An Automated System for Rail Transit Infrastructure Inspection

Final Report

Under Cooperative Agreement
No. RITARS-12-H-UML

Between

U.S. Department of Transportation
Office of the Assistant Secretary for Research and Technology

And

University of Massachusetts Lowell
Office of Research Administration
600 Suffolk Street
Lowell, MA 01852

Program Manager
Mr. Caesar Singh

Project Principal Investigator
Yuanchang Xie, University of Massachusetts Lowell

Report Submitted on: March 29, 2015

DISCLAIMER

ACKNOWLEDGMENTS

EXECUTIVE SUMMARY

This project applied commercial remote sensing and spatial information (CRS&SI) technologies such as Ground Penetrating Radar (GPR), laser, GIS, and GPS to passenger rail inspections. An integrated rail inspection system that can be mounted on hi-rail vehicles has been developed. This integrated system consisted of four major components: *GPR subsystem, laser subsystem, track dynamic model subsystem,* and *a web-GIS based Decision Support System (DSS)*. The GPR subsystem was designed to identify track subsurface problems such as fouled ballast and suspicious underground objects. It was developed using a set of home-made horn antennas; the laser subsystem was for detecting surface track defects such as missing fasteners and cross ties, large cracks in cross ties, and wide rail gauge. It was developed based on a commercial laser product called LCMS; the track dynamic model was developed to identify a critical track structure problem called hanging tie using spectrum analysis. The main instrument for this subsystem was a 3-axial accelerometer; and the web-GIS based DSS was developed using ArcGIS for Server and Microsoft SQL Server. Its purpose was mainly to manage and visualize the results generated by the previous three subsystems.

The first three subsystems were integrated and mounted on a hi-rail SUV. This automated and operational system was then tested at both Massachusetts Bay Transportation Authority (MBTA) and Metro St. Louis. The team also developed algorithms for processing the GPR, laser, and track dynamic model data. These algorithms include a 2D entropy method for GPR data analysis and a 3D template matching method for identifying missing fasteners. Among them, the laser algorithms have already been commercialized by Pavemetrics as LRAIL. The data collected from MBTA and Metro St. Louis was processed by these algorithms and fed into the DSS. In addition to popular GIS tools for visualizing, querying, and editing spatial and attribute data, the project team developed a mobile App to facilitate field asset inspections.

Rail transit agencies in the United States rely heavily on visual observation for their weekly track inspections. This manual method is time-consuming, costly, and cannot effectively identify subsurface safety hazards. The project team reached out to several major rail transit agencies in the United States and demonstrated the developed product. These demonstrations generated substantial interest among these stakeholders. With the aging rail infrastructure, this developed system is expected to substantially benefit the rail transit industry by improving the track inspection efficiency, accuracy, and safety of both the rail transit systems and track workers. During the course of this project, the team identified additional interesting and innovative ideas. Some of them have been successfully implemented such as the mobile App. These ideas are also discussed throughout this report.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

A 2008 United States Federal Transit Administration (FTA) report [*1*] estimated that about 25% of the rail and bus assets in the United States are close to or beyond their useful life. For the nine largest transit agencies, the corresponding percentage is nearly 35%. The National Transportation Safety Board (NTSB) is seriously concerned about the safe operations of the current rail transit industry after investigating the July 2006 Chicago Transit Authority Blue Line derailment accident, concluding that the accident was caused by the failure of the tie plates and fastening systems. A member of the NTSB warned that this accident was a "*wake up call … to all transit agencies … with equipment and infrastructure that age with each passing day.*" Defects in railroad tracks are also responsible for thousands of accidents on an annual basis around the world. In Canada, in 2012 alone a total of 1,011 rail accidents were reported to the Transportation Safety Board (TSB) [*2*] with a large percentage being attributed to track defects. In the United States, the Federal Railroad Administration (FRA) reported that track defects are responsible for more than 33% of all registered rail accidents. Thus, the reliable detection of defects in railroad tracks is of great importance for both freight and transit rail safety.

1.1. Current Practices

A critical aspect of the transit state of good repair is the inspection of rail transit infrastructure. The current rail transit inspections are primarily based on visual observation, which is time-consuming and labor-intensive. Also, they cannot effectively identify subsurface hazards. For instance, at Metro St. Louis, ultrasound is used for rail inspection twice a year, and an inspection vehicle is used to check the rail geometry every few months. For all other structures and facilities (e.g., ties and fastening systems), they are inspected manually on a weekly or monthly basis, which requires considerable time and efforts. This practice is very typical for other rail transit agencies in the United States.

Remote sensing technologies such as laser and Ground Penetrating Radar (GPR) have been used by many state Departments of Transportation (DOTs) for highway pavement and bridge inspections. The unique ability of GPR to rapidly evaluate subsurface conditions has been demonstrated through various applications, especially for highway pavement inspections [*3,4,5,6,7,8,9,10*]. Theoretically, GPR can also be used to inspect rail transit infrastructure (e.g., concrete ties). However, its applicability for rail inspections so far has not been thoroughly investigated. Another promising remote sensing technology is laser, which has been tested in many states such as Tennessee, Texas, and Georgia for mapping highways and identifying pavement defects. Due to the encouraging results obtained, laser recently has received increasing attention from state DOTs [*11,12*]. However, its applications in rail transit industry are limited. Most applications used laser for measuring railway clearance [*13*] and wheel/track profile [*14,15*], which were different from the proposed study. Some other studies [*16,17,18*] applied video cameras for rail inspections. Although companies such as ENSCO and Holland have developed trains and trucks for track inspection that utilize laser technology, it would be too expensive to use these trains and trucks for the weekly track inspections required by most rail transit agencies.

1.2. Research Objectives

Owing to the apparent need and the lack of research, this study proposes to develop an automated system for rail transit infrastructure inspections. This new system integrates several existing Commercial Remote Sensing & Spatial Information (CRS&SI) technologies: GPR, laser, accelerometer, Geographical Information Systems (GIS), and Global Positioning System (GPS). As shown in the figure below, the integrated system is portable and can be easily mounted on a hi-rail vehicle that virtually all rail transit agencies own. The radar system (GPR) is for detecting subsurface defects (e.g., fouled ballast) and objectives (e.g., underground pipelines); the laser system is for detecting surface problems such as missing fasteners, wide/narrow gauge, broken ties, and missing ties; the accelerometers provide input to a track dynamic model to identify hanging ties; algorithms have been developed to process the collected radar, laser, and accelerometer data; finally a decision support system has been developed to manage and visualize the processed data.
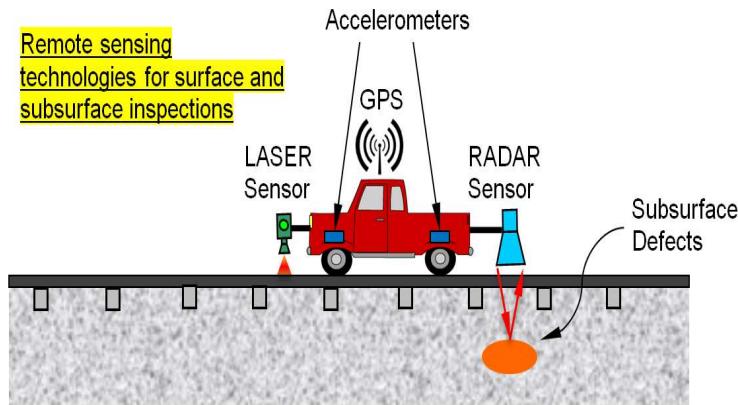
Figure 1 Proposed automated rail inspection system.

This proposed research well matches the CRS&SI program's objectives, particularly "***Apply concepts of CRS&SI technology integration and products to monitor the condition of constructed infrastructure systems and provide quantitative data to reduce/supplement components of current visual inspection procedures.***" The proposed system has the potential to enable rail transit agencies to accurately monitor the conditions of rail infrastructure with much reduced time and monetary costs, provide quantitative structure surface and subsurface information for further analysis, and help optimally prioritize transit asset renewal and replacement activities. It is more effective and efficient than the current visual inspection procedures.

1.3. Research Approaches

To achieve the above objectives, the following subsystems were developed and integrated: GPR, laser, and track dynamic model. The team designed and developed the radar subsystem from scratch. This included both hardware and software systems. The radar subsystem has been extensively tested in the lab before being implemented in the field. This study utilized an off-the-

shelf laser product, Laser Crack Measurement System (LCMS), for the laser subsystem. The LCMS was developed by Pavemetrics Inc. in Canada and has been widely used for highway pavement inspections in many countries. Algorithms were developed to adapt this laser system for rail infrastructure inspection. A track dynamic model has been developed that takes the accelerometer data as the input to identify hanging ties. Once the radar hardware design was finalized, it was mounted on a hi-rail vehicle and was integrated with the laser and accelerometer. This integrated system was tested at both Massachusetts Bay Transportation Authority (MBTA) and Metro St. Louis. For the decision support system, the team compared various Geographic Information Systems (GIS) products and finally adopted ArcGIS for Server. In addition to a web-based decision support system, the team also developed a mobile App for rail infrastructure inspection. In the remaining chapters of this report, these research efforts are described in detail.

CHAPTER 2

LASER SENSOR AND SOFTWARE DEVELOPMENT

2.1. Principles of 3D Laser Ranging

Laser ranging, often known as lidar, has seen rapid increases in use for sampling 3D environments across scales from a few millimeters up to the whole globe [*19,20*]. Examples of applications include detecting surface cracks in industrial facilities, scanning 3D urban cities, and mapping global vegetation height [*21*]. In principle, existing techniques for laser ranging can be roughly grouped into two categories: those based on the time-of-flight mechanism versus those based on laser triangulation [*22,23,24*].

The time-of-flight laser systems pinpoint objects and determine their locations by emitting a laser pulse and timing the round-trip of the pulse between the laser and the target (see Figure 2). The round-trip travel time is then converted to a distance based on the speed of light, which is combined with the position of the laser to calculate the position of the target. As such, the accuracies in ranging and positioning achieved by time-of-flight laser systems depend largely on the preciseness of timing. Generally speaking, determining the travel time for light is difficult, and the accuracies of distance measurements are often on the order of millimeters. The actual accuracies in ranging are further degraded, attributable to the instability of mobile platforms [*19*]. This level of accuracies is generally sufficient for geosciences and engineering applications, such as measuring topography over large areas, but is not precise enough for many industrial applications, including the proposed tasks in this project.
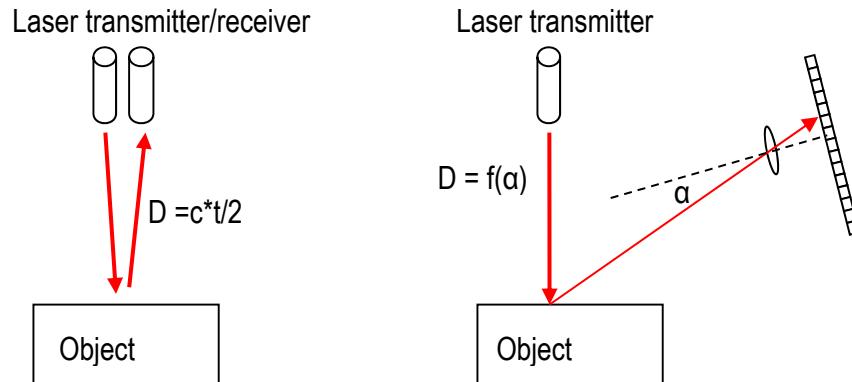
Figure 2 Two typical lidar systems employed for engineering and environmental survey: Time-of-flight vs. tri-angulation lidar.

Alternatively, laser ranging can be implemented via the triangulation method, a technique of centuries old [*25,26,27*]. As in Figure 2, when a point laser impinges on the target, the laser spot on the targeted will be imaged by a camera; then the laser, the laser spot and the camera forms a triangle; plus, the location of the imaged laser spot is determined directly by how far the laser is from the target and therefore can be inverted to compute the laser-target distance based on trigonometry. In general, such triangulation-based laser sensors have higher accuracies than time-of-flight lidars; in many cases, the achievable level of accuracies is tens of micrometers.

This accuracy level makes triangulation-based lidar a choice suitable for detecting defects in railway infrastructures.

In addition to employing a laser spot for 3D ranging, triangulation can be equally applied for laser profilers/stripers [28]. Specifically, the laser beam is passed through a cylindrical lens, which spreads the laser to cast a line instead of a spot on the target. The laser strip impinged on the target is then imaged by a 2D camera. Similarly, according to the triangulation principle, the relative locations of the imaged profile on the camera can be used to derive the true 3D profile on the target. A series of such individual profiles will be collected as the laser moves longitudinally as shown in Figure 3 below; the combination of these contiguous profiles will reconstruct the 3D surface of the target. This is the technology developed and used for this project, as further summarized next.
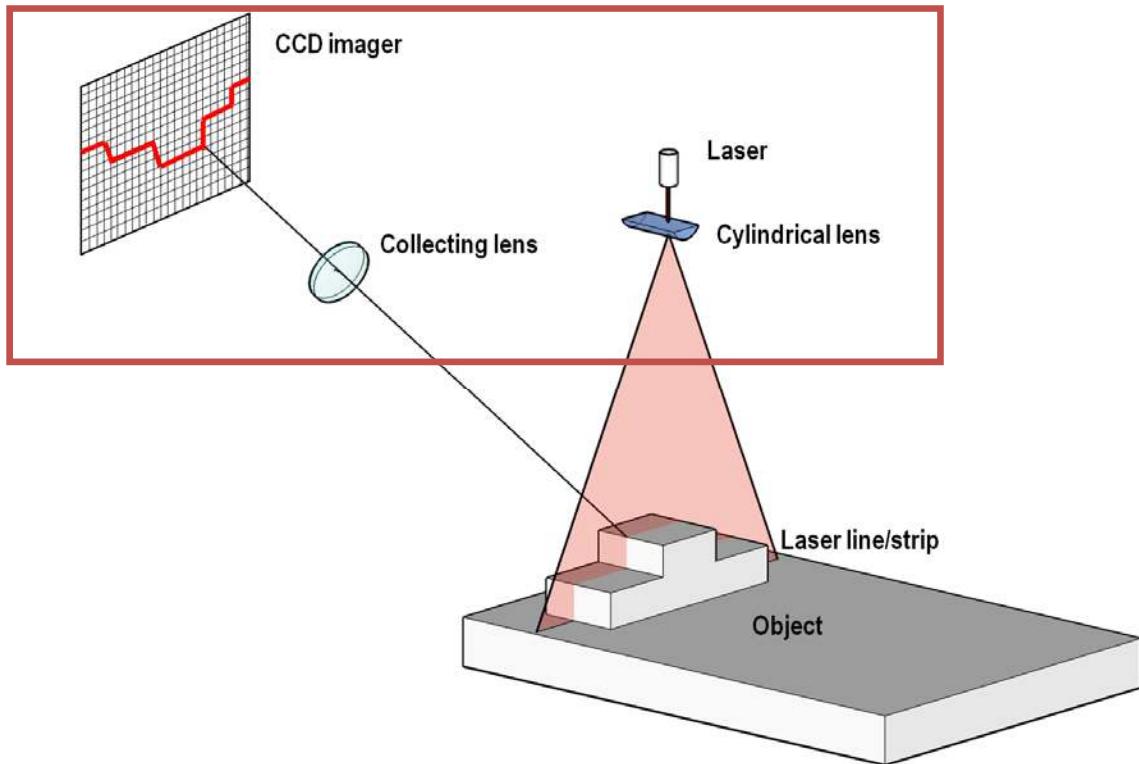
Figure 3 Schematic of a tri-angulation lidar to measure 3D structures of targeted. This configuration is known as laser profiler and serves as a prototype of the laser system "LCMS" we employed for this project: A line of laser is shed onto the targeted object, which is imaged by a camera; the relative positions of pixels imaged will be converted to their real 3D positions via some projective transforms. The major hardware components of the LCMS system are enclosed in the red rectangle.

2.2. Laser Sensor and Its Configuration

Rather than building a laser profiler from off-the-shelf components from the scratch, an existing laser scanning system, known as the Laser Crack Measurement System (LCMS), was chosen as a prototype platform for railway survey in this project [29]. LCMS is a triangulation-based 3D laser profiler. It has been previously used worldwide for highway pavement inspection

[*30*].
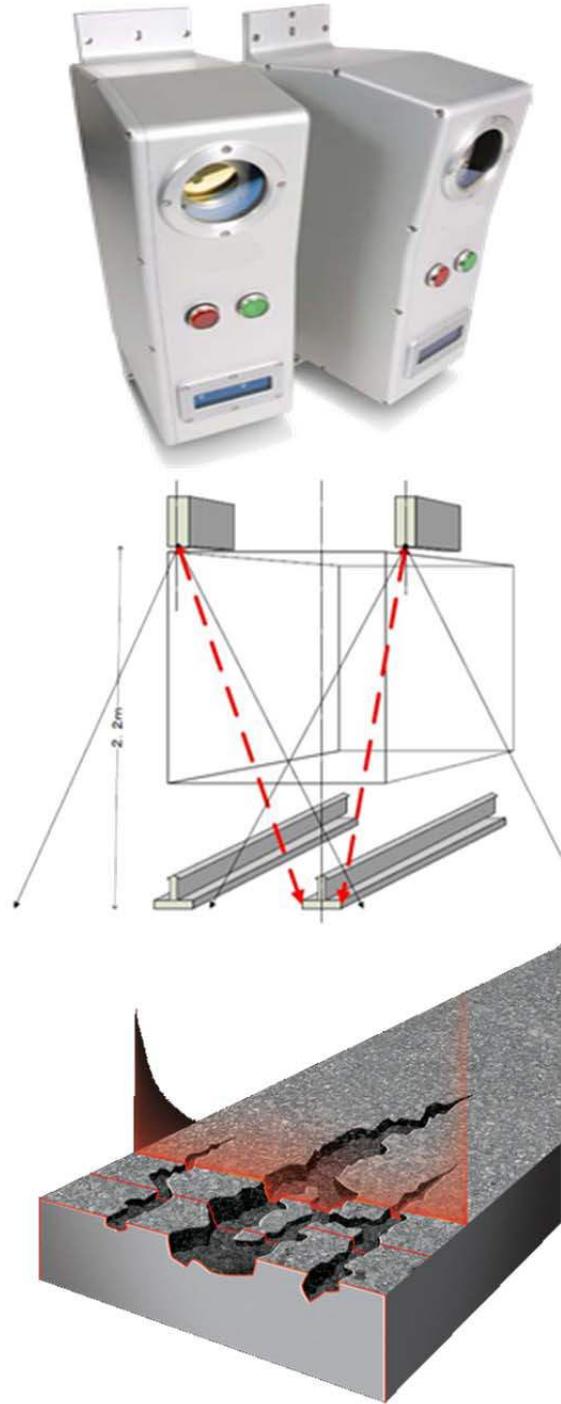
Figure 4 LCMS implements the prototype system in Figure 3. It consists of two sensors, one looking at the left and another at the right. Its configuration when deployed is depicted in the middle, and its use for detecting minute details, such as cracks, is illustrated (bottom). Its specific technical parameters are provided in Table 1.

More specifically, LCMS combines two 3D laser profilers that use high power laser line

projectors, custom filters and a camera as the detector [*31*]. During operations, a strip of laser light is projected onto the railroad track and its image is captured by the integrated camera as shown in Figure 4. The image from the cameras is sent to a frame grabber to be digitized and then processed by a high-performance computer to find the centerline of the imaged laser profile. The centerline can be converted via the triangulation method to a 3D profile of the target. As LCMS is carried by an inspection vehicle and is moved longitudinally over the target, a series of consecutive profiles will then coalesce into a 3d representation of the target. This 3D data comes in the form of a 2D raster image with the gray values representing the z coordinates or range information [*32*]. At the same time, the intensity of the imaged laser line/profile is also recorded, thereby providing another 2D raster image for intensity.

Key specifications for LCMS are summarized in Table 1. In particular, the system can be operated when the carrying vehicle travels at a speed up to 100 km/hr, with a maximum sampling rate of 11,200 profiles/s. The higher the sample rate is, the larger the data throughout. The data rate at the maximal speed of 100km/hr reaches about 150 GB/km, which, if not compressed, constitutes a bottleneck for data transfer and storage. As such, customized lossless data compression algorithms, such as fast JPEG compression, are applied on 3D raw ranging and intensity data, reducing the data rate to a manageable level of 5GB per kilometer.

Table 1 LCMS specifications.

No. of laser profilers	2
Sampling rate (max.)	11,200 profiles/s
Vehicle speed	100 km/h (max)
Profile spacing	Adjustable
3D points per profile	4,096 points
Transverse field-of-view	4 m
Depth range of operation	250 mm
Z-axis (depth) accuracy	0.5 mm
X-axis (transverse) resolution	1 mm

2.3. In-House Calibration and Validation of Laser Ranging Capability

$$\mathbf{x} = \mathbf{H}\mathbf{x}'$$
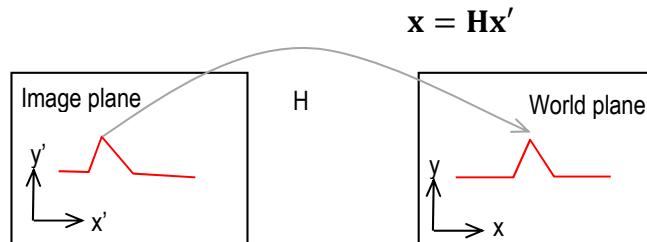
Figure 5 The 2D coordinate transformation from the image coordinate system to the world plane coordinate system, as represented in terms of a matrix H. This matrix is normally estimated for each LCMS sensor based on some known targets. Once estimated, the imaged lines will be transformed back to recover the true world lines.

In principle, there exists a one-to-one correspondence between the points/pixels on the image plane and the points on the laser-shed real-world plane. The correspondence has to be established mathematically, which is a process that depends on the interior orientation parameter

of the camera and the geometrical relationship between the camera and the laser projector. The process for building this correspondence is known as calibration. Once calibrated, the locations of the pixels at which the laser profile are imaged can be directly converted to get the true real-world locations of the associated profile. Algebraically speaking, the coordinate of a point on the image plane, x', is mapped back to that of the associated point on the world plane (i.e., the plane where the laser and its projected laser profile lie), x, via a 2D projective transformation H [*33*]. This process is illustrated in Figure 5.

More specifically, $\mathbf{x}'$ and $\mathbf{x}$, refer to the 3-vector homogenous coordinates of the image and real-world points; $\mathbf{H}$ is a 3 x 3 matrix with a total of 8 degrees of freedom. The transformation matrix H can be estimated from pairs of known image and real-world points, for example, using some calibration targets with known geometries. Many standard algorithms exist for estimating H, such as the famous direct linear transformation method and RANSAC-based nonlinear methods [*33*]. In addition to the use of H to relate $\mathbf{x}'$ and $\mathbf{x}$, another common method is to directly fit a polynomial relationship for calibration.

The nominal horizontal resolution and vertical accuracy for LCMS are 1 mm and 0.5 mm. To further assess and establish its horizontal and vertical accuracies, LCMS was validated in UMass Lowell lab through some experiments based on wooden plates and blocks. These wooden objects were piled and arranged to create many artificial configurations. Two sample configurations are depicted in Figure 6, one for vertical measurements and another for horizontal distance measurement. The horizontal and vertical distances of the wooden objects for each of the configurations were accurately determined using a high-precision ruler, serving as references to validate the laser measurements. Here, for comparison purposes, distance measurements were preferred over absolute coordinates because the LCMS coordinates are not absolute but relative. This is also because what matters most for practical applications is relative locations of interested features rather than their absolute coordinates.
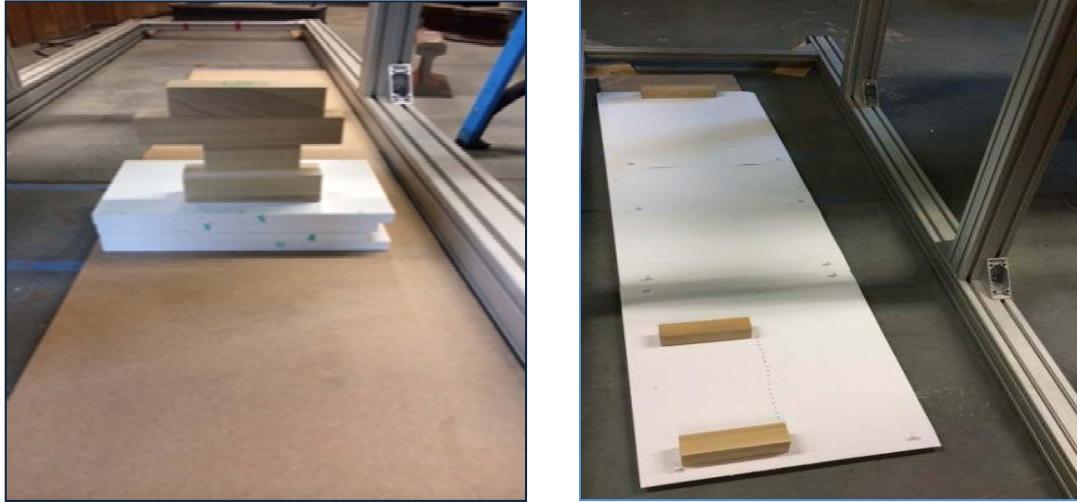
Figure 6 Two example settings we employed in UMass's lab to validate the ranging accuracies of LCMS, one for testing vertical accuracy (left) and another for horizontal (right). The height or the relative distances between the targets (i.e., wooden plates) are measured using rulers, serving as references to be compared against the laser measurements.

