relatively small sets of measurement data, however, it is very difficult and time consuming when processing large amount of impulse responses. In addition, different person inspecting the same data may create different cluster identification results. An automatic cluster identification algorithm will fix the above problems. A few studies on automatic cluster identification have been made in the UWB literature [3]-[4]. The automatic clustering algorithm discussed in [3] involves the setting up of various criteria for the definition of a cluster. Because the criteria proposed in [3] are quantitative, the algorithm must take lots of user input to initialize the program, which makes the procedure semi-automatic. In our UWB channel measurement campaign, we have to process thousands of impulse responses for statistical analysis. This motivates us to develop more automatic procedures for cluster identification.

The focus of this paper is an algorithm that can find clusters automatically in channel measurement data. This algorithm deals with filtering the clusters, setting the threshold and finding the accurate clusters. Our experiments deal with UWB impulse responses with clustering phenomenon so as to speed up the cluster identification process.

The paper is organized as follows. In Section II the cluster identification problem is defined. In Section III, the automatic cluster identification algorithm is explained in detail. In Section IV we demonstrate the algorithm by applying it to a realistic UWB impulse response. In Section V we conclude the paper.

II. Problem Definition

The initial point is that we have a large set of impulse responses obtained from intra-vehicle UWB channel measurements. It is observed that paths tend to arrive in clusters. In order to estimate the S-V model parameters to characterize the intra-vehicle UWB channel, we must find a way to identify the clusters with acceptable accuracy. It is always possible that clusters are identified manually with visual inspection. However, it is very impractical to identify clusters manually for a very large amount of data like ours. The workload can become frustrating and overwhelming. This is the key reason motivating us to find an automatic way to identify the clusters.

Our measurement experiments were performed in time domain. When inspecting the clusters in the impulse responses derived from the measured data, we found that the strongest paths or the peaks do not always arrive at the beginning of the power delay profiles (PDP) of the impulse responses. Therefore, instead of having a sharp beginning, each cluster may own a rising edge. This observation makes it difficult to recognize the starting point of such kind of clusters.

978-1-4244-2699-7/09/$25.00 ©2009 IEEE

Multipath propagation is the fact that paths arrive at the receiver with different time delays. It leads to the overlapping of neighboring clusters in the impulse responses. This situation creates a problem for identifying a correct cluster because it is difficult to identify where the clusters begin or end.

In addition, there is still some noise (unwanted paths) in the impulse responses and this can form fake clusters. This adds the difficulty to recognize real clusters. We have to find a way to remove these fake clusters.

III. Cluster Identification Algorithm

The proposed cluster identification algorithm takes both the time distance and the variations in amplitude into account when recognizing clusters. In the previously reported efforts, the algorithms required many user inputs and made the results more subjective [3]. The algorithm proposed in this paper requires much less inputs from the user. Before any raw measurement data is input to the algorithm, the received multipath signal waveforms are first deconvolved using CLEAN algorithm to extract the impulse responses. CLEAN algorithm assumes that the received signal is a sum of the same shape pulses arriving at different time and with different strengths. Details of the CLEAN algorithm can be found in [5].

Our algorithm has two main steps. The first step is to find clusters using time distance. Then each cluster found in first step is further broken down into smaller clusters using the variations in the amplitudes.

A. Clustering Using the Time Delay

The first step in this algorithm is to find the non-zero paths in the data. P_i will represent the non-zero points, where i is the index of the non-zero point and P is the magnitude.

Next, P_i is clustered according to the amount of time delay between P_i and P_{i+1}. The time delay between P_i and P_{i+1} is found for all i. This is set to D_i. The user will set a threshold, T, for the amount of delay that will be allowed. If $D_i > T$ a new cluster will begin. If $D_i < T$ no action will be taken. Because as time progresses the impulse responses tend to become more separated, it is possible to find that a path is the only path in its cluster and its amplitude is small. In such case, the path will join the previous cluster.

B. Clustering Using Amplitude

The SV model shows that the beginning of clusters should have a large increase in intensity. [1] This property is used to identify the clusters. A line of best fit and the error associated with that line can help determine where this large increase in intensity occurs, and consequently the beginning of the cluster. The line of best fit is found using P_i. The line of best fit is found using the least squares method. This method finds a line where the sum of the squares of the error is minimized. If the error is small, all the data points are close to the line and there are no large deviations from the line. A large deviation could represent a beginning of a new cluster.

With the best line defined above, the goal of this section is to find groups of lines of best fit to cover the data set. The beginnings of new lines will signify the beginnings of clusters. The user will manually select a threshold, T. T will be based off how large the deviations of P_i are to the line of best fit. If T is too big, the large deviations from the line will not be caught. If T is too small, the beginning of a cluster will be identified incorrectly. A flowchart of this procedure is in shown in Fig. 1. This procedure is repeated until all P_i is placed into a cluster.

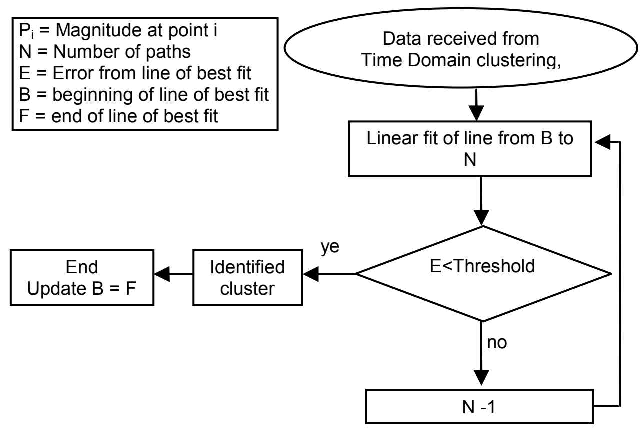

Fig. 1. Flowchart of cluster identification using amplitude.

The lines of best fit are found using the least-squares method. This method minimizes the squared distance x_i that each data point has from the line of best fit. The error used in the algorithm is the norm of the residuals. The norm of the residuals is the root mean square of the vertical distances of the data points to the line of best fit.

$$\textit{norm of the residuals} = \sqrt{x_1^2 + x_2^2 + x_3^2 + \ldots} \qquad (1)$$

After this procedure is carried out, the lines of best fit will be checked to see if any line has a positive slope. According to the SV model the power of signals must always be decreasing. This means if the lines of best fit correctly identified a cluster, they must have a negative slope. Otherwise the identification result is incorrect. This incorrectly identified cluster is usually the rising edge of a cluster where the strongest path is not at the beginning. To correct this, the cluster where the slope of the line of best fit is positive is combined with the cluster after it. This step is not shown in the flowchart.

IV. EXAMPLE

We tested the proposed algorithm on the intra-vehicle UWB channel measurements. This algorithm was able to correctly identify the number of clusters in our data and where the cluster began and ended. This section will show an example of applying the algorithm on one of the UWB impulse responses.

Fig. 2 shows the UWB channel impulse response extracted by the CLEAN algorithm, which is the input to the cluster identification algorithm.

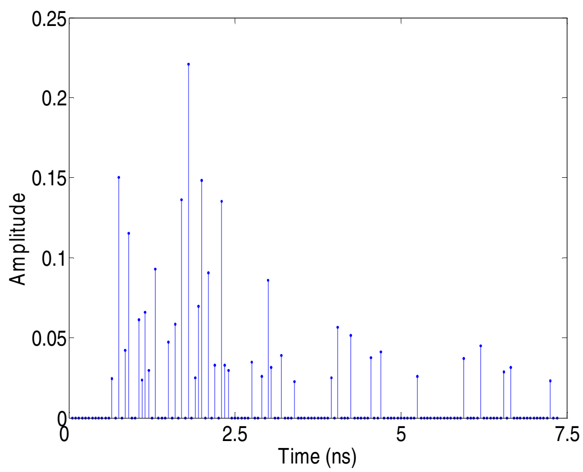

Fig. 2. Example of deconvolved signal.