For comparisons between laser measurements and in-situ measurements, laser profiles for each configuration were pre-processed to derive the height and horizontal distances for the

targets of interest. As exemplified in Figure 6 and Figure 7, vertical distances/heights were derived as the difference between the top and the bottom of the wooden composite object using an iterative line-fitting method; horizontal distances were defined with respect to the edges of two wooden blocks, which were extracted using a simple threshold-based differencing approach. In total, 79 wooden composite objects were measured for vertical distances by LCMS; the resulting laser-measured heights match closely the in-situ values, with a root mean square error of 0.41mm. Fifty horizontal distances were obtained; these values compare favorably to the true reference measurements, with a RSME of 0.7mm.
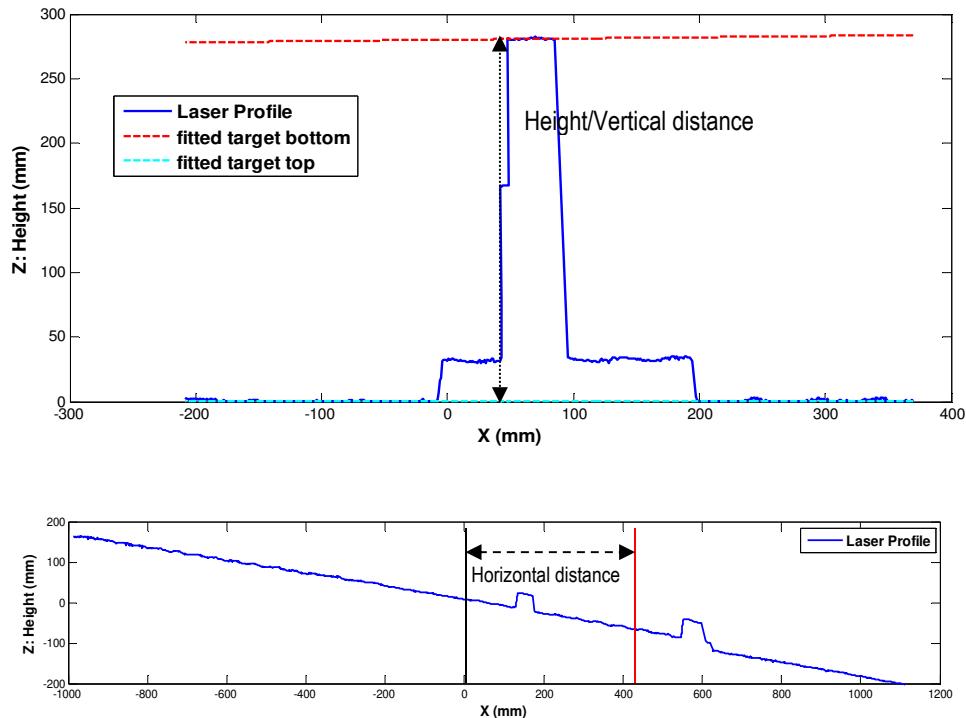
Figure 7 Examples of the acquired laser profiles, one corresponding to the vertical setting in Figure 6, another to the horizontal setting in Figure 6. The vertical height or horizontal distance, as depicted in arrows, were derived from the laser profiles using an algorithm that combines line fitting and difference estimator. Note that because of the sensor is tilted, the laser profile appeared titled too, as exemplified for the laser profile in the horizontal setting.

Both the horizontal and vertical accuracies, as determined by the lab experiment, accords well with the nominal resolutions of the system (see Figure 8). However, the RMSE values reported above are likely to be some conservative estimates for the following reason. In the lab experiment, LCSM was fixed and there were no relative motion between the system and the target whereas in practice, LCMS is always in motion. The movement of LCMS actually can increase the positioning accuracy. At first glance, this appears counter-intuitive because motion often introduces additional uncertainty into the positioning for many time-of-flight flight lidars, but for LCMS which is a triangulation-based lidar, the movement of LCMS relative to the target will smooth out specking noises of the imaged profiles, and correspondingly give a refined delineation of the centerlines of the imaged profiles. This increased precision in pinpointing image points then translates to improved accuracies in locating real-world points, as indicated by

the equation $\mathbf{x} = \mathbf{Hx}'$. For this reason, the errors in distance measurements obtained from our lab experiment are very likely to exaggerate the true uncertainty. Regardless, even with the exaggerated uncertainty, the level of accuracy is high enough for our purposes.
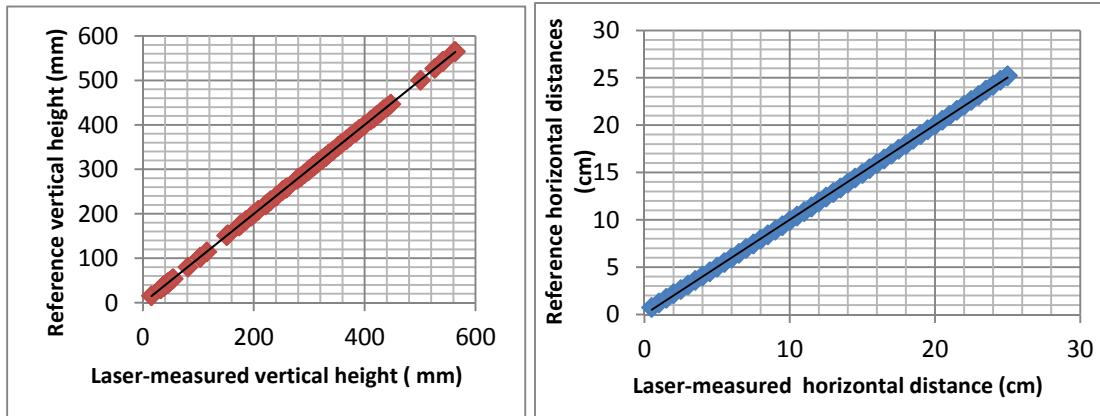
Figure 8 Comparison of laser-measured and reference target characteristics. The two match closely, both for vertical and horizontal ranging, therefore indicating the high fidelity of the ranging data acquired by LCMS.

2.4. Laser Data Collection

Figure 9 Deployment of the LCMS system on a high-rail track in a data collection campaign at St. Louis Metro in 2012.

Two major campaigns have been conducted to collect 3D laser data. To deploy LCMS, the system has to be mounted in a high-rail truck; also, LCMS has to be assembled together with other accessories, which, for example, includes a Dynapar Series HA725 Shafted Encoder, ad a Garmin 18 x 5Hz high-sensitivity GPS sensor, and one high-speed computer. The assemblage and installation was not as easy as it first appeared, due to the modification and customization to the high-rail truck as well as the limited budget for this task.

To address this practical challenge, our first data acquisition was conducted in December, 2012 through a partnership with a contractor called "Mandli" with the support from Pavemetrics (one of the project team members). This partnership afforded the project team an opportunity to

collect the laser data at St. Louis Metro (Metro) instead of MBTA, using a hi-rail pick-up truck customized by Mandli (see Figure 9). Prior to data collection, the Mandli staff calibrated the LCMS equipment; CodeRed Business Solutions (CRBS) helped the team to schedule and coordinate the right-of-way (ROW) safety training and ROW access with Metro's senior management staff. In this data campaign, the first test run was conducted along a segment track that has a majority of concrete ties at a maximum speed of 9.9 MPH, which collected the 3.3 miles' worth laser data at 1mm resolution. The second laser data collection test run commenced on December 18th from a truck pad near Lambert Airport to Central West End Station at 1mm resolution for approximately 10 miles. This segment of track has a majority of wood ties that were also subject to a tie replacement program. The third data collection test run was conducted along the same segment of track, but at a 4mm resolution that allowed a maximum speed of 21.9 MPH except at crossovers for approximately 3 miles. Sample laser data collected from Metro St. Louis are depicted in Figure 10.
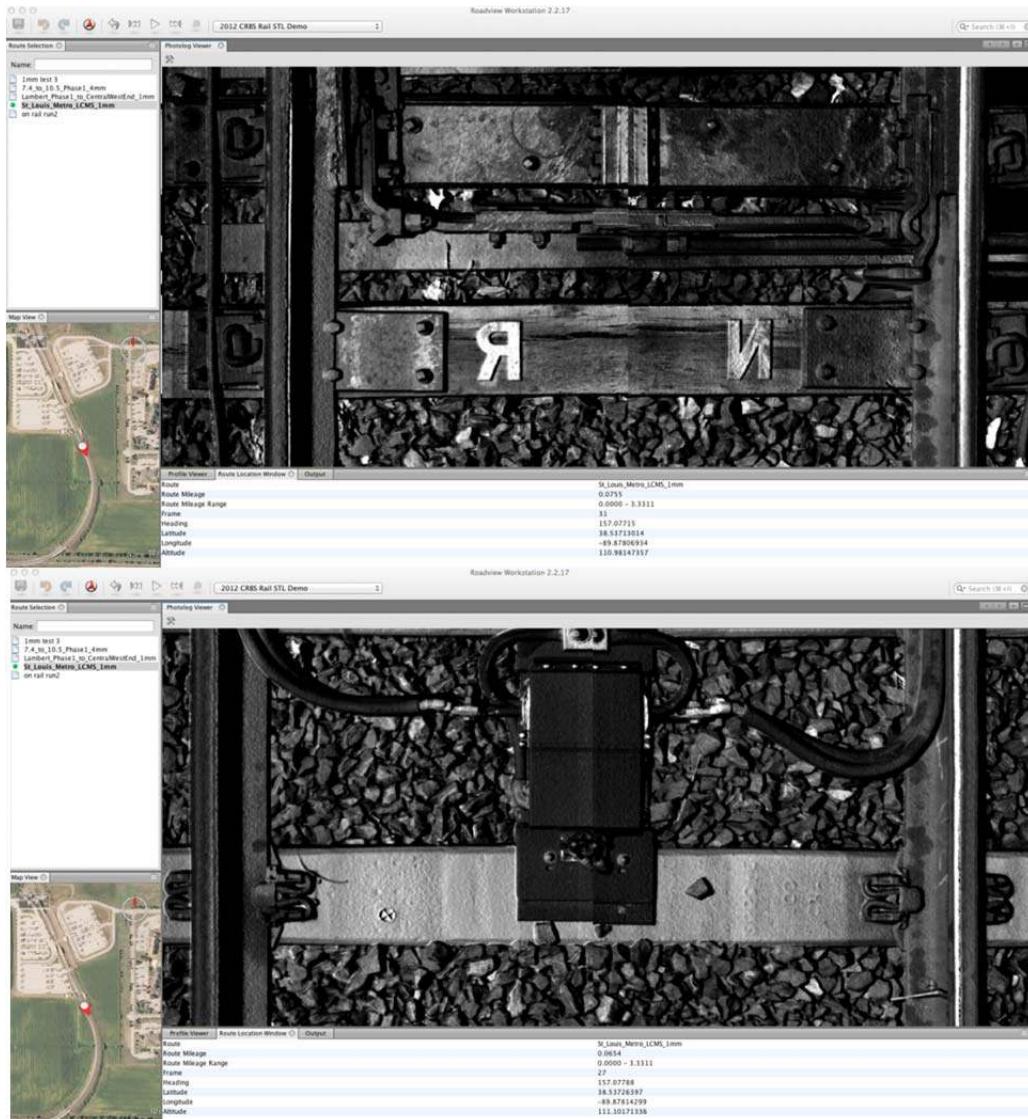
Figure 10 Laser data collection of wood ties at 1mm resolution at cross-over (top) and impedance bond mounted on s concrete tie at 1mm resolution (bottom).
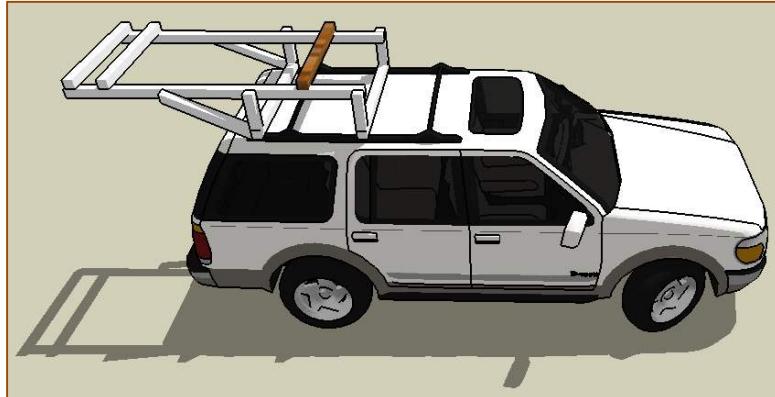
Figure 11 A rack designed specifically for the LCMS system

Our second data collection campaign was carried out in August 2013, both at MBTA and Metro St. Louis. The campaign completed the field tests of both the GPR and Laser systems. To deploy the LCMS system, a hi-rail rental SUV (Colvin's Aerial Trucks, MA) was modified and installed with customized racks for the laser system. The design of the rack was depicted in Figure 11. At MBTA, with the coordination from CodeRed, the hi-rail vehicle accessed the ROW to collect the laser and GPR data from Blandford grade crossing to Summit Street for approximately 6 miles traveling westbound and from Mattapan Station to Cedar Grove Station for approximately 3 miles traveling northbound. At Metro St. Louis, the data collection was conducted from Shiloh-Scott Station to Belleville Station for approximately 6 miles, and from Forest Park Station to Sunnen Station along the Cross County extension for approximately 6 miles. In addition, CodeRed assisted the project team in conducting the track gauge measurements and sample collection of ballast at multiple track segments at MBTA and Metro St. Louis. In total, about 200 GB of laser data have been collected from the two data campaigns.

2.5. Laser Algorithms for Feature Detection & Their Software Implementation

2.5.1. Data Pre-Processing
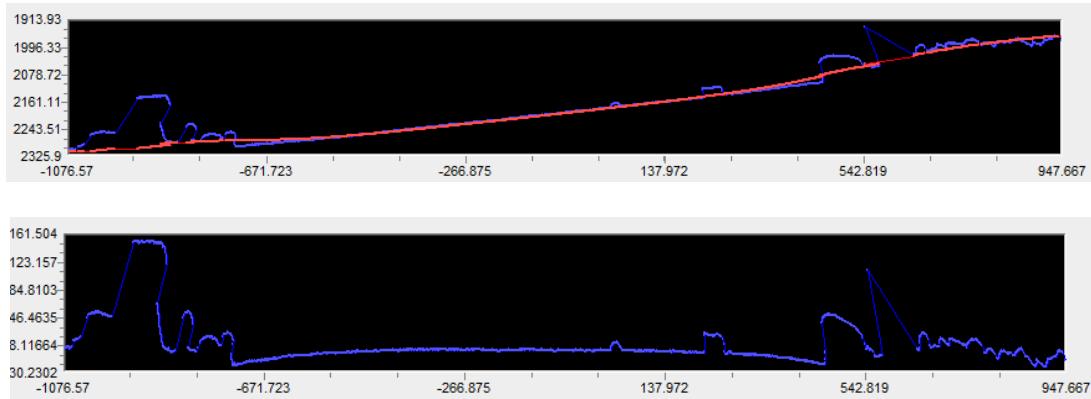
Figure 12 Detrending of individual laser profiles to rotate the coordinate system to a world-based system, as a way to improve the ease of future data processing and analyses for feature detection.

Before applying dedicated algorithms for feature detection, raw 3D laser data need to be

pre-processed. The pre-processing mainly considers three major aspects. First, the cross-section profiles obtained by the laser profilers are recorded in a sensor-based coordinate system and have to be corrected to get a meaningful interpretation. In the sensor-based coordinate, the rail cross-section profiles appear oblique, which is unsuitable for feature detection. Therefore, the sensor-based coordinate should be transformed to a world-based coordinate system wherein the 1D profile appears more or less flat in trend (See Figure 12 and Figure 13). This transformation was achieved by a rotation for an angle. Rather than determining the rotation angle theoretically, it was empirically determined via trial-and-error. After the rotation, features of interest become more conspicuous.
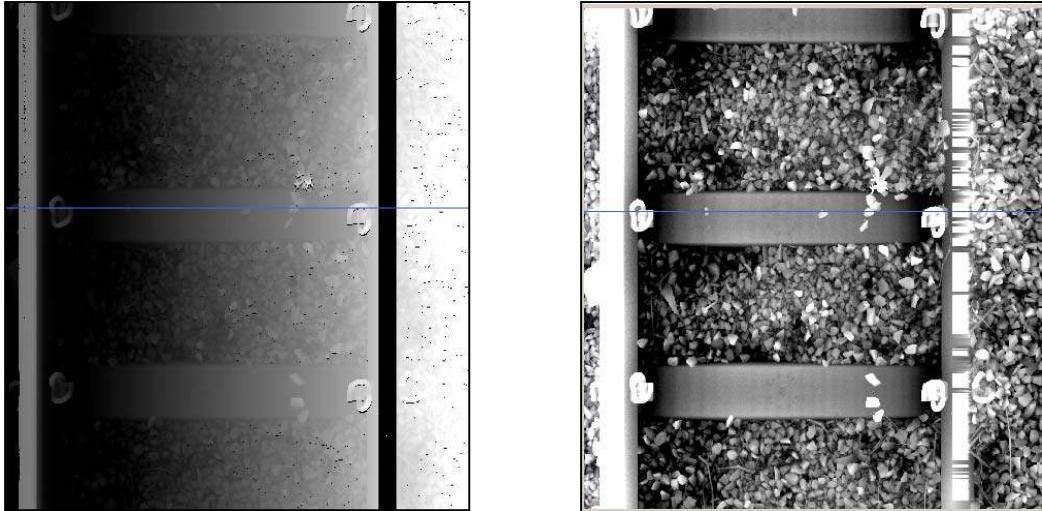
Figure 13 3D laser height image rendered in the original coordinate system (left) and its representation after the coordinate rotation (right).

Second, to obtain a complete view of the entire cross-section of rail tracks, 3d images and data obtained by the two profilers need to be merged. These two profilers, one on the left and another on the right, each acquire part of the cross-section, but they, if combined, cover the full cross-section, with an overlap in the middle. The merging of left and right images requires adjusting the along-track misalignment between the two lasers; this problem was clearly manifested in Figure 14 where the same tie appeared at different locations on the left and right images. The relative along-track shift between the two profilers can be accurately determined by computing the cross-correlation function of two images and identifying the lag that gives the highest cross-correlation.
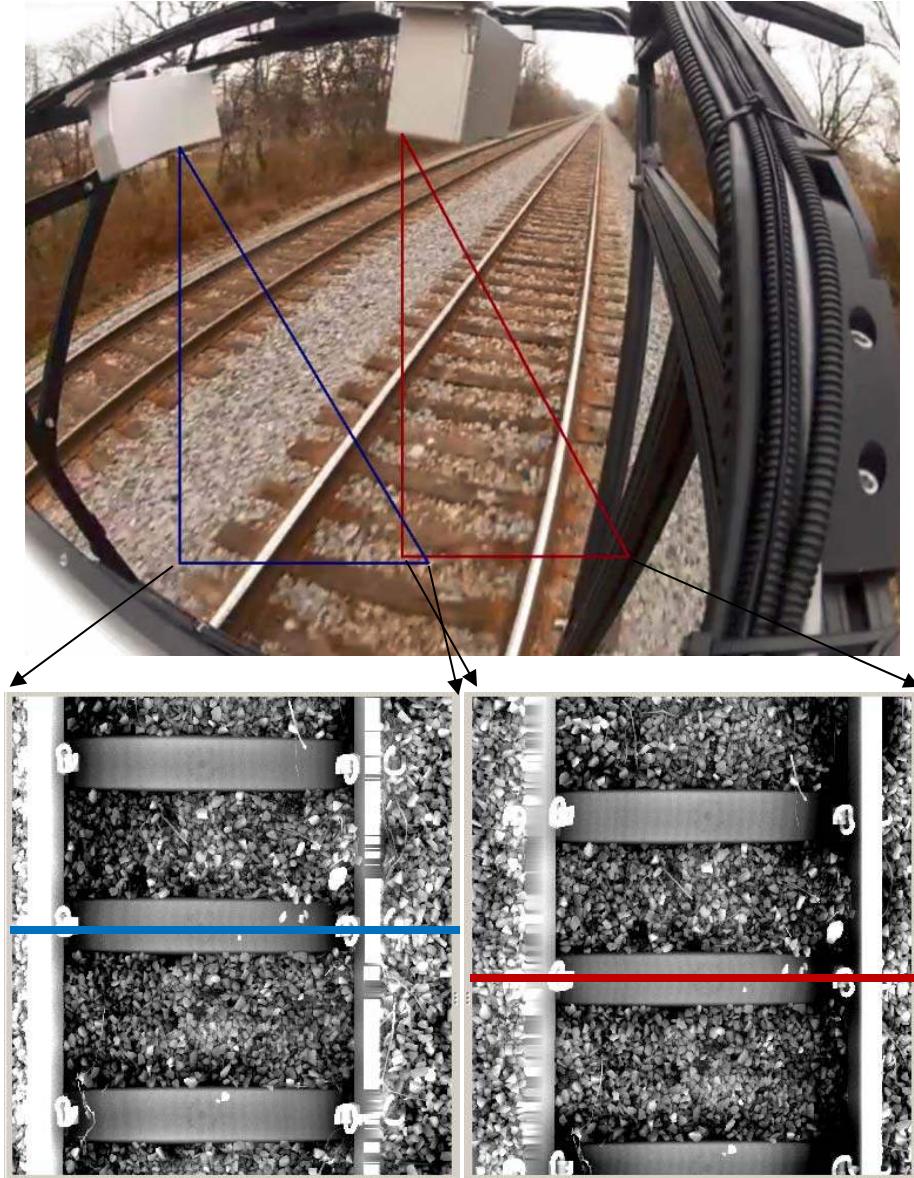
Figure 14 To enhance spatial coverage, two sensors of LCMS on an inspection hi-rail vehicle are deployed to observe the right and left portions of targeted surfaces, yet with a significant overlap. Because of some shift in alignment between the two sensors, there also exists a shift between the left and right 3D images acquired, as indicated the relative locations of the blue and red lines. The shift can be automatically estimated by diagnosing the cross-autocorrelation function between the images and then will be applied to merge the two images into a single one.

Third, individual image sections should be stitched. For convenience of data storage, raw data have been saved as individual sections of ~2 m in length, although the data stream is continuous. If processed individually, features of interest or ties may be split onto two images (Figure 15), potentially introducing artificial edge effects into the feature detection. These edge effects can be easily eliminated by stitching consecutive images together, a process that is done in memory temporarily for convenience of feature detection.
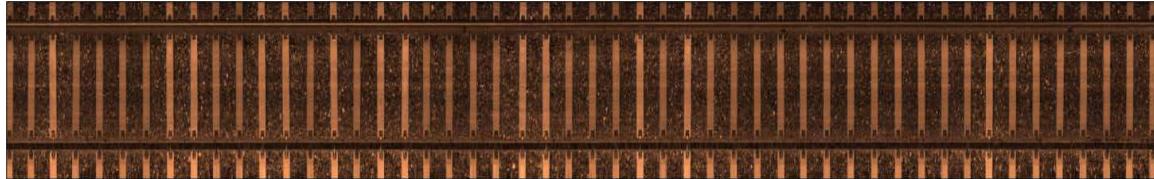
Figure 15 Individual road sections stitched together into a longer road segment for feature detection. Such stitching will help avoid the edge effect.

The 3D shape of the railroad is acquired as the hi-rail inspection vehicle travels along the tracks. The 3D data can be collected at user specified distance intervals (e.g., one image every 1mm or 4 mm) depending on the purpose of the data collection and analysis. This is done through synchronizing the laser line projectors with the vehicle odometer or an encoder installed on the high-rail vehicle.
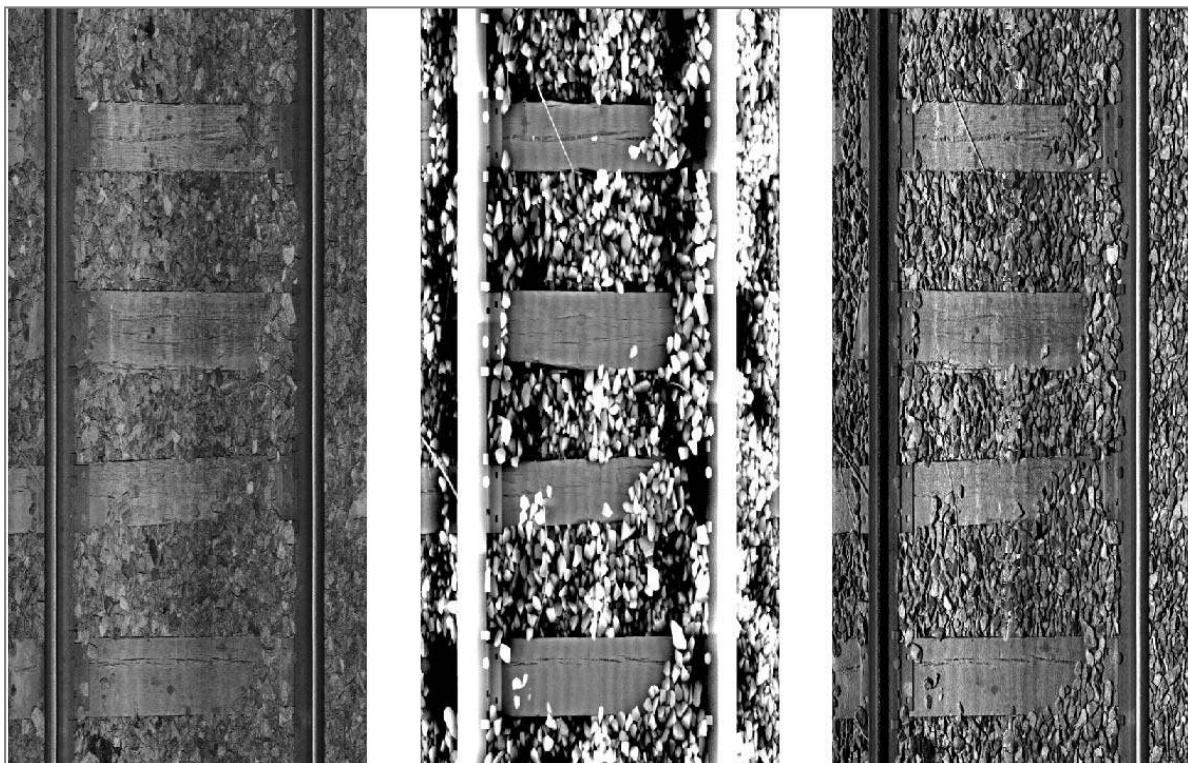
Figure 16 LCMS data type: Intensity (left), range (center), 3D merged (right).

LCMS sensors simultaneously acquire both "Range" (height of each pixel) and "Intensity" (the intensity of the reflected laser light for each pixel) images of the scanned surfaces. By merging the range and intensity data, 3D profiles of the scanned surfaces can be obtained. Figure 16 shows the results of the different types of images (i.e., intensity, range, and merged 3D image) that can be produced from the LCMS data. The range data acquired by the LCMS system measures the distance from the sensor to the surface (e.g., roadway pavement or railroad track) for every sampled point. The center image in Figure 16 is produced by translating range data values into a grayscale image where farther away points (e.g., the railroad ties) are represented by darker pixels and closer points (e.g., the top of the rails) by lighter pixels. The left image is the result from the merging of the right and left images using the method described

15

above.

Next, our feature detection algorithms will be described, with a particular focus on rail gauge estimation, and detection of fasteners and cracks. These algorithms each consist of a set of customized procedures and rules and involve many well-established signal and image processing routines. Examples of the existing routines include least-square fitting, correlative analysis, Fourier transform, mathematical morphology, and image filtering. In particular, the image filters we used include but are not limited to maximum, minimum, median and mean filters of varying sizes and shapes, and edge detectors such as Sobel filter, Canny filter and blob detection [34,35,36,37,38]. All these routines have been well established and documented throughout literature and textbooks; therefore, they will not be detailed here. Rather, the descriptions below will present our general framework on how to combine these existing routines plus some innovative customary procedures for feature detection.

2.5.2. Rail Gauge Estimation

Rail gauge is the measured distance between two rails at a specific point along the track. It is an important inspection data element as variations in gauge are typically associated with other track defects and can cause devastating accidents. For example, missing fasteners can often lead to changes in rail gauge as the rail is free to move and is no longer properly held in place. To address this key need, we developed a new vision-based algorithm for rail gauge estimation. This algorithm utilizes the 3D range data from the LCMS sensor as follows:
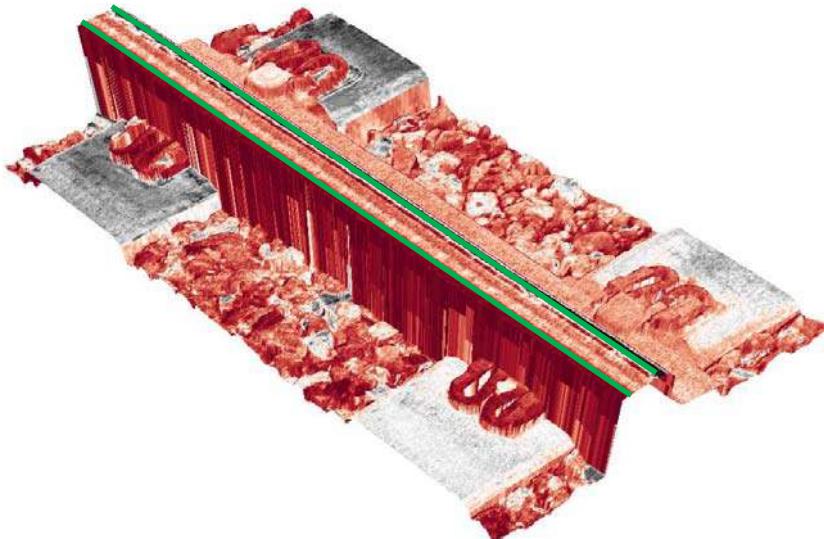
Figure 17 3D Shape of the railroad track acquired by the LCMS.

1. Based on the realistic assumption that the rail head is the nearest object to the imaging system (the LCMS), the algorithm searches for the coordinate of the shallowest point in the generated range data. It is assumed that this point is on the rail head (see Figure 17).

2. A new range image is defined by applying a double threshold operator to the original range image from Step 1. The thresholds are a function of the shallowest point value. As

a result, this refined range image only contains the area of interest for inspection: the rail head.

3. A remapping process is applied to convert the new range image (from Step 2) to a grayscale image (pixel values between 0 and 255).

4. Horizontal and vertical Sobel kernels are applied for edge detection.

5. The Canny method is applied in order to extract the rail head contour.

6. The extracted contour often includes discontinuities. In order to address this issue, a fitting-based technique was developed for closing and refining the rail head contour. In addition, at this step the actual positions of the rail edges are calculated.

7. Finally, we apply a "Blob Coloring module" to detect and recognize (extraction of characteristics) blobs in a given binary image. A blob is a region of a digital image in which all the pixels are similar to each other. The final result of this step is that the edges of the rail in a given image are detected and highlighted.

Figure 18 illustrates the result of this process for a segment of rail; the rail edges are correctly detected and highlighted in red. Once the rail positions are localized, the next step in the inspection process is to identify a new Region Of Interest (ROI) to detect and recognize other features requiring inspection such as tie-plates and fasteners.
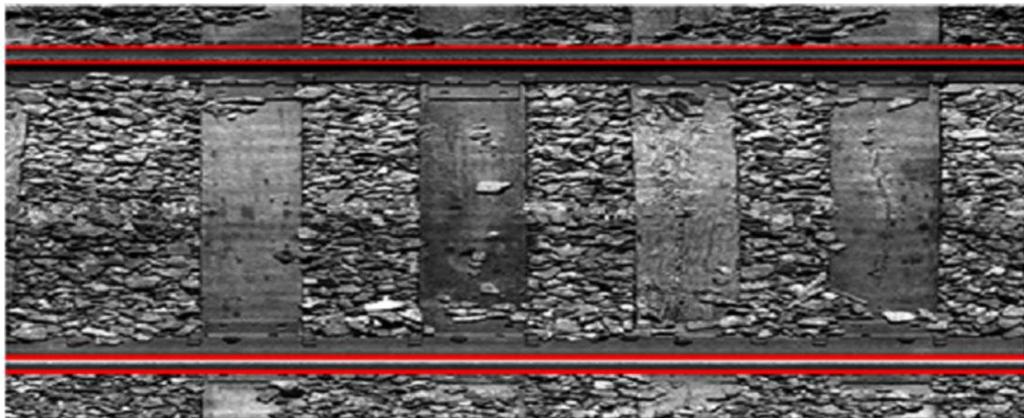
Figure 18 Detection and localization of the rail positions.

2.5.3. Detection of Fasteners

Missing fasteners are a critical flaw to detect during track inspection. To detect missing fasteners, the first step is to detect all of the fasteners in the region of interest. A vision-based approach (as opposed to 3D depth) was developed as follows:

1. First, a new 3D ROI is established based on the detected rail positions from the rail gauge inspection algorithm. The new ROI dimensions are established using the now known dimensions of the rails. Rail height is used to determine the ROI in terms of

17

vertical/depth space and tie plate (see Figure 19) width is used to define the horizontal ROI. As is the case for the rail heads in the rail gauge detection algorithm, a new range image which targets fasteners specifically is produced by this process.

Figure 19 Fasteners to be detected. In the algorithms, these fasteners are of similar shapes and therefore will be detected through a template-matching process.

2. Similar to the rail gauge algorithm, a remapping process is applied in order to convert the range image from Step 1 to a grayscale image (pixel values between 0 and 255)

3. The Canny method is applied in order to extract the contours of fasteners.

4. Two morphological operations (dilation followed by erosion) are applied in order to fill small holes and to eliminate isolated points in the detected fasteners.

5. Finally, a "Blob Coloring module" is applied which detects and recognizes (extraction of characteristics) blobs in a given binary image. In the final output, fasteners are detected and highlighted.

The developed algorithm is applied to some sample data collected from St. Louis. The experimental results, shown in Figure 20 reveal the successful detection of fasteners. More importantly, four missing fasteners (three on the right rail and one on the left rail) are not detected as false positives for the left image in Figure 20. Also, all fasteners except for those three missing ones on the right image of Figure 20 are correctly identified and highlighted.
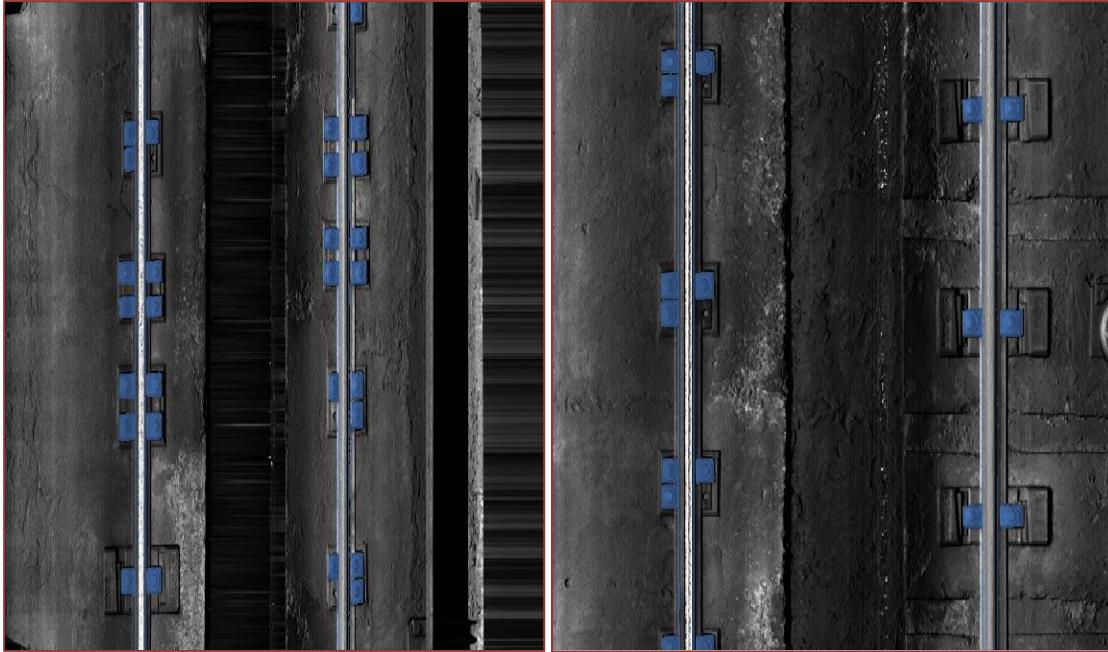
Figure 20 Detection of fasteners in railroad tracks.

2.5.4. Detections of Cracks

Railroad track structure often suffers from the presence of cracks (see Figure 21). Due to the natural variability of cracking patterns, detecting cracks reliably is far more complex than the approach of applying a threshold on a range image as was adopted for rail gauge and fastener detection. This is especially true for hairline cracks as shown in Figure 21. Given the collected 3D LCMS data, the algorithm developed for crack detection involves:

1. The first step is to apply a set of mathematics-based detection operations upon the 3D data set (i.e., the range image) in order to obtain a binary image where the remaining active pixels are potential cracks.

2. The binary image from Step 1 is then filtered to remove many of the false detections which are caused by various design features and general distress on the railroad ties/sleepers which are not cracks.

3. At this point in the processing, most of the remaining pixels can correctly be identified as cracks. However, many of these crack segments need to be joined together in order to avoid multiple detections of the same crack.

4. After the detection process, the next step consists of the characterization of the identified cracks.

5. Finally, the severity level of cracking is determined by evaluating its width (opening). Severity can be further grouped into categories such as low, medium and high by defining width ranges for each level.

19
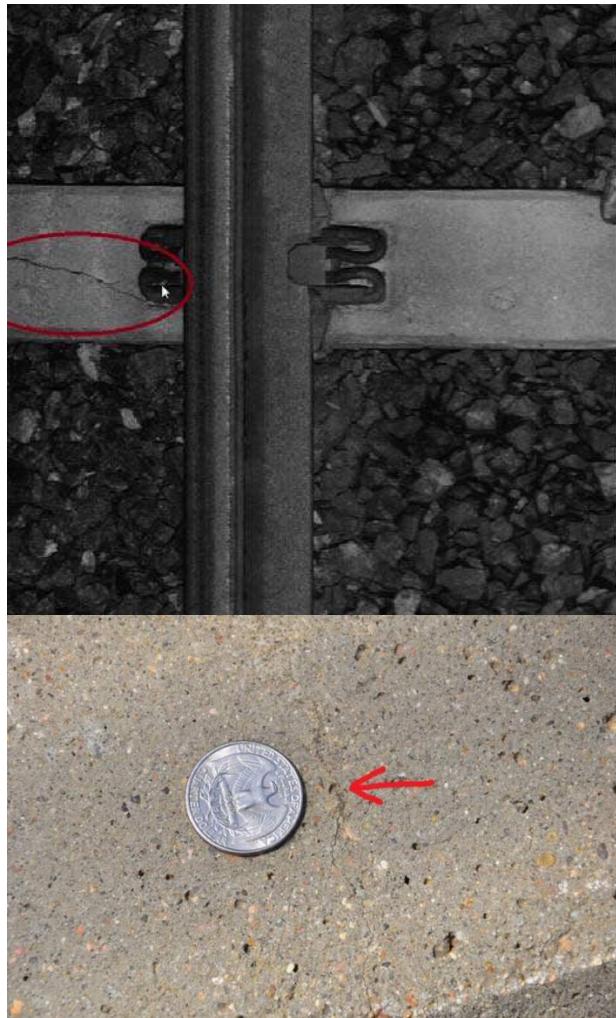
Figure 21 A crack (circled in red) in a concrete-tie (top) and a hairline crack photographed in a concrete tie at Metro St. Louis.

The developed algorithm has been applied to both concrete and wooden ties. For the hairline crack (see Figure 21) in a concrete tie in St. Louis, we were unable to detect it. However, for large cracks in concrete ties and especially wooden ties, the developed algorithm has demonstrated reasonably good performance. Figure 22 shows the range data of a sample wooden tie and the detected cracks are highlighted in red. In fact, those hairline cracks in concrete ties do not pose safety concerns that require immediate attention.
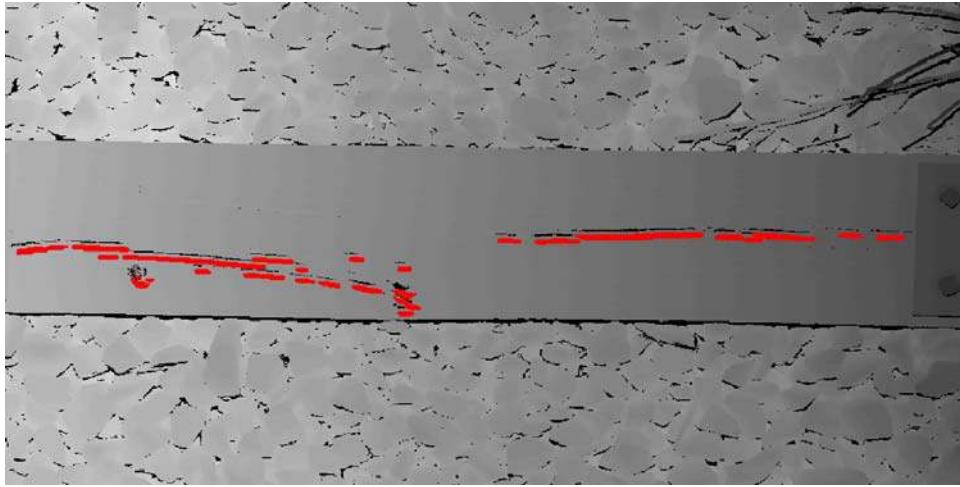
Figure 22 Detection of cracks in wooden cross-ties.

2.6. Conclusions

The reliable and automatic inspection of railroad track structures is critical for both freight and passenger rail safety. Most rail transit agencies rely heavily on manual inspection methods, which are time consuming, labor intensive, and sometimes are unreliable. Although some agencies do employ automatic tools (e.g., ultrasound and geometry car) for track inspections, these tools are usually very expensive and rail transit agencies (especially small- or medium-sized agencies) cannot afford to own them. They can only hire contractors to do the inspection work using these tools on an annual or bi-annual basis, even though they do have the needs to inspect tracks more frequently. The use of computer vision techniques has recently been identified as a promising method for inspection.

This project introduced a new computer vision based framework with the goal of automatic inspection of railroad tracks. The current version of the framework is based on the LCMS system that consists of two high-performance laser profilers. The developed framework is capable of estimating rail gauge, detection of existing/missing fasteners and the identification of cracks in railroad ties/sleepers. The developed framework has been tested based on the rail data collected from Chicago and Metro St. Louis. The result suggests that the framework is able to accurately measure rail gauge and identify missing fasteners and cracks. Also, the developed framework is much more affordable than existing ultrasound and geometry cars. It can be readily mounted on a high-rail vehicle and is a promising alternative to existing manual and automatic inspection methods.

The LCMS system we deployed can collect rail profiles at 1mm horizontal accuracy and 0.5 mm vertical accuracy; therefore, much more information can be extracted from the generated detailed 3D profiles. The current focus is more on rail gauge, fasteners, and ties. In addition to these features, the collected data may also be used to assess the conditions of ballast, track bed slope, and overhead power and signal lines. Moving forward, it is expected that the proposed framework will serve as a platform to develop a complete system for the automatic inspection of railroad tracks.

CHAPTER 3

GPR HARDWARE AND SOFTWARE

The GPR subsystem consists of hardware and software systems. In this chapter, the focus is on the hardware system and the associated control software development. Some radar data processing algorithms are also presented at the end. In the next chapter, a more detailed discussion on the radar data processing algorithms is provided.

3.1. GPR Hardware Development

3.1.1. System Architecture

As shown in Figure 23, the GPR hardware consists of five important functional units: (1) RF (radio frequency) transmitter, (2) ultra-wideband (UWB) antennas, (3) data acquisition comprising high-speed real-time digitizer, high-speed data transmission, and storage unit, (4) multi-core computer, and (5) FPGA digital controller along with a wheel encoder.
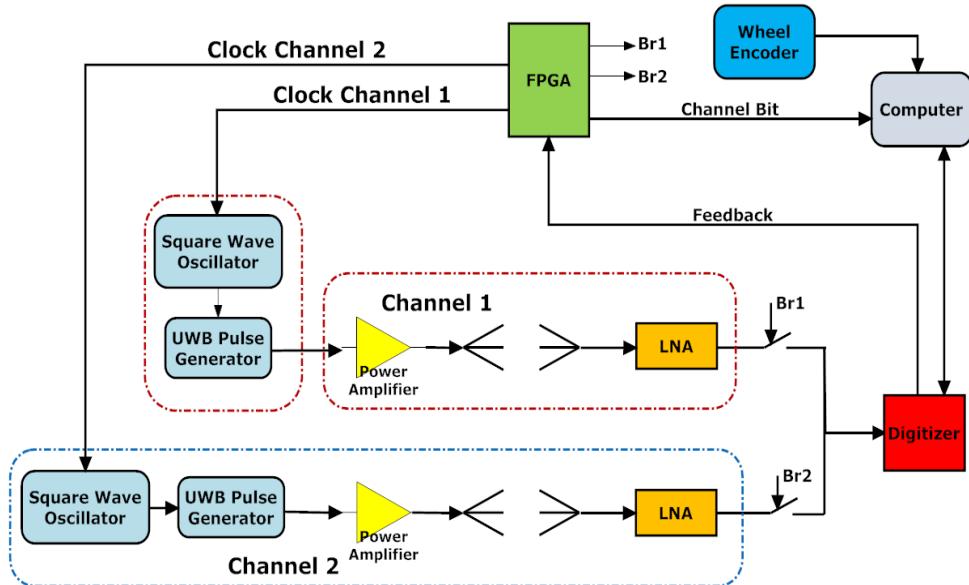
Figure 23 High-speed UWB GPR system.

The RF transmitter comprises of an UWB pulse generator that generates high-amplitude Gaussian pulses with the pulse repetition frequency (PRF) determined by the FPGA. The high-speed digitizer comprises of a 10-bit 8-GHz real-time sampling ADC and a high throughput data transmission unit connected to a computer via PCIe connection. The computer streams the data from the digitizer and tags the data with header details. The wheel encoder measures the scan distance via quadrature pulses generated based on distance travelled. The FPGA receives the pulses from the wheel encoder to perform travel distance triggered scans. The distance information is transmitted to the computer for creating the data headers. The UWB antennas in this GPR system are compact and have good impedance matching over a wide bandwidth for

22

effective signal transmission and reception. Some selected key specifications of this GPR system are summarized in Table 2. In the rest of this section, the designs of selected key functional units are elaborated.