D_i , the amount of time delay between P_i and P_{i+1} , is calculated. The data was split into clusters through threshold. The threshold that was used was T = 8. The result is shown in Fig. 3. The lines indicate the clusters.

Each cluster shown in Fig. 3 is further clustered based on amplitude. The error threshold that was used was 0.1. The first cluster, which was identified as 0.65ns to 2.3ns by the time delay, was clustered as 0.65ns to 1.3ns and 1.5ns to 2.3ns. The other clusters remained one cluster. Fig. 4 shows the lines of best fit for the cluster 0.65ns to 2.3ns. Notice that the second line has a positive slope. This incorrectly identified cluster was consolidated with the cluster following it. The final clusters with respect to both time and amplitude are shown in Fig. 7.

The accuracy of this algorithm was compared to the algorithm described in [3]. In cases where there were large time delays between clusters both algorithms would be able to identify clusters correctly. However, when the time delays between clusters that are not very large, our algorithm is able to identify the start of the cluster more accurately. This is shown in Fig. 5 and Fig. 6. The reason that our algorithm identifies the start of the cluster more accurately is that the start of the cluster may be a large time delay away from the maximum path. Also the start of the cluster could be small in magnitude.

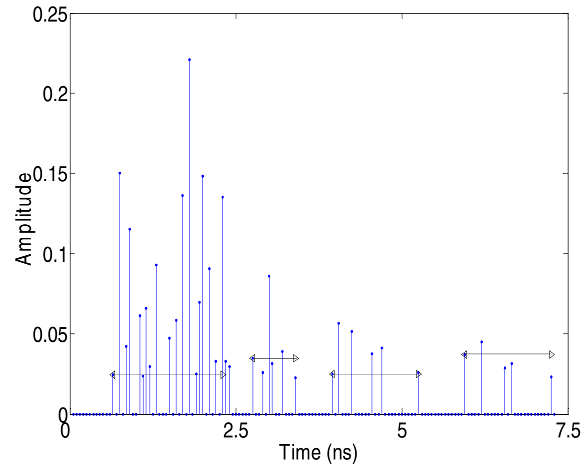

Fig. 3. Cluster identification according to the time delay.

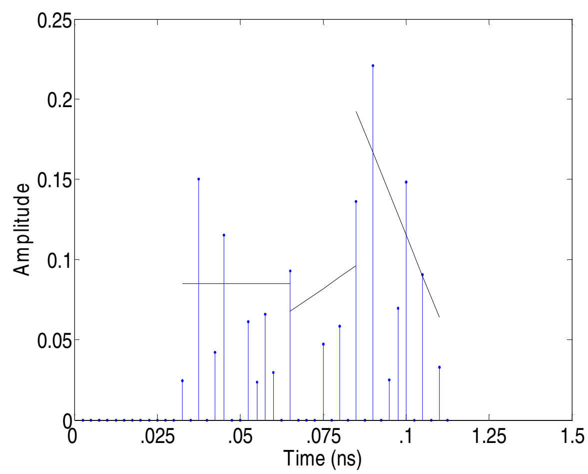

Fig. 4. Cluster identification with respect to amplitude. The lines of best fit indicate the beginning and end of a cluster.

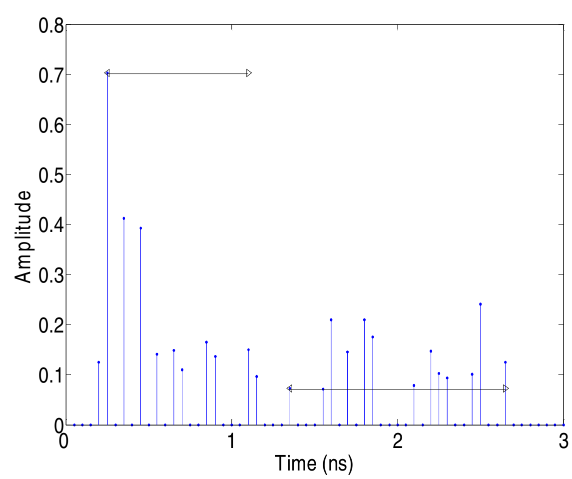

Fig. 5. The clusters using the algorithm described in [3].