Table 2 UWB air-coupled impulse GPR system specifications

Key Specification	Value
Vertical sampling rate	8 GHz
Vertical sampling window	40 ns
Pulse repetition frequency	0 to 30 kHz
Horizontal resolution	Up to 1 cm at 100 km/h survey speed
Signal bandwidth	600 MHz to 2 GHz (1 GHz central frequency)
Penetrating capability	Up to 1 meter

3.1.2. UWB Pulse Generator

A UWB Gaussian pulse is generated using a microstrip delay line, a Schottky diode, and a step-recovery diode (SRD) connected in series or parallel to the delay line. Pulse generators with SRD connected in series with a microstrip delay line generates pulses with low ringing. But the output amplitude of the pulses is smaller and requires power amplification before being transmitted via the antennas. Pulse generators with shunt-connected SRDs produce higher amplitude pulses that are prone to higher level of ringing. However, by placing the pulse forming network (comprising the delay line and Schottky diode) in the input stage of the SRD, the level of ringing can be reduced drastically. Using a transistor switching circuit and a power transistor operating in saturation region to drive the SRD, negative Gaussian pulses with 23V peak-to-peak amplitude can be generated over a wide range of pulse repetition frequency.

The circuit diagram of the UWB Gaussian pulse generator is presented in Figure 24. It comprises of four parts: (1) Op-amp to amplify the digital clock from an FPGA, (2) Edge triggered timing circuit, (3) BJT driver circuit to generate a negative driving pulse, and (4) Gaussian pulse generator using a SRD.
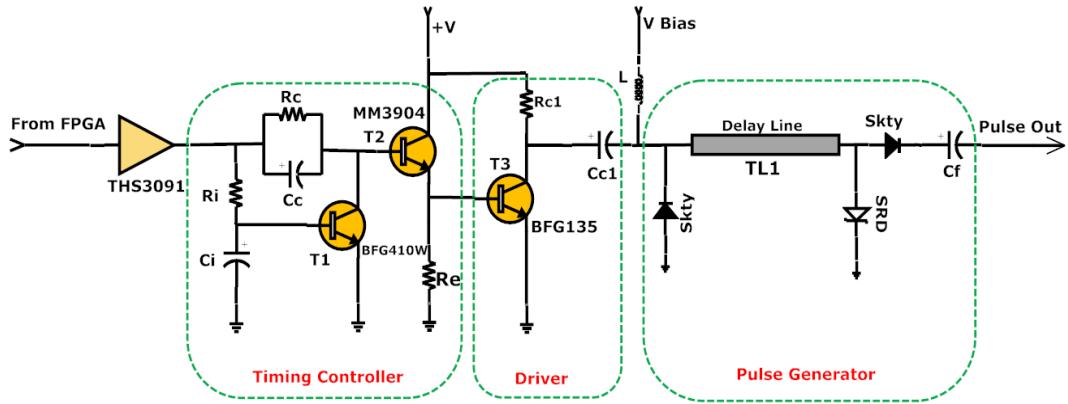
Figure 24 UWB Gaussian pulse generator with shunt-connected SRD.

The first part of the circuit consists of a current feedback operational amplifier, THS3091, to amplify the clock signal from an FPGA. The timing control circuit ensures that irrespective of the PRF of the FPGA clock signal, a negative pulse with defined width is provided to the driver stage. The UWB Gaussian pulse is generated using the combination of a step-recovery diode (SRD), a Schottky diode, and a delay line. In steady state, the Schottky diode is reversely biased and the SRD is forward biased by a small bias voltage Vbias (~1.25 V). When the negative pulse is applied from the driver circuit, the effective voltage (Vbias-Vneg) reversely biases the SRD and forward biases the Schottky diode. Due to the presence of the intrinsic layer, the SRD continues forward conduction for short time duration. When the charges are depleted, the SRD turns off instantaneously and generates a very sharp transition edge in the order of 10s of picoseconds. Also, the SRD offers very high impedance which reflects the signal away from the SRD in two directions, one towards the output and the other one back towards the input stage. Since the Schottky is forward biased and effectively grounded, the signal that travels via the delay line to the Schottky diode is reflected back to the output with opposite polarity. The two signals combine and produce a UWB negative Gaussian pulse.

The width of the UWB pulse is determined by the signal propagation delay across the delay line. The filter capacitor, C_f, transmits only the UWB pulse signal while blocks the driver signal and eliminates the low frequency noise. The waveform measured in a DSO (digital sampling oscilloscope) after 20 dB attenuator (factor of 10) is shown as the pulse generator component in Figure 25. The PCB (printed circuit board) implementation of this GPR pulse generation circuit is shown in Figure 26.
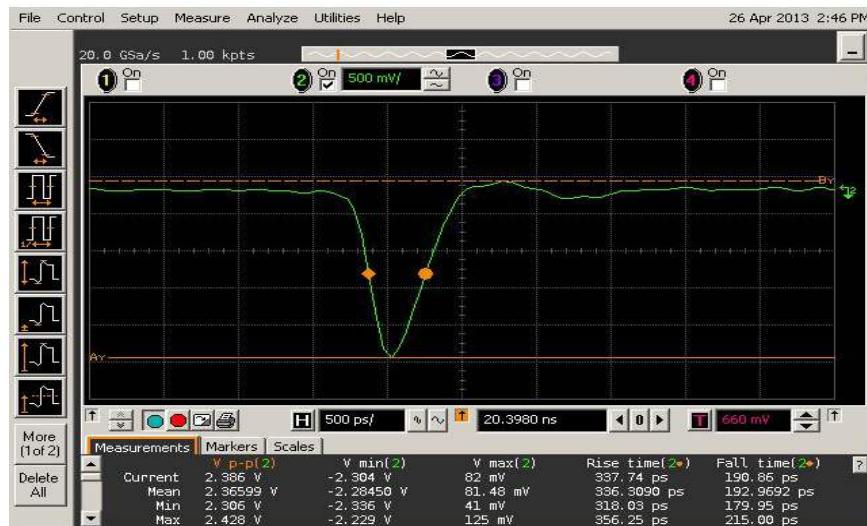
Figure 25 Output waveform of the Gaussian pulse generator measurement upon 20 dB attenuation.
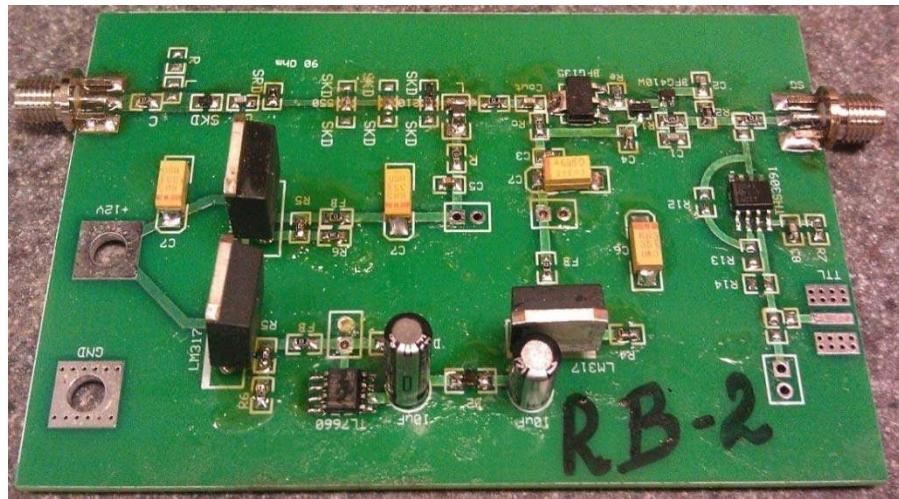
Figure 26 PCB implementation.

3.1.3. UWB Antenna

The UWB antenna in a GPR system needs to have good impedance matching to minimize the internal reflections and to smooth the transition of the circuit impedance. The objective is to maximize power transmission. In this project, the GPR system is equipped with a novel Good Impedance Match Antenna (GIMA).
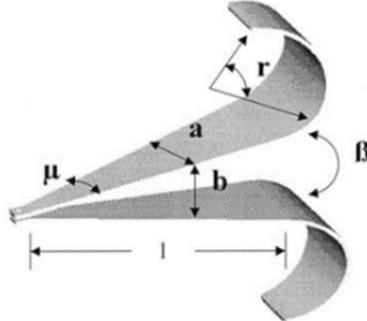
Figure 27 Drawing of UWB horn antenna (where μ = 13', β = 6', r = 150', a = 60 mm, b = 60 mm, and l = 180 mm).

25
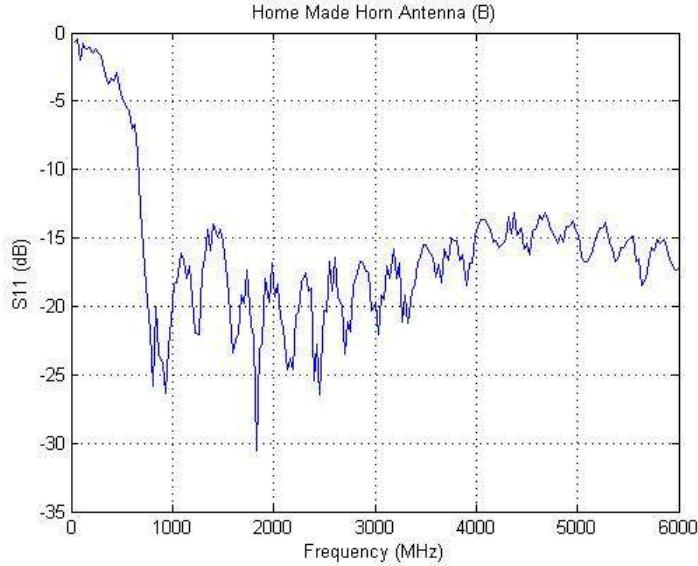
Figure 28 Reflection loss of the UWB horn antenna from 30 MHz to 6 GHz.

As shown in Figure 27, the GIMA is modeled as a waveguide with two metal conductors of varying width and separated by an insulator. The variation in impedance is achieved by smooth variation of the width of the metal plates and the separation gap (β) between them. The rolled edges of the metal plates are to minimize edge diffraction. Figure 27 shows the shape and the key dimensions of the GIMA. The antenna return loss S11 plot is shown in Figure 28. It can be seen that the GIMA return loss is less than -10 dB from 500 MHz to 6 GHz, which attests to its high-quality impedance matching across a wide frequency band. The hardware implementation of the GIMA is shown in Figure 29.

Figure 29 UWB horn antenna with foam supporting the structure.

3.1.4. High Speed Digitizer Configuration

The digitizer used in this system is a 10-bit real-time sampling ADC (analog to digital converter) (Agilent Acqiris U1065A module) with a sampling rate of 8 GSa/s. Most existing commercial GPRs use equivalent sampling technique where only one or a set of data points are acquired from a single pulse. In order to reconstruct the entire pulse, a number of repetitive pulses are used. Therefore, to obtain good horizontal resolution, the pulse repetition frequency

should be in the order of MHz. For our GPR system, by employing the single-shot full-waveform sampling digitizer, high range resolutions can be obtained with low repetition frequencies even at high travel speeds.

At 8 GSa/s sampling rate, the digitizer rapidly accumulates large volumes of data. Such data volumes and rates are difficult to manage with ordinary read and write operations. A solution was proposed to configure the digitizer into the Simultaneous Acquisition and Readout (SAR) mode. In SAR mode, the internal memory of the digitizer is divided into three banks as shown in Figure 30. Each bank is further divided in to 333 segments to hold data samples from one pulse in each segment.

- Acquisition bank 1
- Bank 1 buffered
- Acquisition Bank 2
- Readout bank 1
- Bank 2 buffered
- Acquisition bank 3
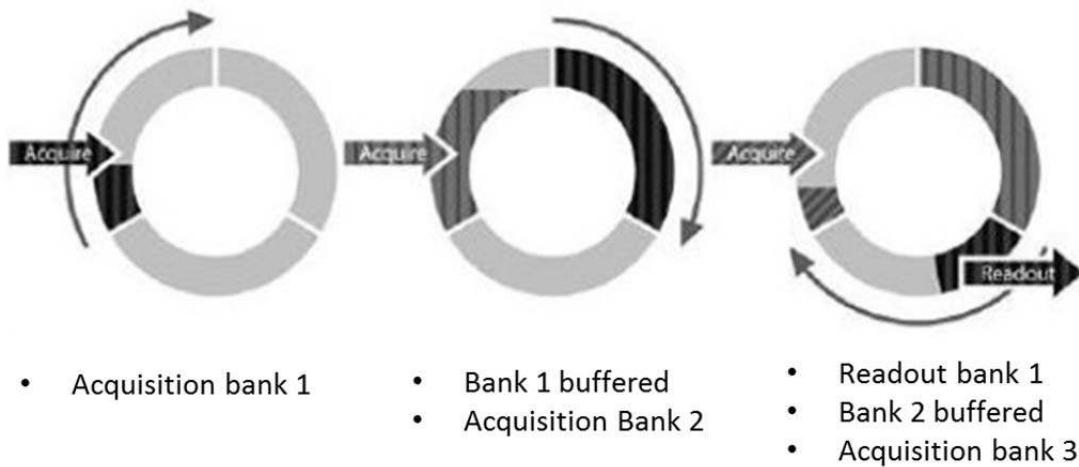
Figure 30 Digitizer configured in SAR mode.

The SAR mode allows simultaneous data writing and reading in the digitizer memory. In other words, the stored data can be read out from one memory bank while the new incoming acquisition data can be written into another available memory bank. The bottleneck in this configuration is the transfer of data from the digitizer to computer memory. To address this issue, a parallel processing algorithm is used to speed up the reading process. A quad-core processor reads data from the digitizer using a high-speed PCI-Express bus, adds header information to the data, and writes the data in computer's hard disk. The header information comprises of the data from the wheel encoder to measure travel distance, system time stamp, and channel tag. Using the SAR configuration, a seamless signal acquisition, data transfer, and storage is successfully achieved for a pulse repetition frequency of up to 40 kHz.

3.2. User Interface and Software Development

3.2.1. Control Program & User Interface

The Graphical User Interface (GUI) of the control program for our GPR system is developed using NI LabVIEW. The GUI is shown in Figure 31. Using this GUI, users can configure the follow GPR system parameters:

- Storage Path – The directory where the .tdms file (GPR data) is saved;
- Delay Time – The trigger delay time is used to adjust the GPR signal to appear in the middle of the sampling window;
- Trigger Number – It specifies for how many trigger pulses received, the digitizer is triggered to collect data once; and
- Pulse Number – In each data collection cycle, the number of GPR traces to be saved in hard disk.

Under normal GPR working condition, the "Error" panel should show no errors if the digitizer is successfully calibrated; the "DEPP EN" panel should display 1 if the FPGA module works properly; and the GPR data is saved as .tdms files in hard disk.
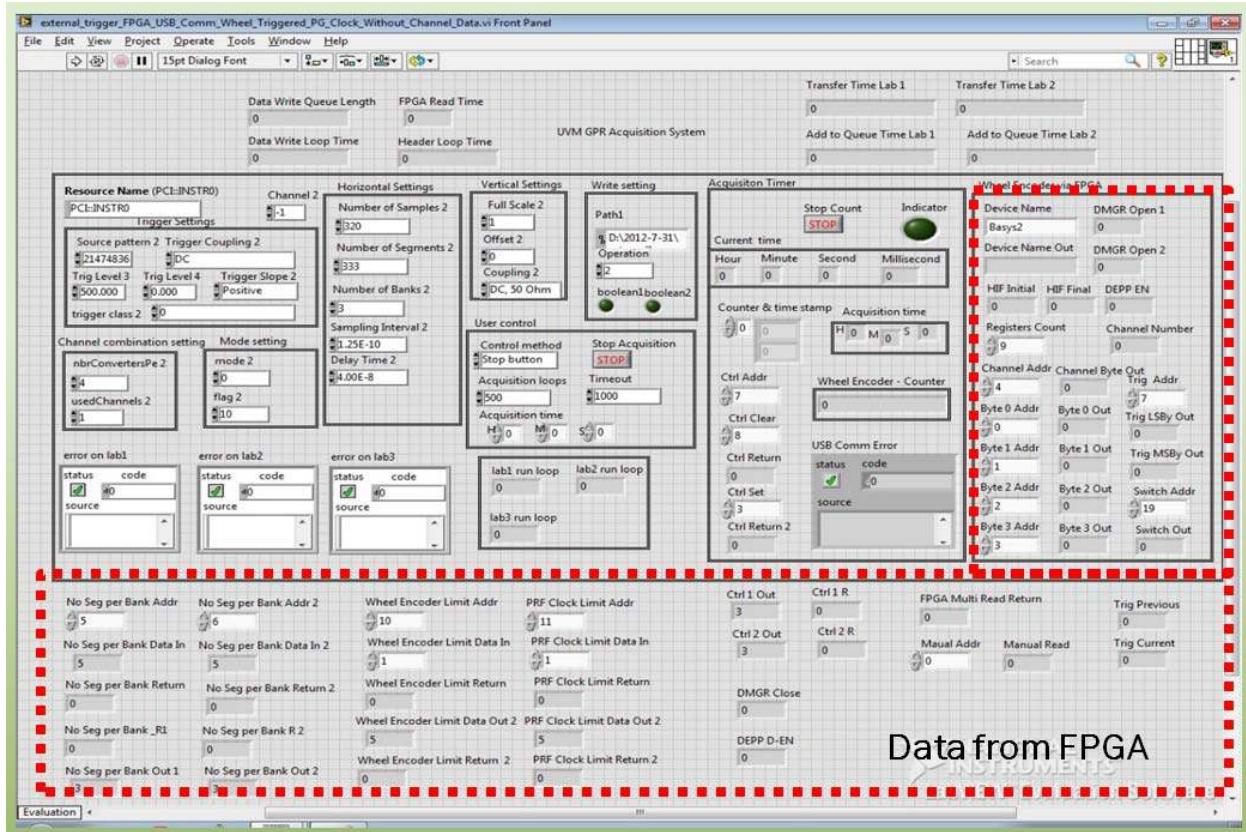
Figure 31 Control program GUI in LabVIEW.

3.2.2. Large Data Stream Display

During GPR scanning, a large amount of data can be accumulated very quickly. The analysis below illustrates the developed GPR data acquisition module specifications.

- Pulse repetition frequency: 30 KHz;
- Each A-scan trace: 320 samples;
- Each sampling data size: 2 bytes (16 bits);

28

- The size of each A-scan trace: 2 bytes × 320 samples/trace = 640 bytes;
- For one hour (60 minutes) of GPR scanning, the total amount of acquired data is: 30 KHz × 640 bytes/trace × 3600 seconds = 69.12 GB.

If the GPR is mounted on a vehicle that travels at 30 mph, it means about 70 GB data is collected every 30 miles. To save such a huge data set in a single file and process it directly is very challenging. Therefore, we choose to divide the large data set into small segments so that signal processing can be conducted on each segment separately. Moreover, dynamic segmentation is introduced that renders users the ability to arbitrarily select data corresponding to any areas of interest. Our radar data is stored in TDMS format, which consists of three parts:

- **Lead In**: It contains basic information about the file (e.g., a tag that identifies the file as TDMS, a version number, and the overall length of the segment);
- **Meta Data**: It contains the names and properties of all objects in this segment, such as index information used to identify the location of raw data segment; and
- **Raw Data**: Scanning data collected before data processing.

The first image is shown exactly the same with the portion of second image
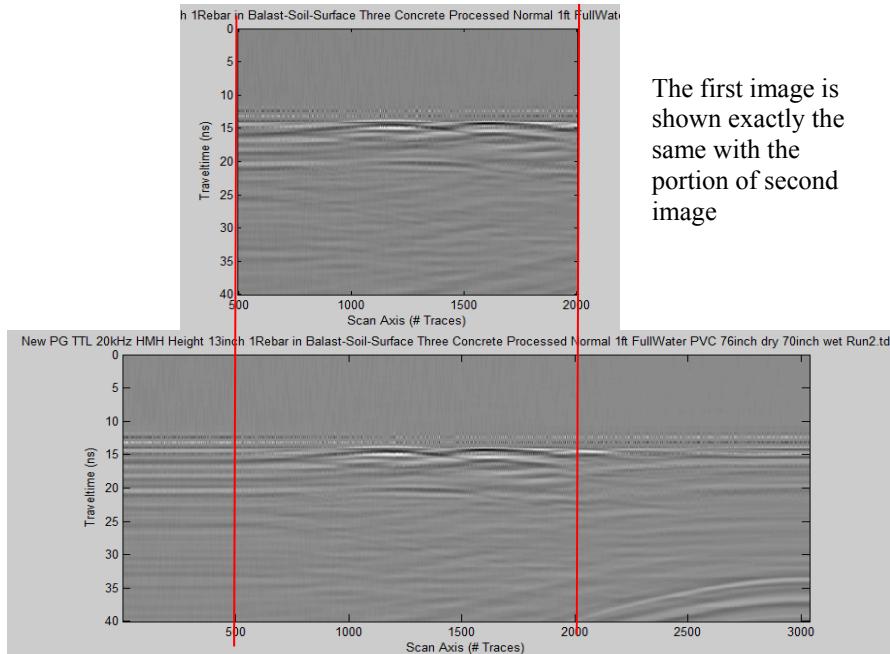
Figure 32 GPR data segmenting example.

To display the collected GPR data, a computer program has been developed to read the "Lead in" and "Meta Data" portions of TDMS files and to store them in computer memory. From these two portions, the computer program identifies the GPR raw data length and the beginning address of the GPR data. A pointer is then used to localize the beginning of GPR data. One can arbitrarily select the segment of GPR data to be processed, and tag the beginning and the end of that data segment as "StartPoint" and "StopPoint", respectively. The selected data segment can be post-processed to obtain B-Scan image. Figure 32 shows one example where a small data segment can be selected from a large data set. For the quad-core computer used in this study, up to 900 data segments (or B-scan images) can be opened simultaneously. Using the

29

"Previous" or "Next" buttons of the developed GUI, we were able to easily switch among B-scan images. For processing large data sets, the following steps should be followed:

- Open and read the TDMS file block by block. At each time, only one block of data is stored in computer memory;
- The smallest unit in one block is 1 segment (333 data traces). The maximum number of segments that can be displayed in one block is 900 due to the memory size limitation of the quad-core computer used in this study;
- Choose the number of segments to be displayed in one block (e.g., n segments). Each time, n segments of data will be displayed in a B-scan image; and
- Use the "Next" and "Previous" buttons (see Figure 35) to display the next and previous n segments of data, respectively.

For demonstration purpose, a 1.125 GB data set was processed. We were unable to open such a large data set using normal data processing methods. Using the above data processing tool, we were able to open the data set and read data block by block. The lead in information of this GPR (.tdms) file is shown in Figure 33.
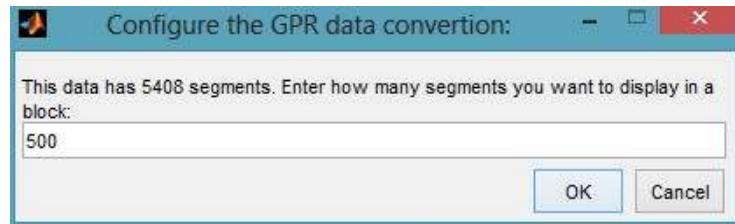
Figure 33 Demonstration - data file information.

This data set has 5,408 segments which are saved in TDMS file format. As shown in Figure 33, the program is configured to open 500 segments in each block, which results in 11 blocks. Figure 34 through Figure 36 show some demonstration screenshots for this data set.
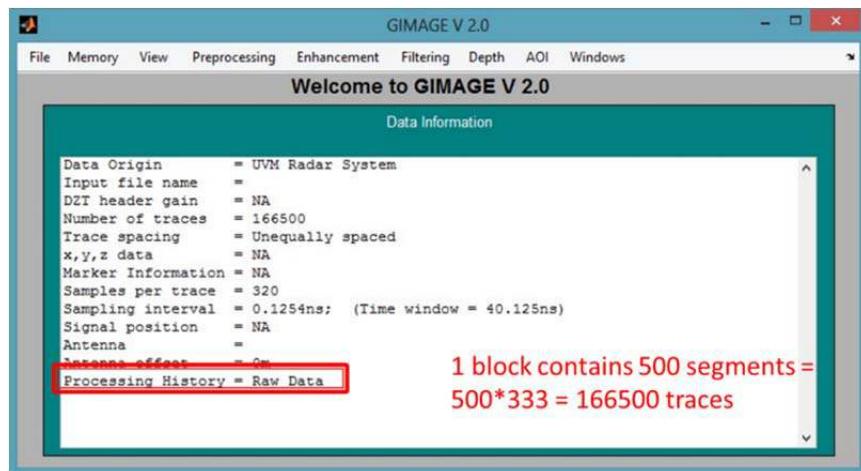
Figure 34 Demonstration - data information.

Since this is the second block, now both 'Previous' and 'Next' buttons are enabled.
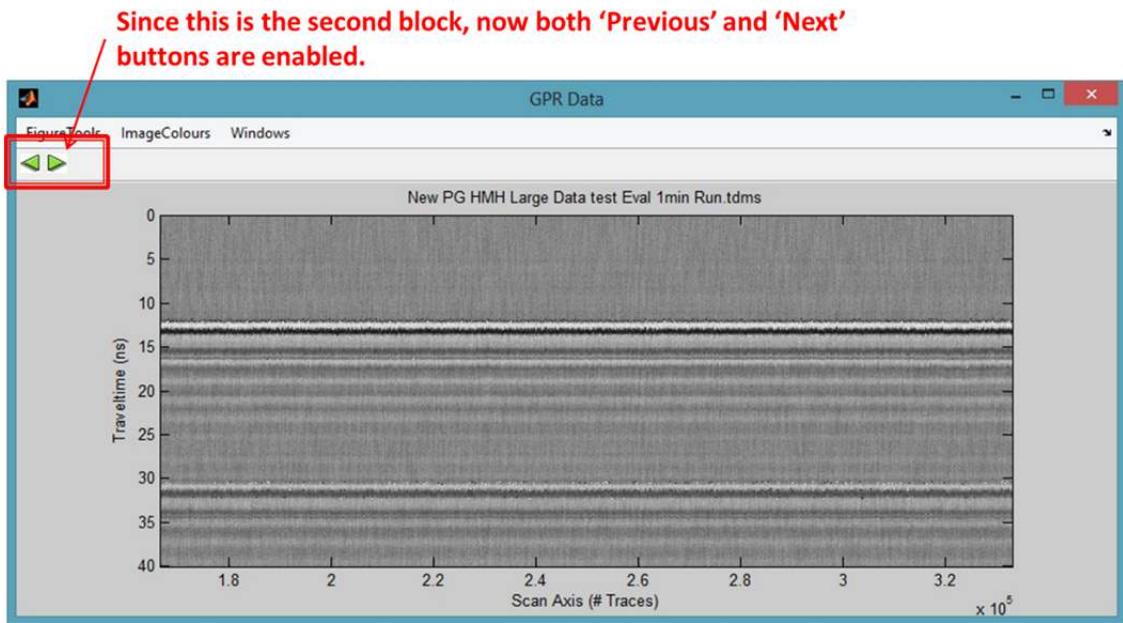
The X-axis indicates the location of this block in the whole data.

Figure 35 Demonstration - B-scan block.

Since this is the last block, now the 'Next' button is disabled
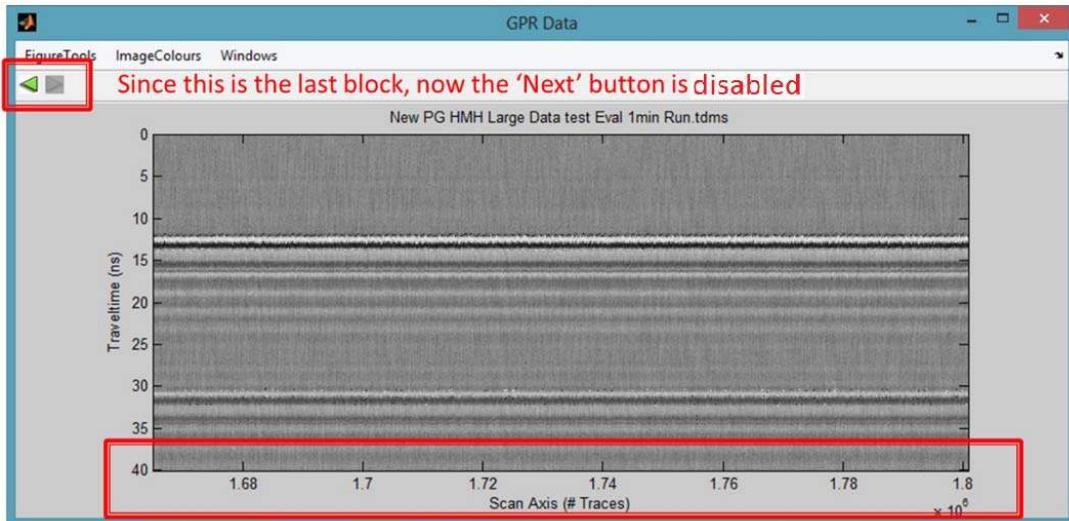
The last block only has 408 segments, so the number of data traces are 408*333=135864

Figure 36 Demonstration - last block.

3.2.3. Dual Channel Control Configuration

The developed software system has two parallel scan channels that operate in "ping-pong" mode. Since both channels share the same data acquisition unit, the high speed ADC has to collect data from the two channels alternatively. A challenging issue here is when the collected data is continuously streamed into the ADC and stored in hard disk, how to identify the

source channel of each data segment precisely? To solve this problem, the frame synchronization approach used in communication systems is adopted. The proposed approach adds a synchronization head to the data to distinguish the two channels. At the beginning of each scan, one channel is set to have oscillating signal while the other is kept silent for certain clock cycles. During radar scan process, such a synchronization operation is performed periodically (e.g., once every 10 minutes) to prevent radar from losing synchronization due to system metastability or other interference effects. Figure 37 shows an example of separating radar data from different channels under time trigger mode. Figure 38 shows an example of separating radar data from dual channels under wheel encoder trigger mode.

Figure 37 Dual channels radar data under time trigger mode.

The whole GPR data containing Channel 1 and Channel 2:
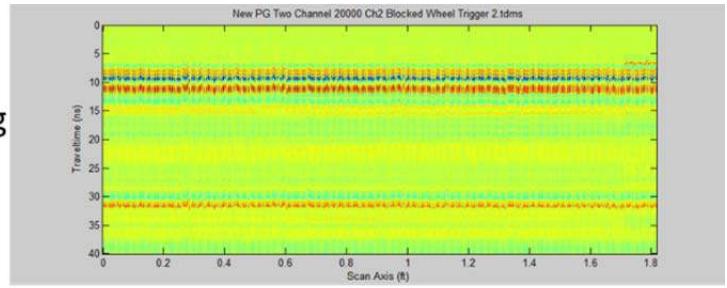
Channel 1 data:
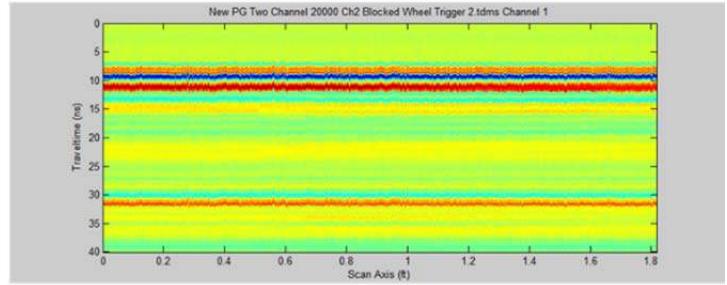
Channel 2 data with self-check blank data:
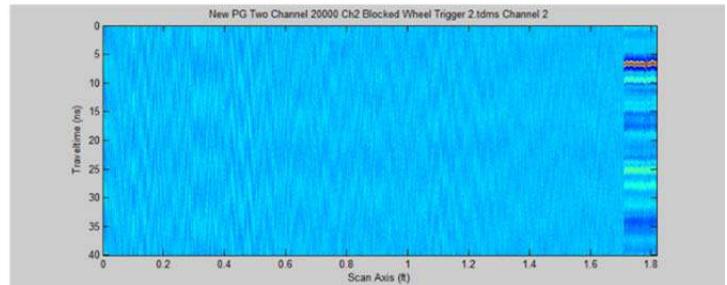
Figure 38 Dual channels radar data under wheel encoder trigger mode.

3.3. Signal Processing Algorithms

3.3.1. Railroad GPR B-Scan Image Enhancement

In addition to the user interface and the control program, a five-step signal processing approach has been developed to improve the B-scan image quality. The flowchart of this signal processing approach is shown in Figure 39. The first step is to remove random noise by performing averaging operation; the second step is image resolution improvement. In B-scan images, most object features have small dimensions and may be fuzzy to be recognized. To resolve this issue, a Bicubic interpolation algorithm is developed to leverage data resolution so as to improve image quality. The Bicubic interpolation considers not only the pixel value of the interpolation point but also the derivatives of 16 pixels (4×4) surrounding it. Compared to other interpolation algorithms such as bilinear and nearest neighbor, the Bicubic interpolation can generate much smoother results; the third signal processing step is the adaptive gain enhancement. Due to the signal attenuation during transmission, the reflection signals from deep layers are often blurred by the strong reflections from shallow layers. In this study, an adaptive signal gain manipulation method is applied to compensate deep layer reflection signal attenuation; Hilbert Transform is often used in communication demodulation, which can

33

effectively extract the modulated baseband signal from the modulation. In our GPR image processing, Hilbert Transform is implemented to extract the pulse signal envelope based on the signal power characteristics; the fifth step is background removal. In B-scan images, the strong air-ground surface reflection and other clutters can blur subsurface object signals. The background removal step is to extract subsurface object signals from the background images. For ballast contamination assessment, an analytic method using Hilbert Transform is also applied to extract reflection signal power information.

- **System and Environmental Noise Removal**
  - ✓ **Ensemble Averaging**
- **Image Resolution Improvement**
  - ✓ **Bicubic Interpolation Algorithm**
- **Signal Attenuation Compensation**
  - ✓ **Adaptive Gain Adjustment**
- **Signal Envelope Extracting**
  - ✓ **Hilbert Transform**
- **Background Removal**
  - ✓ **Average Subtracting Filter**

Figure 39 B-scan image enhancement flowchart.

To evaluate the performance of the five-step signal processing approach, experiments were conducted with two different setups. The first setup was designed for mimicking a railroad segment with wood ties and subsurface pipelines. The second setup was for inspecting railroad segments containing fouled ballast.
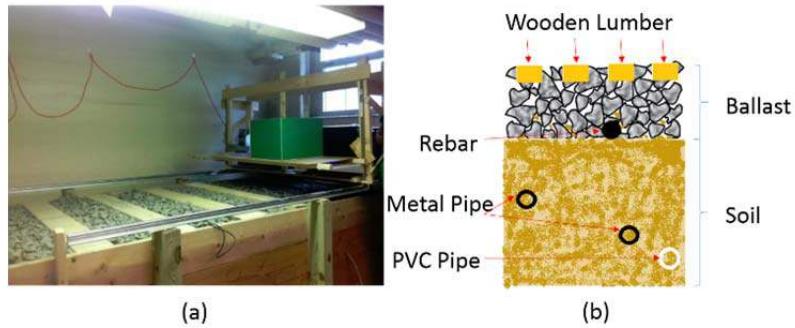
Figure 40 Setup with wood ties and underground pipelines: (a) lab setup, and (b) subsurface features.

Figure 40(a) shows the test platform used to emulate a segment of railroad. The wood ties (lumbers) are spaced 1 foot apart. Figure 40(b) illustrates how the subsurface structures are designed: (1) The ballast layer right above the soil is 8 inches thick; (2) One rebar is buried at the ballast-soil interface. The rebar diameter is 2 inches; and (3) Two metal pipes (diameter = 2 inches) and one PVC pipe (diameter = 4 inches) are buried inside the soil layer. Their burying depths are 18 inches, 24 inches, and 26 inches from the surface of the ballast layer, respectively.
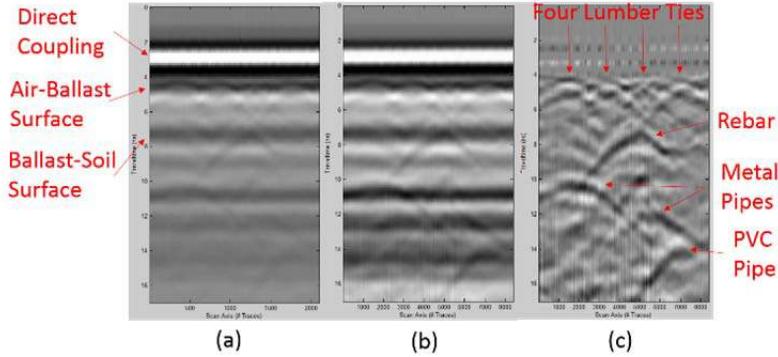
Figure 41 B-scan results for wood ties and pipelines: (a) raw B-scan, (b) interpolation + adaptive gain enhancement, and (c) background removal.

Figure 41(a) is the raw B-scan image after noise removal. Figure 41(b) is the image after interpolation and adaptive gain adjustment. Figure 41(c) is the image obtained from Figure 41(b) after background removal. It can be seen that after removing background clutters, the features for surface and subsurface objects become much more pronounced. As shown in Figure 41(c), these features include four hyperbola curves for the wood ties, one hyperbola for the rebar, two hyperbolas for the metal pipes, and one hyperbola for the PVC pipe.
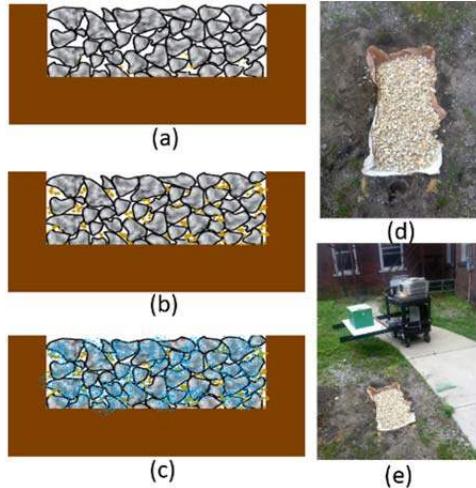
Figure 42 Ballast condition assessment test configuration: (a) dry and clean ballast, (b) dry and fouled ballast, (c) wet and fouled ballast, (d) ballast buried in ground, and (e) GPR system.

To further test the GPR system and the five-step signal processing approach, the developed GPR system was used for assessing ballast conditions. As shown in Figure 42, three different test scenarios were considered, including dry and clean ballast, dry and fouled ballast, and wet and fouled ballast. For the dry and clean ballast scenario, a 2 feet long, 1 foot wide, and 3 inches deep hole was filled up with dry and clean ballast. For the dry and fouled ballast scenario, soil and sand was mixed with the clean ballast. For the wet and fouled ballast scenario, approximately one gallon of water was evenly spread over the fouled ballast surface.

35

For the dry and clean ballast scenario, large air voids exist in the ballast, causing strong signal scattering effect. For the dry and fouled ballast scenario, since the air voids are filled by a mixture of soil and sand, the scattering effect is relatively low. While for the wet and fouled ballast scenario, the moisture content further reduces the strength of the reflected radar signal, which results in the weakest scattering effect and the lowest reflection signal amplitude.
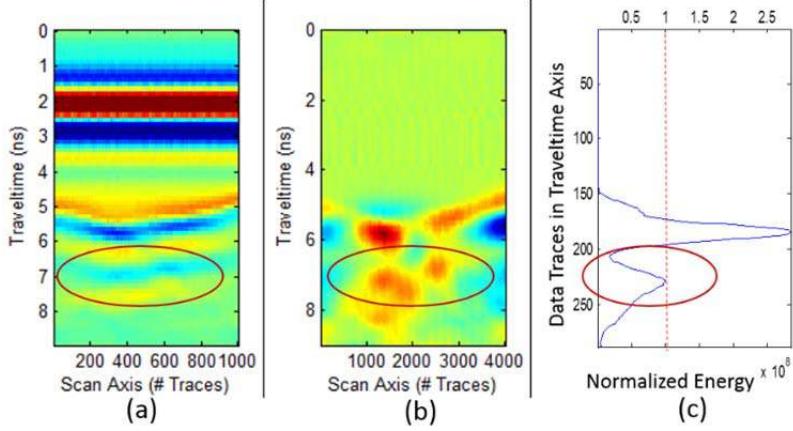
Figure 43 B-scan images for dry and clean ballast: (a) raw B-scan, (b) processed B-scan, and (c) normalized energy map.
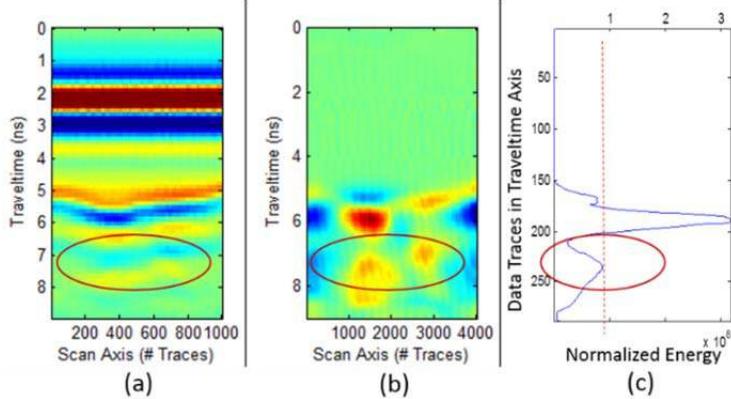
Figure 44 B-scan images for dry and fouled ballast: (a) raw B-scan, (b) processed B-scan, and (c) normalized energy map.
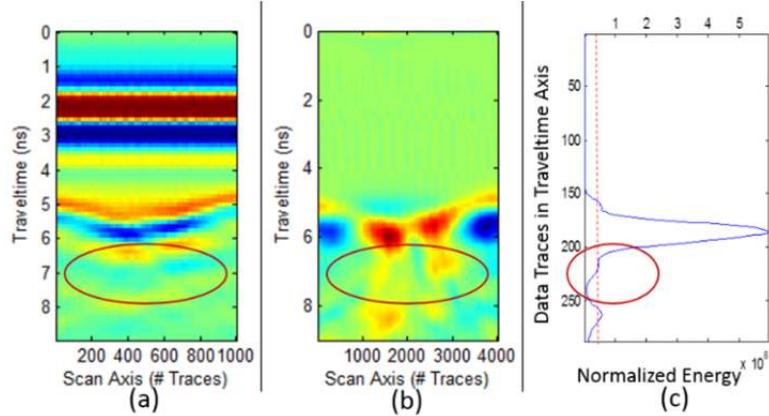
Figure 45 B-scan images for wet and fouled ballast: (a) raw B-scan, (b) processed B-scan, and (c) normalized energy map.

Figure 43(a), Figure 44(a), and Figure 45(a) show the de-noised (i.e., the first step in Figure 39) B-scan images for dry and clean ballast, dry and fouled ballast, and wet and fouled ballast, respectively. In these images, the direct coupling signal between the transmitter and the receiver antennas locates between 1 ns and 4 ns. For the ballast, its top surface's reflection signal locates at 5 ns; the reflection signal of ballast locates between 6 ns and 8 ns; and its bottom surface's reflection is at 8 ns. A comparison of the three images does show some differences. However, they are insignificant and can be easily overlooked. Therefore, the remaining four steps of the five-step signal processing method were applied to these images.

After applying the remaining four steps, the power distribution features of reflection signals are extracted and plotted in Figure 43(b), Figure 44(b), and Figure 45(b) for the three scenarios. Their normalized power maps are shown in Figure 43(c), Figure 44(c), and Figure 45(c), respectively. For the dry and clean ballast, its normalized energy is close to 1, while the corresponding value is 0.9 for the dry and fouled ballast and 0.5 for the wet and fouled ballast. As the above analysis suggests, these quantitative power parameters can effectively characterize ballast conditions.

3.3.2. Interest Region Detection Using 2D Entropy

About 300 GB GPR data was collected during the field tests. Given such a large data set, it would be difficult and time-consuming to review the B-scan images manually. Automating the data analysis becomes both important and necessary. Since the developed GPR system is mainly for subsurface defect inspection, an A-scan decomposition method was first utilized to automatically separate cross-ties from ballast layer before any further data analyses. After these two layers are separated, different algorithms can be developed to analyze them separately rather than applying the same algorithm to both of them. For the subsurface layers, a 2D entropy method was developed to automatically identify problematic or suspicious features for further analysis. Before the 2D entropy analysis, four pre-processing steps are applied:

- Step 1: Stacking every 100 traces waveforms;
- Step 2: Background removal using subtracting average;

- Step 3: Low pass filter with 2 GHz cutoff frequency to remove high frequency noise; and
- Step 4: Stacking every 10 traces waveforms.

After the pre-processing steps, signal decomposition based on cross correlation was applied to mark the cross-ties. Using the direct coupling pulse as the reference signal, an A-scan waveform can be decomposed into combinations of this reference signal with varying amplitudes and time delays. The decomposed signal patterns represent the reflection signals from scatters (e.g., underground objects and layer interfaces). In our B-scan images, the first echo was the antenna direct coupling signal, and the second echo was the reflection signal from cross-ties. Based on the signal decomposition result, the cross-ties were marked out and separated from the ballast and soil layers underneath it.

Following the signal decomposition, the 2D entropy method was utilized to characterize the singularity and similarity of subsurface structures. High entropy values indicate high similarity, while low entropy values mean high singularity. Based on this principle and the entropy analysis results, heterogeneous regions in B-scan images were marked out automatically. This 2D entropy method has been applied to the data collected both in the UMass Lowell radar lab and from Metro St. Louis.

Figure 46 Test configuration in UMass Lowell radar lab.

Figure 46 shows the test platform containing both dry and wet ballast in the UMass Lowell radar lab. The region inside the red circle contains wet ballast while the remaining region contains dry ballast.

New PG HMH Fouled Ballast Soil 4ft 4in ThickSoil 2ft 4in Moisture soilregion 2gal 0min 1.tdms
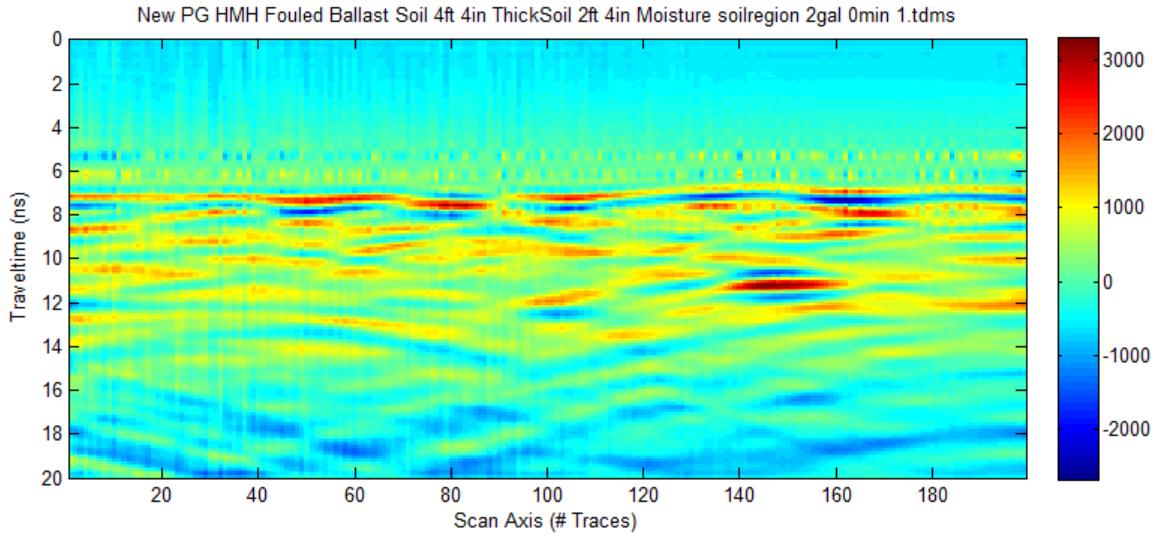
Figure 47 Pre-processed B-scan image based on UMass Lowell lab data.