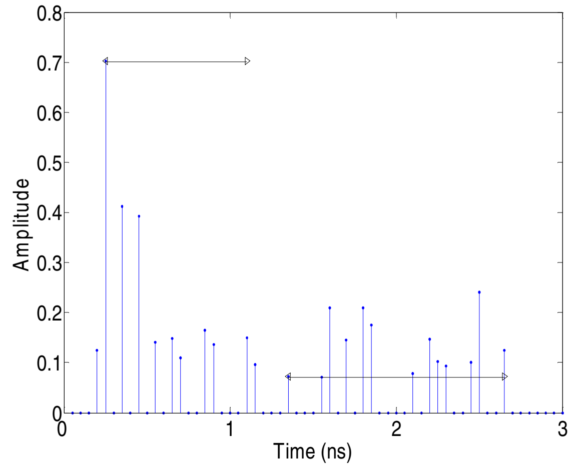

Fig. 6. The clusters using our algorithm.

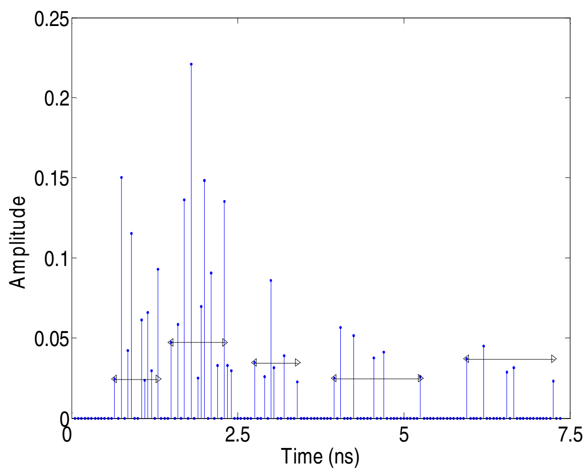

Fig. 7. The final clusters. The horizontal lines mark the clusters.

V. CONCLUSION

Our UWB measurement experiments produced a large amount of data. We need to identify the clusters in these data. It is impractical to handle such a large amount of data manually. Therefore we designed an algorithm that would identify these clusters automatically. While implementing this algorithm we make sure that it identifies them accurately, as it would be if the identified clusters were done manually. This algorithm was able to take both the time delay and amplitude into consideration into clusters. The result of implying this algorithm to our measurement data shows that it works accurately.

ACKNOWLEDGEMENT

This work is supported by the National Science Foundation under Grant No. 0721813. The authors would like to thank Dr. F. Mili for her encouragement and care. The authors also wish to give their appreciation to all reviewers for their invaluable time to review this paper.

REFERENCES

[1] A. Saleh, and R. A. Valenzuela, “A statistical model for indoor multipath propagation,” *IEEE Journal on Selected Areas in Communications*, vol. 5, pp.128-137, Feb. 1987.

[2] D. Shutin, “Cluster Analysis of Wireless Channel Impulse Responses,” *Int. Zurich Seminar on Communications,* pp.124-127, Feb. 18-20, 2004.

[3] O. H. Woon and S. Krishnan, “Identification of Clusters in UWB Channel Modeling,” *Proc. Of IEEE Vehicular Technology Conference 2006,* pp.1-5, Fall, 2006.

[4] J. Salo, J. Salmi, N. Czink, and P. Vainkainen, “Automatic Clustering of Nonstationary MIMO Channel Parameter Estimates,” *International Conference on Telecommunications’05*, May 2005, Cape Town, South Africa.

[5] R. G. Vaughan and N. L. Scott, “Super-resolution of pulsed multipath channels for delay spread characterization,” *IEEE Trans. on Communications,* vol. 47, pp. 343-347, March 1999.