New PG HMH Fouled Ballast Soil 4ft 4in ThickSoil 2ft 4in Moisture soilregion 2gal 0min 1.tdms
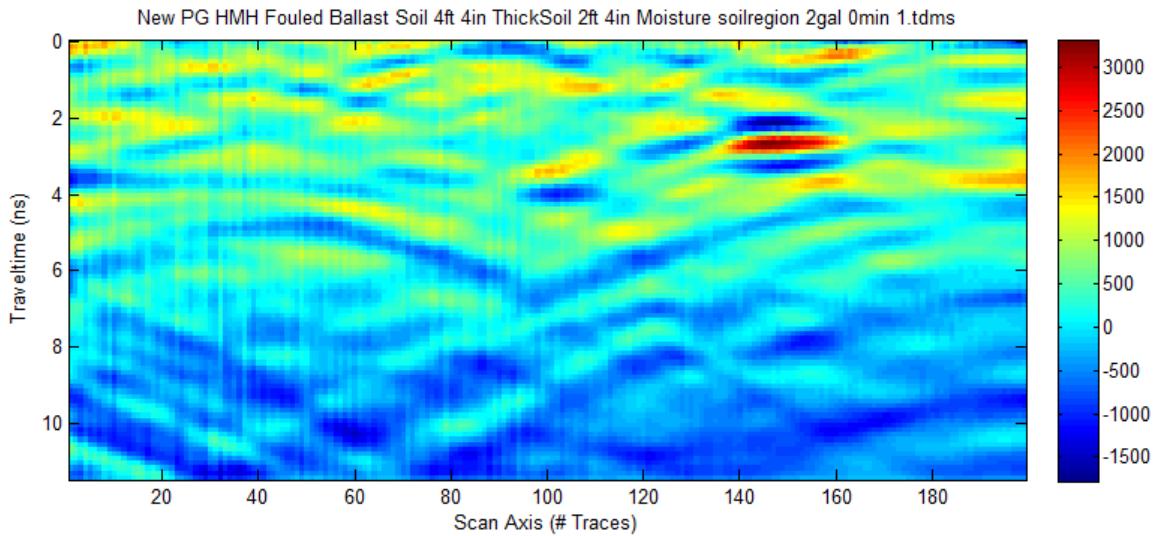
Figure 48 Subsurface layers.

Figure 47 shows the pre-processed B-scan image for the lab test configuration shown in Figure 46. Based on the A-scan decomposition result, the subsurface layers are located and shown in Figure 48.
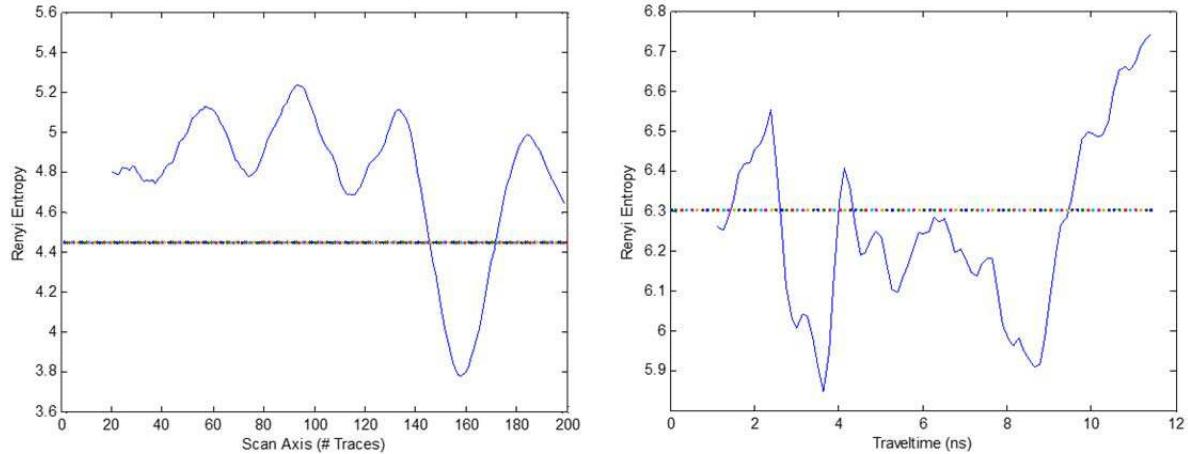
Figure 49 2D entropy analysis result.

The 2D entropy was calculated for the subsurface layers and the result is shown in Figure 49. The left figure shows the entropy values in the horizontal direction while the right one shows the entropy values in vertical direction. The two dashed lines in Figure 49 illustrates the thresholds. The thresholds were determined by the OTSU method. The regions below the threshold lines are considered singular.
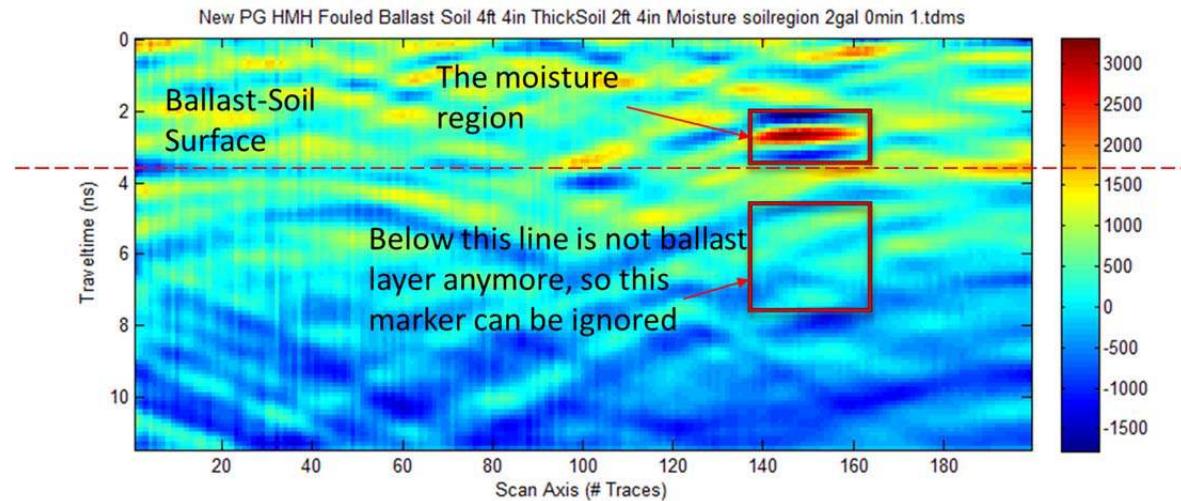
Figure 50 Wet region marked by red rectangles.

Figure 50 shows the 2D entropy analysis result for the UMass Lowell radar lab data, which is for wet and fouled ballast. The red horizontal line is the ballast-soil interface. The region below this line is for soil. The focus in this experiment is the ballast layer so the singular region in the soil layer is not considered. As Figure 50 suggests, the 2D entropy method was able to successfully identify the wet ballast region and automatically mark it by a red rectangle. The location of the red rectangle matches the actual wet ballast region in the lab setup.

UVM GPR system mounted on
the railroad vehicle

Our team working on the
railroad experiment

Figure 51 Field tests.

Figure 51 shows the GPR field test in St. Louis. An approximately 8-meter section from the St. Louis MetroLink's Cross County line was selected to demonstrate the developed GPR system and the signal processing algorithms. This 8-meter section consists of 300 segments and each segment consists of 333 data traces.
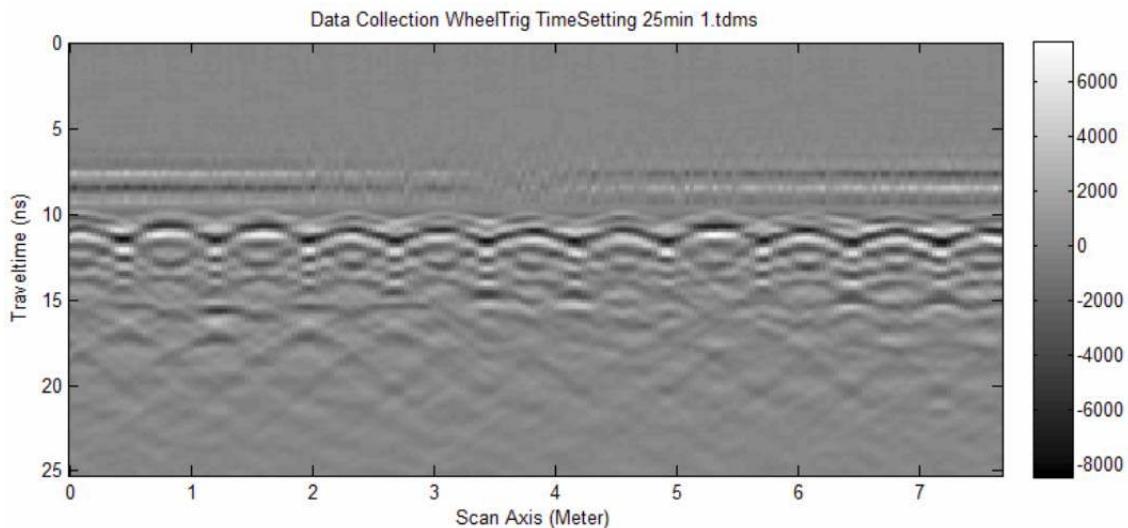
Figure 52 Processed B-scan image.
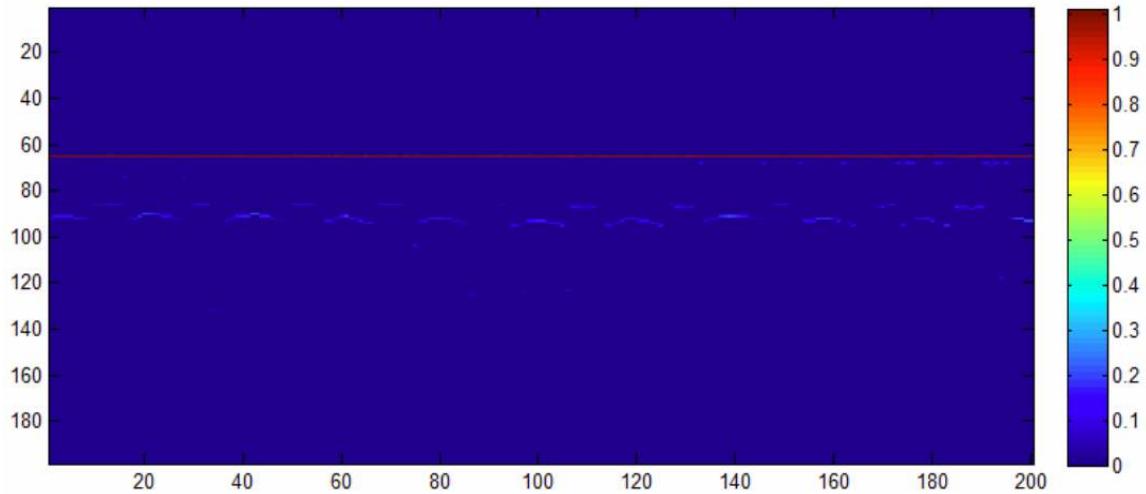
Figure 53 Marked cross-ties.

The pre-processed B-scan image for this 8-meter section is shown in Figure 52. After applying the signal decomposition method, the cross-ties were singled out and the result is shown in Figure 53. The red line is the marker for antenna direct coupling signal. The light blue curves are the markers for cross-tie reflection signals.
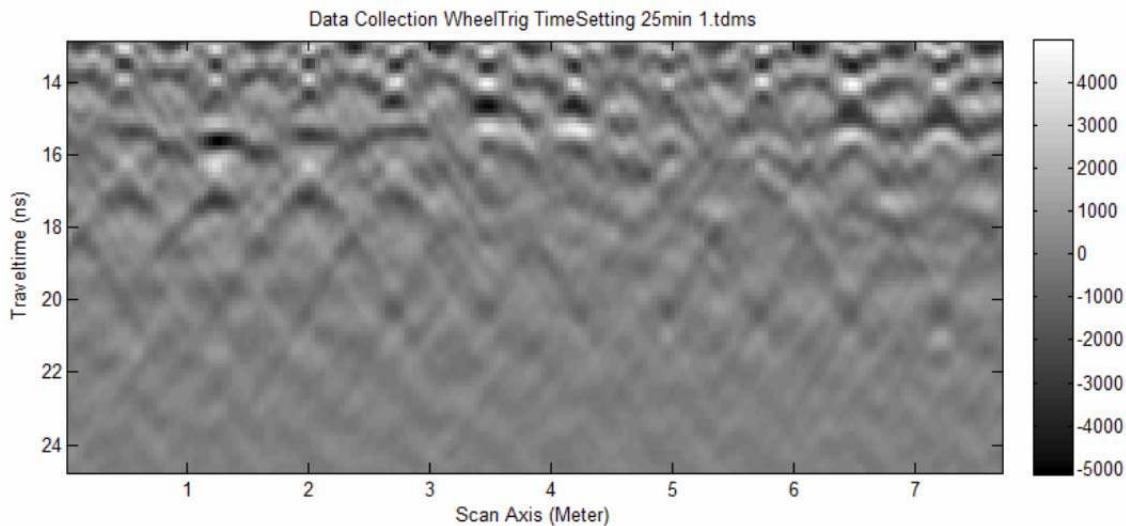
Figure 54 Localized subsurface layer.
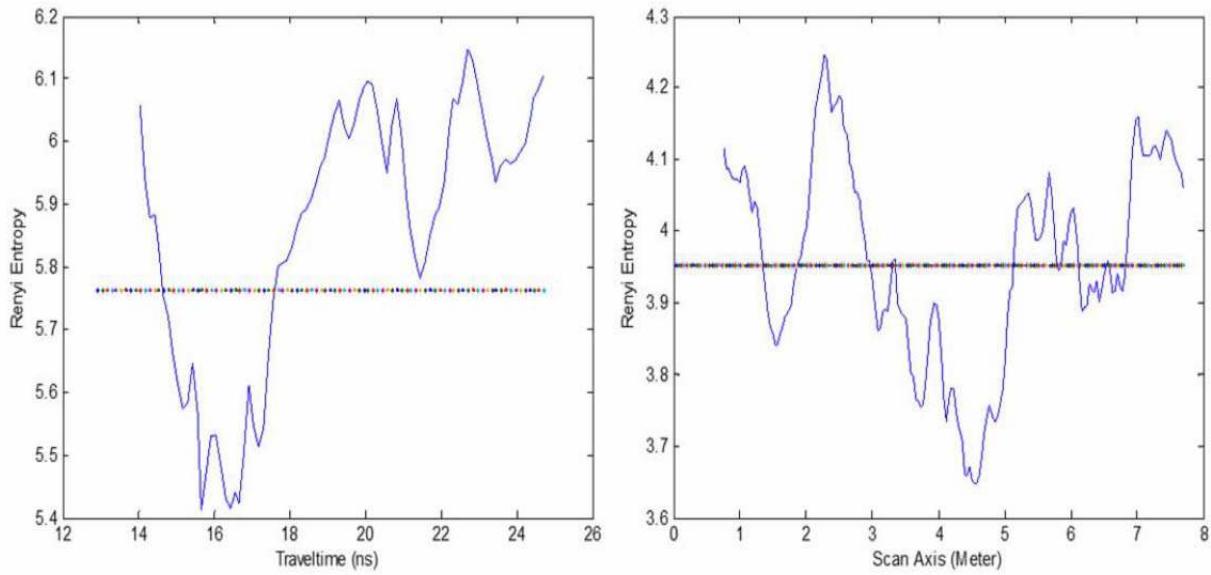
Figure 55 2D entropy analysis results.

Based on the marked cross-ties, the subsurface ballast layer was localized to be from 13 ns to 24 ns as shown in Figure 54. The 2D entropy was calculated for the ballast layer as shown in Figure 55. Based on the 2D entropy result and the threshold values determined by the OTSU method, three suspicious subsurface ballast regions were marked by red rectangles in Figure 56. These three regions in this section may contain abnormal features (e.g., pipelines or fouled ballast) that deserve further investigation.
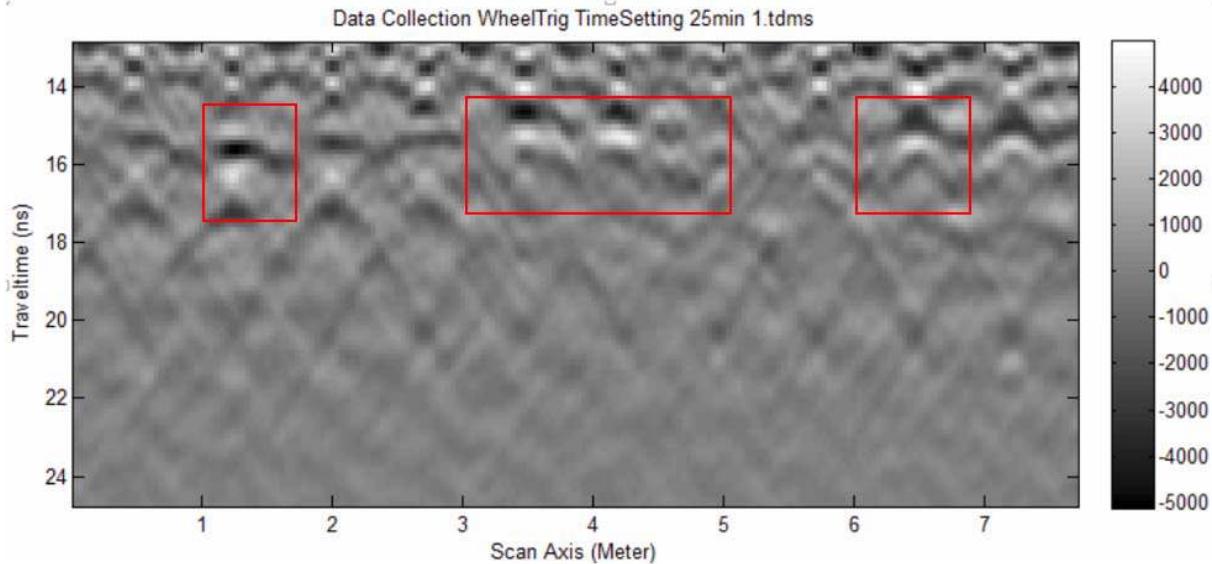
Figure 56 Processed radar image with suspicious areas marked by red rectangles.

As shown in Figure 56, the developed signal processing algorithms were able to

automatically identify abnormal regions from the field data. Unfortunately, the research team was not allowed to dig in the field (This was why a wooden box was constructed in the UMass Lowell radar laboratory) due to safety reasons. Otherwise, it would be very interesting to find out what was there in the field and use it to validate the developed signal processing algorithms.

3.3.3. Batch Processing

A Matlab code frame has been implemented to run the abovementioned algorithms. This code frame has been applied to process the collected radar .tdms data files. Each rail transit line was divided into 10-meter blocks and a JPEG file was generated for each block. A flow chart describing the Matlab code frame is shown in Figure 57.
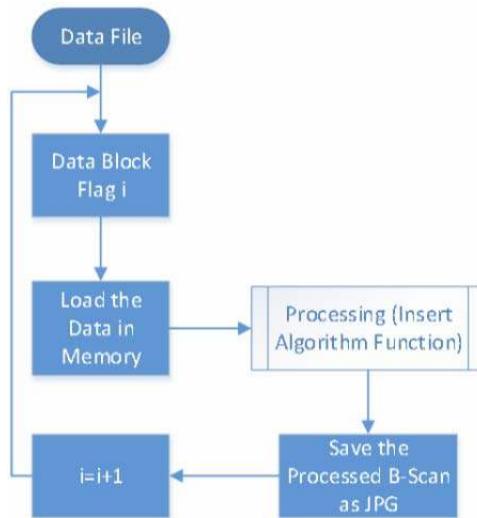
Figure 57 Flow chart showing the Matlab code frame.

During batch processing the GPR data, the Matlab program went through data blocks one by one using an address pointer as the indicator. For each data block, the GPR data was processed by the above enhanced 2D entropy algorithm. The processed result was then saved as a JPEG image.

Chapter 4

GPR DATA ANALYSIS

4.1. Introduction

Ground penetrating radar (GPR) has become an increasingly useful tool in inspecting railroad track conditions. The use of GPR in ballast assessment is desirable, since it provides a quick and cost-effective method to examine ballast conditions in large stretches of track without destroying the track by digging. Analysis of GPR scans used in this manner requires processing of the data to reveal subsurface features of the ballast, including basic signal filtering and noise removal, and often other techniques to highlight and locate any anomalous areas within the ballast. The development of such a data analysis system is required in support of the GPR developed for this project.

GPR has been widely used for inspecting highway pavements and other structures made of homogeneous materials, which in general consist of one substance and do not contain large voids. The challenge with applying GPR to railroad ballast inspection is that ballast contains many relatively large voids. These voids give ballast its desired characteristics for distributing the traffic load across the track. However, these voids make it difficult to apply GPR in a traditional manner, since they create heterogeneous material layers [*39*]. For highway GPR applications, it is possible to use the reflection coefficient method to identify several boundary layer changes under the pavement surface quantitatively. While for railroad applications, the heterogeneity of the ballast makes this method unreliable past the first surface interface. A further complication is that the ballast causes signal scattering with certain ranges of radar frequencies due to air gaps. The scattering causes signal losses and distortions that make subsequent layers difficult to measure [*39*].

The fundamental filtering for B-scan processing involves the development of low-pass and high-pass filters, to remove any Direct Current (DC) offset of high-frequency harmonics outside the GPR operational bandwidth. Also, railroad ties and overhead transmission lines can create a noisy environment for the GPR system. Therefore, it is necessary and important to perform background noise removal to isolate GPR signals representing the ballast layer. The filtering techniques are effective in cleaning up the B-scan data, but further work is still needed to locate the depths of anomalies in the ballast layer. Using the dielectric properties of ballast and the two-way travel time of GPR signal, it is possible to calculate the distance/depth where a significant GPR signal reflection occurs.

This chapter focuses on the processing and enhancement of raw B-scan images generated from railroad track radar scans. The purpose is to produce images that can be used to access subsurface ballast conditions. This task requires the development of many quantitative tools to support the enhancement of B-scan images and to locate the depths of subsurface anomalies. The scheme for developing these tools is as follows: A basic set of digital filters was developed for initial B-scan image processing. Once these filters were developed, a basic software package with Graphical User Interface (GUI) was created to display the filtered B-scan images and allow users to control the filters. Next, work was done to develop algorithms capable of calculating the depths of anomalies in ballast. These algorithms were integrated into the software package.

4.2. Literature Review

We started the literature review with researching the general functionality and theory of GPR systems. It provided a necessary framework for understanding the specific requirements of this task, which includes filtering raw B-scan images and identifying the depths of anomalies in ballast.

Although the purpose of this task is not to design GPR hardware, understanding radar type, configuration, and frequency band helps set appropriate parameters for filtering techniques and determine the information that will be available in the collected data [39,40,41]. Also, it is important to understand the physical characteristics of the medium with which the GPR interacts. These include the attenuation rate of GPR signal and complex dielectric characteristics. Dielectric characteristics manifest through the reflection strength and propagation rate of the GPR signal across the medium [39,42]. The filter selection of a GPR data processing system generally requires the ability to perform band-pass filtering and background noise removal, though this is largely dependent on the quality of the raw B-scan data. Background noise removal and ringing removal are typically needed for air-coupled radar systems. More complex filtering methods are usually left to the discretion of the analyst and are dependent on the purpose of the application [39]. For example, someone more interested in image clarity may opt for advanced filters that improve the visual quality of radar scans, but degrade signal integrity. Likewise, if filters are set too wide, too narrow, or severely asymmetrical, the collected data will be of little value for mathematical signal processing algorithms [40].

Depth determination repeatedly hinges on the calculation of the dielectric characteristics of the material, since dielectric characteristics affect the velocity at which the GPR signal travels through the material. One primary method for determining material dielectric characteristics in highway and geological surveys is the reflection coefficient method [39,43,44,45]. This method is based on three assumptions about the material: 1) its magnetic permeability is negligible, 2) it is a good insulator, and 3) its void space ratio is small and results in a negligible imaginary portion of the dielectric characteristics [45]. The third assumption is often the most problematic for applying this method to railroad ballast. Also, this reflection coefficient method depends on the acquisition of total radar signal strength to establish the ratio of the reflected signal strength [40,46]. The total radar signal strength can be acquired using a metal plate calibration process. This calibration process produces the best results when a calibration data file is acquired before each intended data collection [40].

While the reflection coefficient method has been applied to railroad ballast inspection, it can be problematic due to high signal attenuation and the difficulty in identifying boundaries, especially when the dielectric characteristics of two adjacent layers are similar [47]. Another method for calculating the dielectric characteristics is the "known depth" or "layer thickness estimation" method. This method uses two points in time and the depth of a layer to establish the effective dielectric characteristics of the layer [42,46]. It requires that the two-way travel time is accurately measureable and the thickness of the layer is known.

4.3. Core Development

The development of the GPR signal processing tool was focused on several key features. First, some digital filtering techniques (Section 4.3.1) for raw data were developed to render the radar data serviceable. Focus was then placed on developing an algorithm to apply these filters to bulk field data and produce .jpeg images of the filtered data (Section 4.3.2). With the ability to

visually examine the radar data, emphasis was placed on developing methods of calculating the ballast dielectric (Section 4.3.3) and layer depth (Section 4.3.5), with and without calibration methods (Section 4.3.4). Lastly, a method of spectrogram analysis was included through a Short-Time Fourier Transform (STFT) function (Section 4.3.6).

4.3.1. Filter Development and Graphical User Interface (GUI)

GPR B-scan images are often blurred by surrounding noise and interference. A DC offset may also occur with the signal. Hence, the following filtering techniques were developed to address these issues: a low-pass filter, a high-pass filter, and two background removal filters.

Digital signal filtering is a logical consequence of the combination of computer computational powers and the ability to mathematically model analog filters created with both passive and active electrical components. The primary difference occurs mathematically, with analog filters typically represented as continuous differential equations, while computers sample a continuous signal and process it in a discrete format. By converting the continuous equation to the discrete version, it is possible to model any analog filter as a digital filter. Design in this case can be mathematically complicated and is facilitated with tools like MATLAB, which provide a series of functions that allow users to tailor filters for applications. The low-pass and high-pass filter designs followed this route. These filter types are similar, and the same process for their implementation was followed using MATLAB.

Low-pass filters are used in signal processing to allow only the frequency content of a signal below a desired frequency to pass through the filter. High-frequency noise and harmonics were addressed by the development of a low-pass filter using MATLAB. The high-pass filter can be considered the mirror of a low-pass filter. It will allow all frequency content above a selected frequency to pass through the filter. The low-pass filter was designed using a zero-phase Butterworth filter. A high-pass filter is effective for removing the low-frequency noise that can occur from sources outside the signal generated from the radar. The same method for design and filter type that was used for the low-pass filter was used for the high-pass filter. The combination of these two filters results in band-pass filtering that is effective for viewing the important contents of the signal reflections with the radar in operation. . Both of these filters are vertical methods of filtering data; the filters apply to each signal trace as it passes down through the two-way travel time columns.

While low-pass and high-pass filtering were effective at removal of the spurious frequency components, issues with strong horizontal features were still encountered. These effects, which are typical, are referred to as background noise but come from different sources. "White noise" is a term that refers to a signal that contributes constant power across all frequencies (for further discussion on white noise and other noises, see Carlson and Crilly [48]. Ringing occurs from residual signals of previous radar pulses. Ideally, this residual energy would be lower than thermal noise, but this is rarely the case [39]. Both of these interferences will present as an average contribution to a broad spectrum of signal power and, as such, will escape high-pass and low-pass filtering. This noise can be addressed with background noise-removal filters that operate by using a built-in averaging function in MATLAB.

While the filtering just described was very effective at removing pollution to the B-scans from undesirable frequency contributions, random jittering still occurred in the image, which was escaping both of the noise-removal processes since random jitter would not correlate to any signal. This noise may present from thermal noise or timing "skips" in data recording at high

frequency. Another issue that needed addressing is the possibility that deeper reflections may appear overly washed-out from attenuation of the signal passing through various mediums or suffering from scattering. These issues were addressed by an anti-jitter filter and an automatic gain control function.

After these filters were developed, a GUI (Figure 58) began to take shape in tandem. Primarily the GUI made checking the effect of the filters easier, allowing a quick way to load a data set, filter a data set, and observe the effects of the filter on the data in a meaningful way. It became recognized that the GUI was also a very pragmatic way of presenting and using our filters for the overall project, and was adopted as an objective for our team.
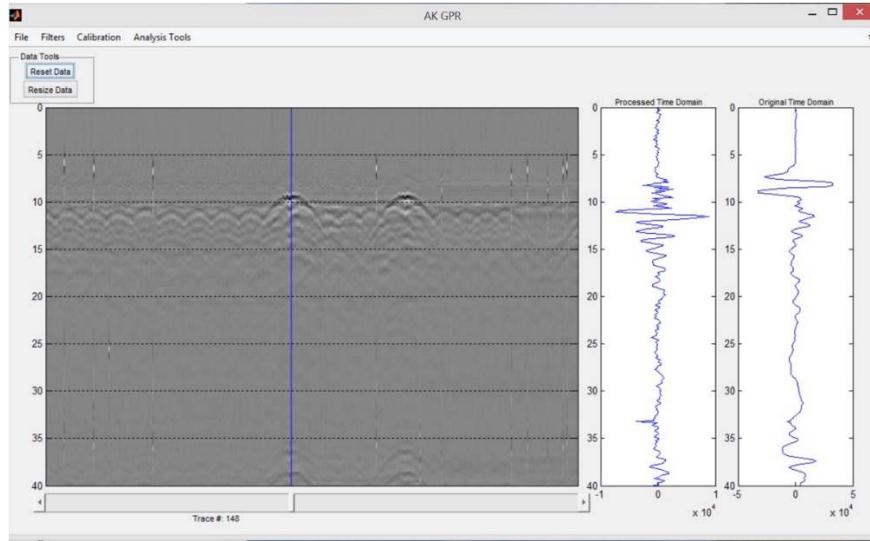
Figure 58 Example of the GUI interface

The GUI was designed to show time domain "waterfall" trace data, both raw and filtered, to the right of the B-scan, allowing the user to simultaneously look over the region of interest with the B-scan, while also looking at individual trace signal behavior. The ability to reset the filtered data and resize the B-scan image to allow the user to "zoom in" on an area of interest was also added.

4.3.2. Automatic B-Scan Processing

The filtering processes enabled performance of basic analysis of the radar data. An additional need of the project was the ability to handle the processing of field data in bulk. Test field data were generated from three different tracks in August 2013. This effort resulted in around 400 GB of field data. One of the objectives of the overall project was to develop a web-based interface for a database, where the track scans could be uploaded and accessed by other users. Processed radar scans were intended to be part of that database. The data recorded by the radar system are initially saved in a .TDMS formatted file. However, loading the .TDMS directly requires a selection of a span of traces to represent the B-scan, since typically the .TDMS files are too large to be opened in their entirety. Saving this information to .MAT is useful because it allows the direct importation of relevant data as a structure into MATLAB for further processing.

One specific filter processing algorithm was developed for this project, however, is not presented hereinafter. The filtered image is saved to .jpeg format ultimately. This algorithm was never formalized into a function of the GUI, as it was mainly intended to be a utility to address the problem of producing bulk images of the processed radar data from the field. With 400 GB of data to process, the total automated images alone took around three days of continuous processing.

4.3.3. Dielectric Calculation Methods

The most important physical characteristic that GPR attempts to take advantage of is the relative permittivity or dielectric constant of materials. This property is a measurement of a material's ability to store and release electromagnetic energy in the form of electrical charge. This storage and release of energy is important for GPR work, as it governs signal attenuation and can provide insight into the type of material contained in the subsurface [39]. The dielectric constant will also affect how fast a signal can travel through a medium. Once the dielectric of the material is determined, the distance can be calculated between two points in time.

Initially, three different methods for calculating dielectrics were planned for the program. The first method measured the spreading of an anomaly's hyperbola or "wake" as the radar passed over it to calculate the dielectric constant. It is common in soil and archeological surveys. The limitation for use of this method, however, comes from its dependence on the presence of a strong uniform reflector (like a metal rod or pipe) within the medium. Small, localized pockets of water might produce the desired pattern, but the majority of ballast fouling will not present in the radar scan in this manner. Soil and water fouling typically will not be uniform in shape and may not reflect uniformly depending on the degree of fouling in the area of fouling.

The second method used reflection coefficients, thus was referenced as the reflection method. It is commonly used in soil and roadway analysis. The method depends on determining the maximum possible signal reflection (usually determined with a metal plate calibration), and finding the ratio of the peak reflection values of the signal over the maximum reflection strength. It works best when signal attenuation is small within the material being measured. The more attenuation of scattering of the signal that is present, the worse this method will perform in calculating the dielectric of a layer. This method has been experimented with heavily, as it would provide the simplest method for automating the ballast layer dielectric calculation and would work well with the desire to calculate depths of anomalies.

In the third method, the depth of an object and the velocity of the signal (to be referenced as the "depth method") were considered. This method takes advantage of the fact that the radar signal is an electromagnetic wave that, in vacuum, would travel at the speed of light, but the permittivity will affect the velocity of propagation for the wave. For application with ballast, this method would require identifying the air-to-ballast interface region and the ballast-to-soil interface region, and knowing the approximate depth by design. The main benefit is that it will allow the accurate calculation of the location of anomalies within the ballast, since the calculated dielectric value will scale the velocity of the signal to represent the ballast as it is presenting in the radar data.

4.3.4. Calibration and Surface Location Identification

In order to begin analysis of subsurface ballast conditions, it is necessary to identify the surface location of the ballast. Visually, this requires filtering to remove the background noise and time shift to reveal the reflection features within the radar data. While this step was accomplished with the filtering tasks, it was also desired to identify programmatically the surface location and attempt to identify automatically other anomalies within the ballast. To calculate depths, it is necessary to identify the dielectric value of the ballast layer.

A metal plate calibration method was developed to obtain the maximum amplitude of reflection to allow development and exploration of the reflection amplitude method (Section 3.3). The method obtains the maximum amplitude of reflection by placing a metal plate of some dimensions large enough to fit fully under the area of the antenna, and running the radar for a short period to sample the reflected signal. This radar data file is then loaded through the GUI for processing. The calibration method developed was meant for the reflection method and was designed by this group specifically for finding the maximum amplitude of reflection and estimating the surface location. The process uses the entire radar data presented in a .MAT or .TDMS file. Once the entire calibration file has been processed, the algorithm calculates the average value of the maximum amplitude column and the average time value of the time column. These two values are stored within the GUI for later use.

The surface location time is expected to be slightly earlier than the ballast time since the metal plate will sit above the ballast and also have some thickness. This estimation is still useful since the difference should be very close to the actual surface time. The actual surface location is found by implementing the same method when the depth calculator is loaded in the GUI. In this case, the frame of data presented in the radar image is passed to the calculator in its raw form, and the same filtering process is reapplied to the region being evaluated. The same identification and averaging of maximum values and times occur. These values were used with the calibration data to compare the surface time and use the reflection coefficient method to estimate the dielectric of the ballast.

4.3.5. Depth Calculation

The culmination of the development of the filtering methods, dielectric calculation methods, calibration, and surface location is the ability to integrate these processes for the sake of identifying fouling regions within the ballast layer. A depth calculator tool was developed to accomplish this task. The calculator operates around the known depth method (Section 0), in which the equation can be rearranged if the dielectric value of the ballast layer is known to calculate the depth within it. The calculator is accessible from the GUI. When the calculator is loaded, it first accesses a raw copy of the radar data, applies the filtering equivalent to that used in the calibration method, and finds the surface time of the ballast layer. The user is then presented with the depth calculator panel (Figure 59).
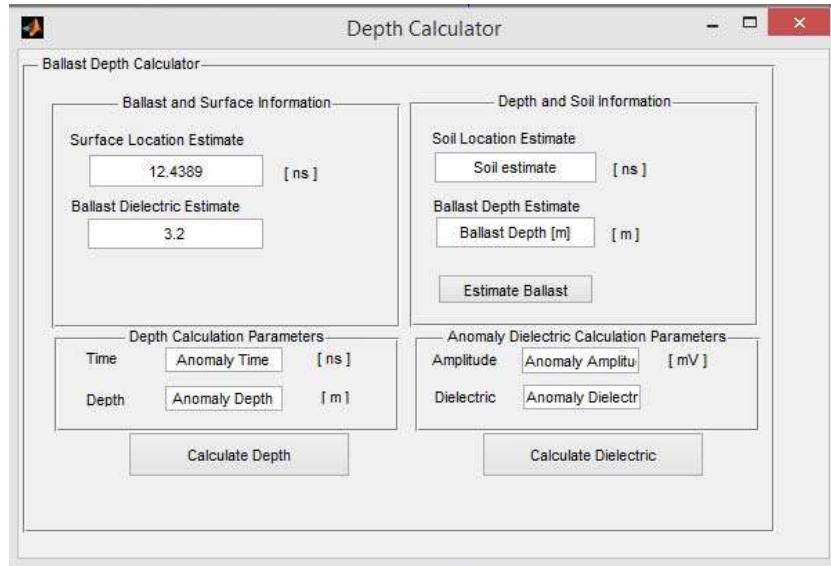
Figure 59 Depth Calculator interface

The surface estimate provided by the calculator is hard set from the algorithm. The ballast dielectric estimate is a preset default (3.2 in this case) that can be modified by the user. The dielectric value for the ballast can be entered either by the user or calculated with the "Depth and Soil Information" panel. Entering the ballast to soil interface time, which can be observed from a B-scan, and the depth of the ballast, obtained from track information, can be used after a surface time has been entered to get a close functional estimate of the ballast dielectric value. Depth within the ballast layer is then calculated by entering the ballast dielectric and the anomaly time in the "Depth Calculation Parameters" and pressing the "Calculate Depth" button. An optional box for calculating the anomaly of the dielectric is provided on the right.

4.3.6. Short-Time Fourier Transform Analysis

An STFT feature was included in the analysis algorithms to provide the ability to look at the spectrum content of the signal. The Fourier transform as a general description is a function that converts a time domain signal into a frequency representation of the signal. This transform is particularly useful in communication systems engineering for evaluating the content of signals; it allows an individual to see what frequencies are contributing to the observed time domain signal and analyze the magnitude of these contributions. This feature can be useful for radar, in that the spectrum content can reveal characteristics of the material through which the signal is traversing. The feature also can verify that the correct frequency is being used by the system.

A limitation of the Fourier transform is that it loses all time information when applied to a time signal. It will reveal the sinusoidal content to the signal by frequency, but not when that content manifests the overall superposition of the time signal. A solution to this limitation is to introduce a window to which the Fourier transform is applied. Instead of applying the transform to the total signal, a localized potion of the signal is analyzed at a given time. The tradeoff to this approach is that it loses an overall view of the total frequency content. Window size in this respect is important, as it can show the spectrum content in a given portion of a signal, but not

exactly when that frequency content is added. Smaller windows allow better time resolution but worse frequency resolution, and vice versa for large windows. This tradeoff is acceptable in radar analysis where it is desirable to see the localized changes to the signal as it progresses through time. Window size can be manipulated by the user to provide enough frequency information, which is pertinent to a given temporal location of the signal. This technique of windowing is referred to as STFT; it provides a method of analyzing the content of individual signal traces making up the radar scan.

The information provided from applying STFT to a radar signal is in seeing the energy of the frequency content as it passes through time, which will be represented passing through materials. The losses will be observable through signal attenuation, which should be most significant in the higher frequencies and tend to slope toward the lower frequencies. This attenuation rate will also be material-dependent, and changes can indicate when layer type has changed. Reflection points on the other hand should be indicated by an increase of frequency power near the time at which the reflection occurs. Manipulating the window sizes and overlap percentage can reveal other information or help clarify the time regions in which frequency changes are occurring. Other information may also be available to the user through clever manipulation of the STFT function.

4.4. Application of the GUI Interface and Algorithms

The algorithms and GUI developed over the course of the project were tested during various stages of the projects. Three major opportunities for application presented themselves in terms of available laboratory data and field data. The GUI was used in a general analysis of a fabricated test track provided by the University of Massachusetts, with the data acquisition performed by the University of Vermont. In addition, a comparison of evaluation methods between teams was performed for a length of test track. A full discussion of these applications is provided in Sections 4.1 through 4.3

A general use application of the software was performed on the test track in an attempt to identify a fouled region within its ballast layer. Two more tests were performed: one where water was periodically added to the test track and scans were performed, and another, where the track was inundated with water and scanned periodically while it dried.

The field data were used in two ways for checking the GUI. One short section of track was selected to examine in an exploratory sense: load the data and use the GUI to evaluate the section like a field examination. The second use of the field data was to conduct a comparative analysis with the University of Vermont (UVM) over a much larger span of track. Since no ground truth was available for the field data, the accuracy of the results cannot be verified.

4.4.1. Analysis of Lab Data

The total functionality of the algorithms and the GUI can be demonstrated through a usage example of the program on the lab data generated by the UVM. The example in this section is intended to provide a full walkthrough of the basic usage of all the functions provided in the interface to perform analysis with a set of radar data. The test bed was laid out as depicted in Figure 60.
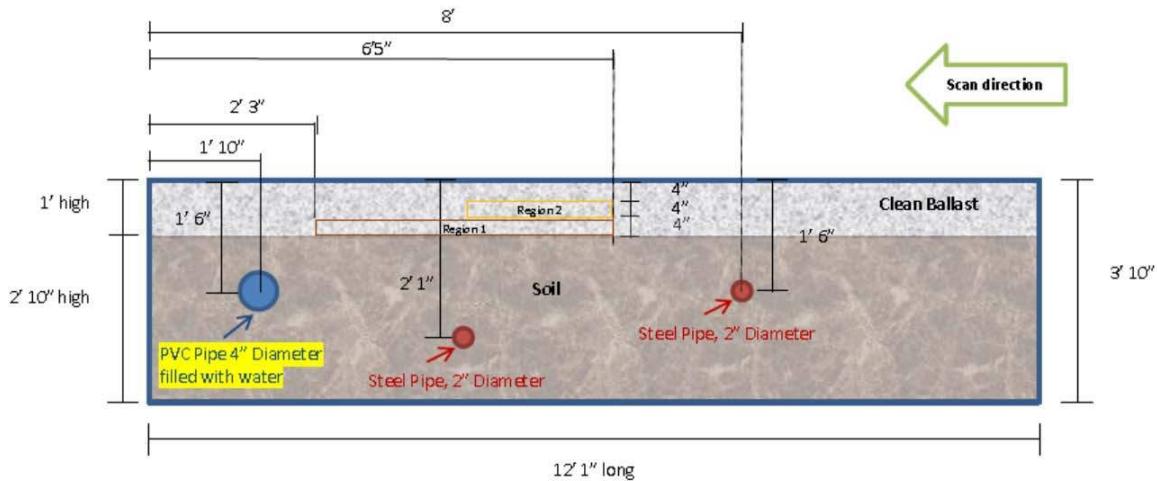
Figure 60 Test bed schematic identifying region 1 and 2 is the objective of this example.

The ballast material occupies the first foot of the test track, with the remaining 2 feet 10 inches filled with soil. Two metal bars are buried 1 foot 6 inches and 2 feet 1 inch into the soil under the ballast. There is a PVC pipe buried 1 foot 6 inches into the soil layer as well. The ballast has been fouled in two areas at two depths. Region 1 is fouled with soil for about 4 inches, and region 2 is fouled with sand on top of the soil, fouling region 1. This example is focused on identifying regions 1 and 2 for the sake of ballast analysis.

After the test file is loaded, the GUI will present the raw radar scan (sometimes referred to as a B-scan) image. Then the Auto Apply feature is used under the Filters menu to clean up the B-scan with low-pass and high-pass filters, KLT noise removal, and anti-jittering. Subsurface features are now visible in the B-scan. The dark band being pointed to in Figure 61 is indicative of some ballast fouling. The image can subsequently be resized to help enhance the view of the region in which it is contained. The B-scan can be resized horizontally to show the trace segments between 43670 and 150076 by using the resize data button. These segments were selected by moving the trace marker close to where the red lines indicate and recording the trace numbers. When the desired horizontal parameters are entered, the B-scan will automatically update to display the traces within this region.
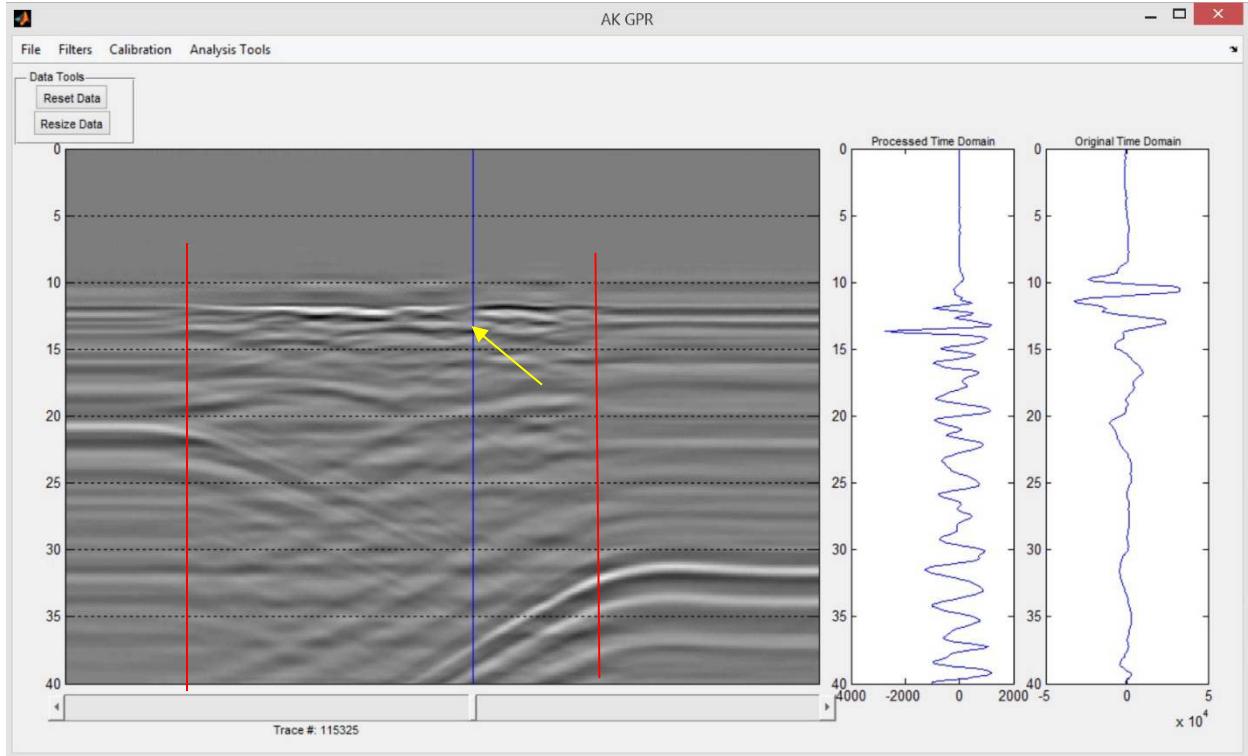
Figure 61 Filtered B-scan image showing a dark band under the surface of the ballast as pointed out by the arrow. The trace marker has been moved to select a trace containing one of the more prominent portions of the dark band. The red lines indicate the boundaries of the area of interest.

Since fewer traces are represented but the image area remains fixed, resizing has the effect of appearing to "zoom in" on the region of interest. Before further enhancement is done the calibration file is loaded with the Calibration menu. The program will process the selected calibration file, applying low-pass, high-pass, and background removal filtering to the calibration data, and then identify the surface location and the maximum reflection amplitude from the metal plate, which will allow depth calculation later. Each step of the process will be indicated through a status bar update. The results can be obtained by selecting the Calibration Results option in the Calibration menu.

Our interest is in finding the fouled portion of the ballast region, so by taking advantage of the estimated surface location and the assumed dielectric value we can expect the soil region interface to be around 16 ns. We can use this to resize the B-Scan further to highlight the ballast region alone. The resize data button is used again, and this time vertical parameters are given as 8 ns and 17 ns to provide a window slightly above the air to ballast interface and just below the expected ballast to soil interface (Figure 62). The resulting enhancement shows clearly an interface at about 12 ns and a strong reflection between 13 to 14 ns near the blue trace marker:

54
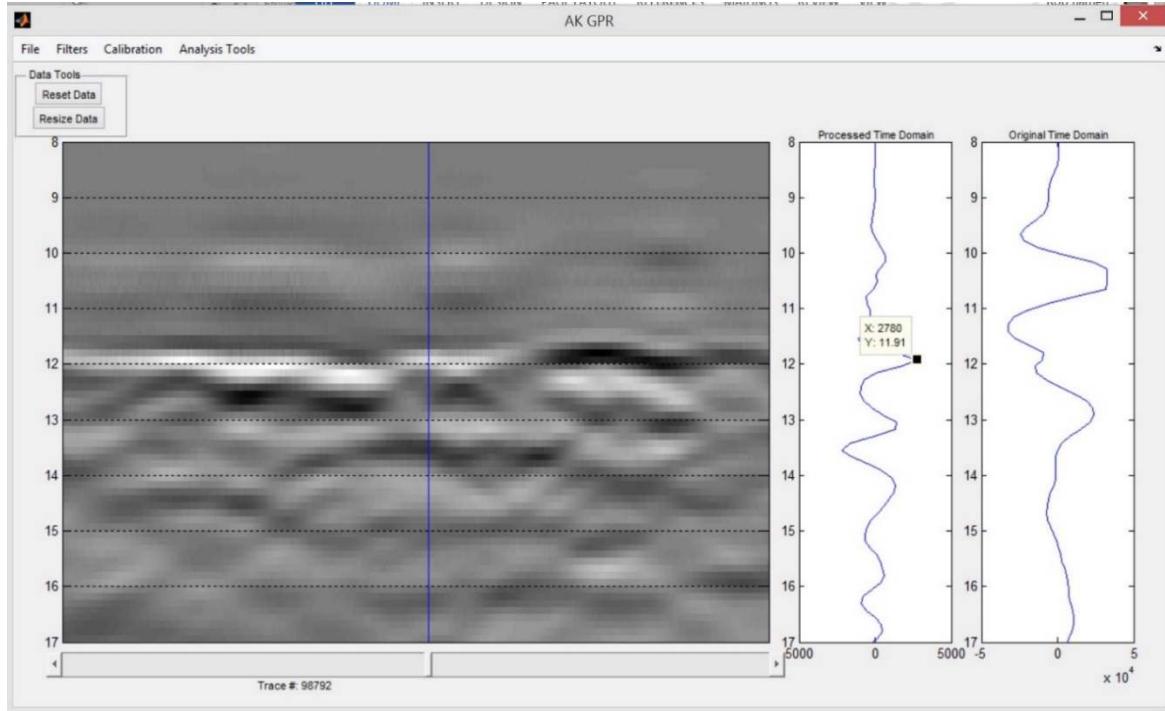
Figure 62 Resized B-scan to emphasize the contents of the ballast region. There are now two distinct reflections in the trace selected by the marker.

To examine this trace specifically we begin by selecting the STFT tool from the Analysis Tools menu. This will use the raw data and the parameters provided through resizing and examine the spectrum content of this trace. A window size of 2 ns was selected with 80% overlap to provide good time resolution while looking for the regions where there is a broad spectrum increase in frequency content (Figure 63).
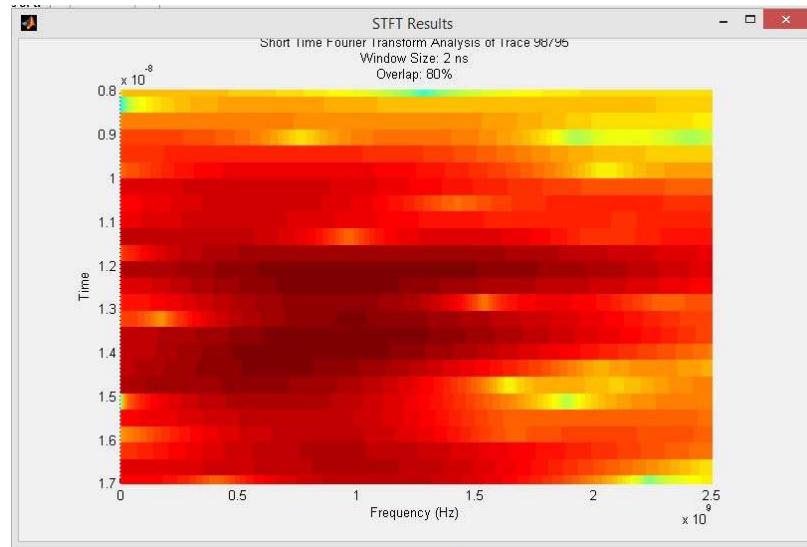
Figure 63 STFT results of the trace using a 2 ns window and 80 overlap to look for the reflection regions of the radar pulse.

In this case there is a lot of energy in the trace near 500 MHz near the 12 ns mark which suggests this is the region where the surface reflection is occurring. There is another region of strong energy increase in reflection around the 14 ns region. This suggests that there is another reflection occurring here. Looking back at the B-Scan and selecting the peak on the Processed Time Data indicates the reflection is occurring at 13.54 ns (Figure 64).
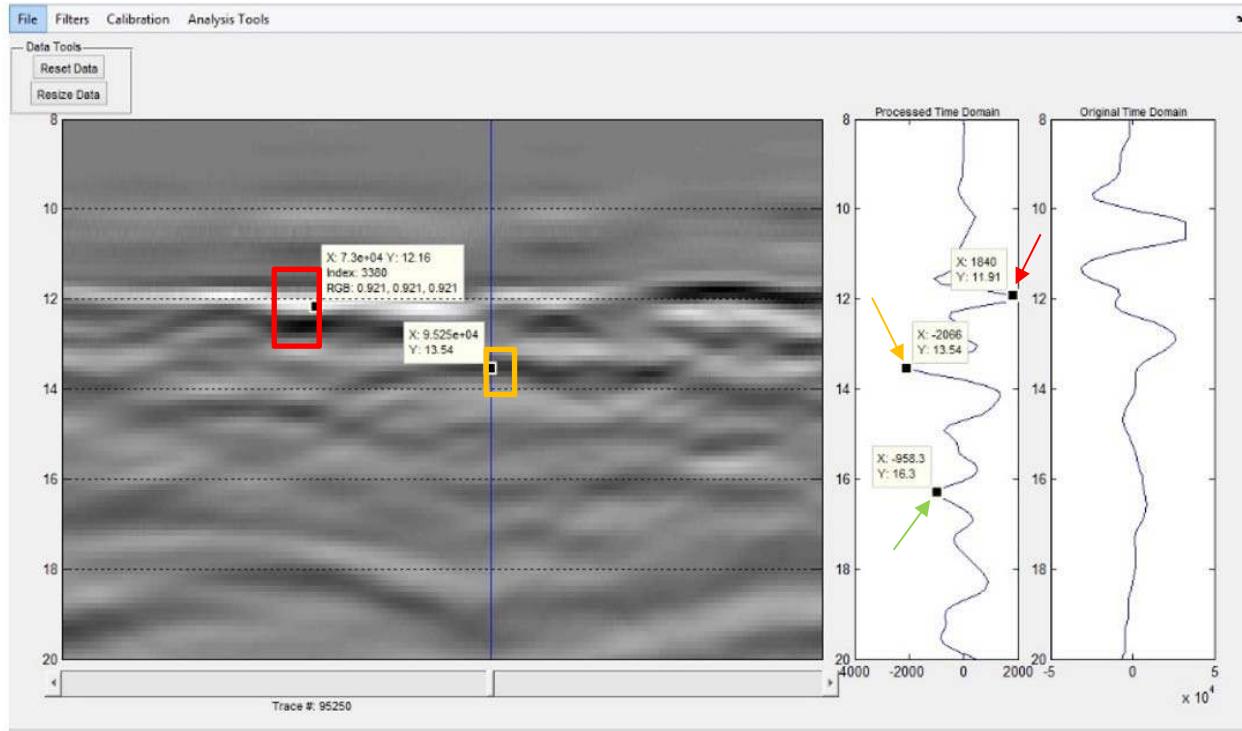
Figure 64 The red box indicates the region where the surface of the ballast begins. The red arrows correspond to the reflection in our trace where the surface is located. The orange box highlights the anomalous reflection region and the orange arrow and point marker provide the data for the reflection at that point, giving a time of 13.54 ns. The green arrow indicates the reflection of the ballast to soil interface.

This corroborates well with the STFT information. The Depth Calculator is opened from the Analysis Tools menu in order to calculate how deep into the ballast this is occurring. For the depth calculation it is necessary to determine an accurate dielectric value of the ballast layer. In this case this is done by utilizing the known depth method. :

The "Depth and Soil Information" panel was used to allow a better estimate of the dielectric value of the ballast. In this case we took the surface bottom to be the estimated 16.3 ns, and the surface time to be the calibration estimate of 12.16 ns. Ballast depth in meters is 0.3048 meters. The dielectric in this case is estimated to be 4.145 (Figure 65). When the dielectric constant for the ballast is calculated with the depth and soil information and the peak information is entered for the magnitude and time, pressing calculate reports the depth as 0.1016 meters, or 4.0 inches. Fouling is expected between 4 to 8 inches given the test track arrangement.

56
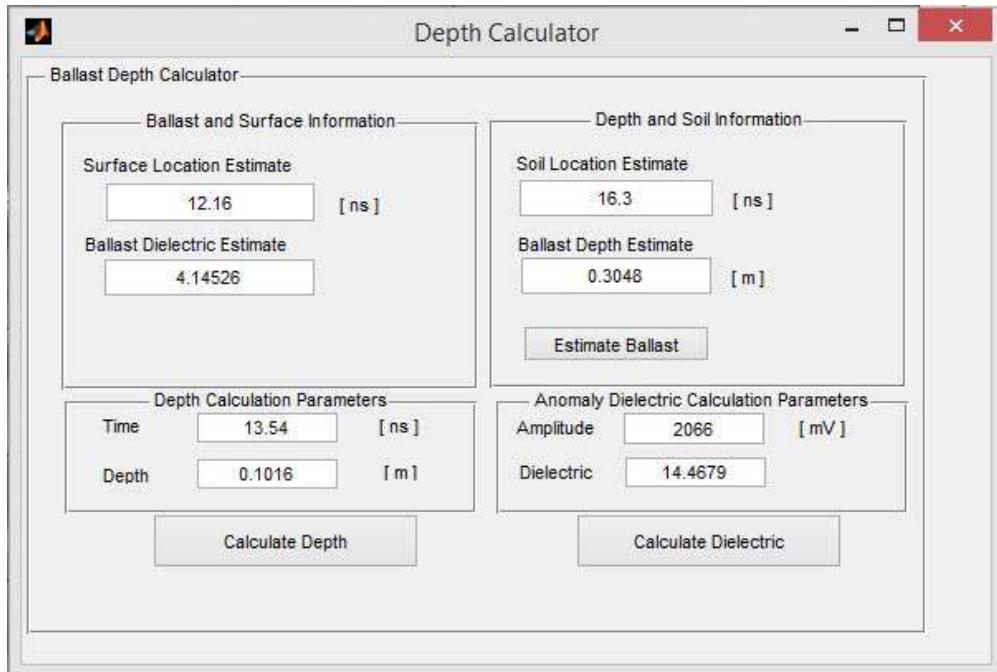
Figure 65 Depth calculator given a dielectric constant estimate of 4.145, an anomaly time of 13.54 and an amplitude of 2066 as indicated in figure 50. The results show the anomaly is occurring at 0.1016 m with a dielectric value of 14.5.

An important note is that the anomaly dielectric constant calculation is currently not verified to be accurate. The large magnitude is indicative of water saturated soil (14.468).

The dielectric constant for the ballast is calculated to be 4.14526 and the peak information is entered for the magnitude and time, pressing calculate reports the depth as 0.1016 meters, or 4.0 inches. Fouling is expected between 4 to 8 inches given the test track arrangement. This result indicates that the dark band around the blue trace marker is part of the fouling region. An important note is that the anomaly dielectric constant calculation is currently not verified to be accurate. A complete tutorial for using the GUI is found in section 4 of the User's Guide.

Additional tests were done with the lab data to cause water fouling of the ballast. In one protocol, several gallons of water were added periodically to the test track, and GPR scans were taken immediately following. In a second protocol, 4 gallons of water were poured over regions 1 and 2, and scans were taken periodically while the ballast was allowed to dry. The GUI was used to visually detect the changes within the ballast during these protocols. In both of these cases, many scans were taken over the course of the lab experiment. The most illustrative examples of the results have been selected for this section.

The first protocol results are illustrated in Figure 66 through Figure 68. These figures represent the radar scans taken immediately after 2 and 5 gallons of water were poured across the test track, which correspond to the beginning, midpoint, and end of this protocol. The biggest difference between Figure 66 and Figure 67 is the noticeable sloping of what is presumed to be soil fouling in the ballast region. Immediately present is a very strong but short reflection that occurs around the 16 ns marker at the bottom right. This reflection is indicative of pooling water near the ballast-soil interface.
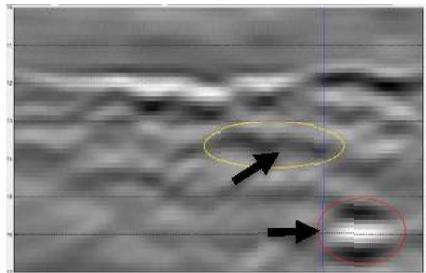
Figure 66 Two gallons of water, scanned immediately, highlighted areas are the fouled region and possible pooling at the soil interface.
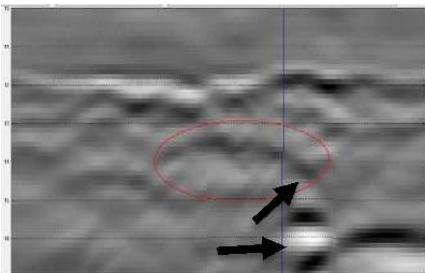
Figure 67 Five gallons of water, scanned immediately. Highlighted areas have changed from Figure 12 showing a more sloping pattern from the fouled region to the pooling region.
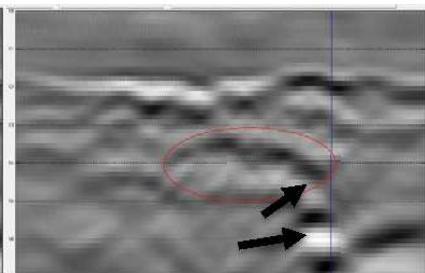
Figure 68 Nine gallons of water and a 10-minute wait before scanning. The original fouled region is much more prominent from saturation with the pooling region still present.

Figure 68 shows the state of the ballast after 9 gallons had been added to the track and after 10 minutes had passed. The suspected fouling region is very pronounced in this case and visibly slopes downward throughout the addition of water. This increase in prominence is likely from water saturation in the soil, and washing of saturated soil from the fouling section toward the area that appears to be pooling water at the ballast-soil interface. The radar system and GUI were successful in being able to pick up the changes occurring within the ballast as varying degrees of water fouling occurred.

After the ballast had time to recover from the treatment just described, a second experiment was conducted to evaluate the radar's sensitivity to moisture levels. Four gallons were poured over the fouled region, and scans were taken immediately and for up to 20 minutes to examine the region as it dried out. Again, the beginning, midpoint, and end of the scans are presented.
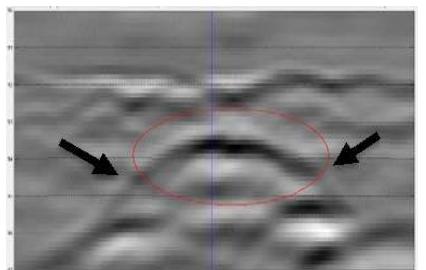
Figure 69 Fouled region immediately after 4 gallons of water were added over the fouled regions.
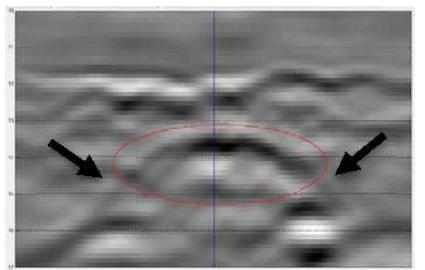
Figure 70 Ten minutes after water was poured over the fouling regions. A detectable decrease in reflection is occurring on the sides.
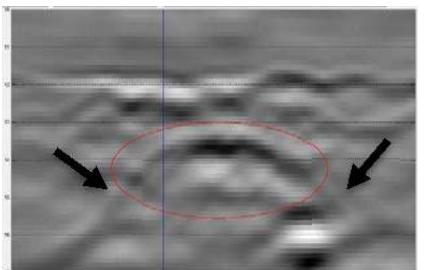
Figure 71 Twenty minutes after water was poured over the fouled regions. Noticeable reduction in the reflections on the side has occurred.

Figure 69 represents the state of the fouled region immediately after 4 gallons of water were added over the fouled region of the test track. The top of the fouled region is distinct as are

58

trailing edges, which slope down toward the ballast-soil interface where darkening due to pooling appears to be occurring on the bottom right, as with the previous protocol. Figure 70 is the state of the fouled region and ballast after it had been allowed to dry for 10 minutes. The side bands of the fouled region have lightened, as would be expected. The conditions of the ballast after 20 minutes of drying are shown in Figure 71.

In both of these cases, the top of the ballast layer was calculated to be 3.6 inches into the ballast region. This is reasonably close to where the fouled region was supposed to be by design (Figure 60). Both of these protocols verified that the radar and the processing algorithms are sufficient to detect changes within the ballast region and provide reasonable depth calculations. The data in this case were limited to what the actual change in moisture was over the course of the tests, as moisture content was not measured in these cases.

4.4.2. Analysis of Field Data

The GUI was used to conduct some basic analysis of field data collected along the Boston Green Line. Verification of our results was limited by the absence of any ground truth for the track. A test sample of 400 segments was selected for analysis. The associated calibration file reports that the surface should be around 9.78 ns. The surface visually appears at approximately 10 ns with what appears to be another boundary layer occurring just beneath 15 ns. The surface estimate from the depth calculator estimates the surface is at 9.03 ns. The discrepancy is likely due to the large amount of jittering that remained even after anti-jitter filtering. Limiting the vertical span between 7 and 20 ns "zooms in" on the suspected ballast region.

Within the radar scan, three clear features appear within the data. The waterfall trace in Figure 72 has had markers added at the reflection points that correspond with these banding lines. The top reflection is likely near the surface boundary. This particular trace occurs at 10.41 ns, probably because this trace's reflection happens to be lower than the average. The bottom band associated with the 17.16 ns marker is likely the ballast-to-soil interface. There is a light but distinct band roughly in the middle. Depending on how the ballast was constructed, this band could indicate light soil/fines fouling. Resizing the image to show between 20 ns and 40 ns allows examination of sub-ballast features.
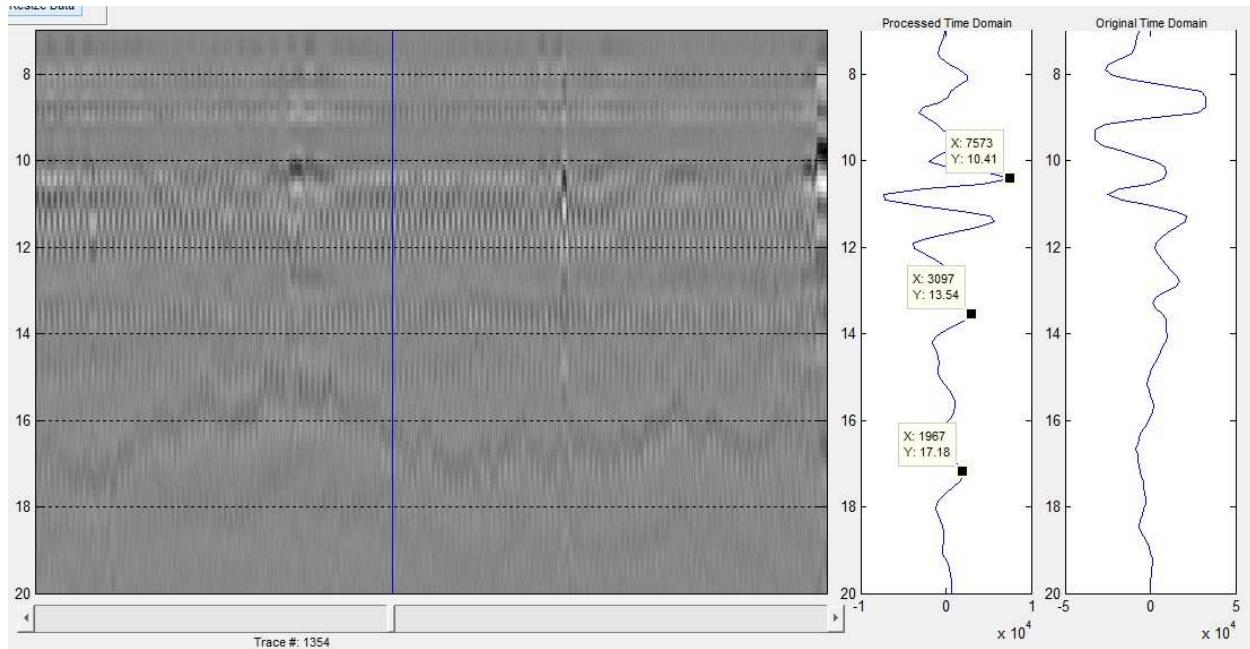
Figure 72 Enhancement of the filtered radar scan limiting the scan between 7 ns and 20 ns. The waterfall has markers at the three significant reflections in a given trace, which correspond with features in the radar scan.
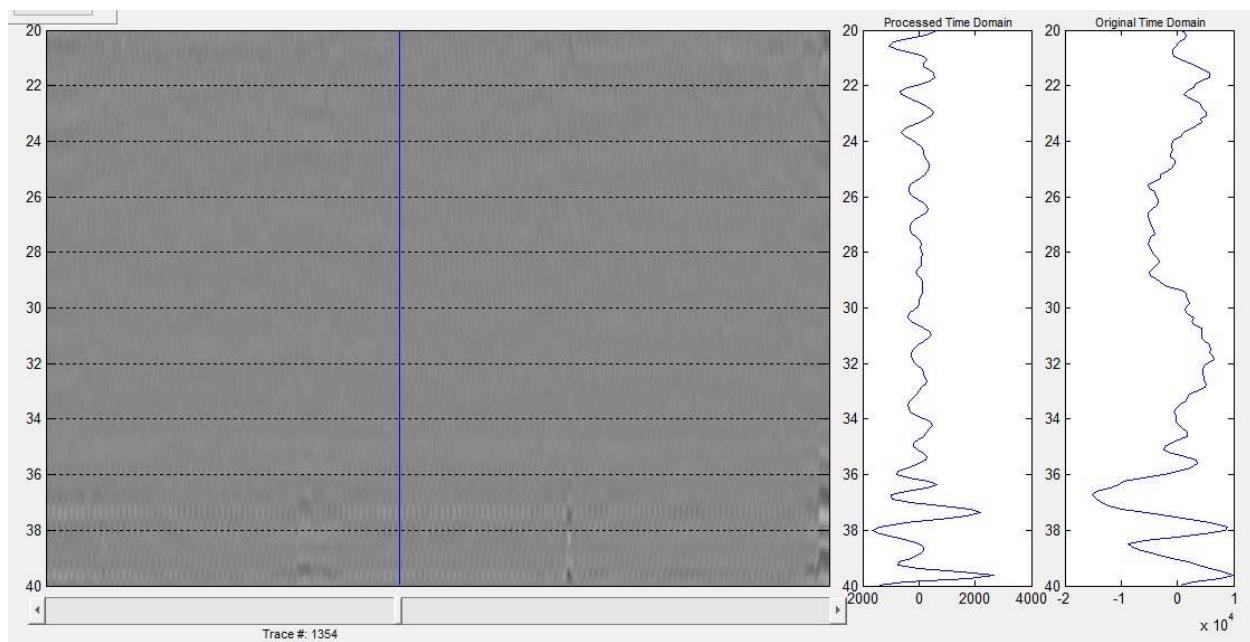
Figure 73 Resized radar scan to show the region between 20 ns and 40 ns, the expected sub-ballast layer.

Figure 73 represents this sub-ballast region. This region looks clean and homogenous until near 40 ns, but this distance should be much deeper than the radar is expected to be effective and can be discounted.

Even without enough information about the track for full quantitative analysis, it is possible to use the GUI to qualitatively identify areas of concern within the ballast. In this particular span of track, the bulging and dipping of significant portions of the ballast-soil layer may be interesting and warrant further investigation of the area for deformities or concerns. Severity of dipping or infringement of the soil layer into the ballast could be determined by having an accurate dielectric estimate of the ballast layer. Unfortunately, this estimate would require more information about the track than is available. Using upper and lower dielectric values of 3.0 and 4.5 [42] and calculation equation from "depth method" (Section 3.3) indicates the ballast region could be anywhere from 0.401 m to 0.273 m, which is a valid range for ballast depth depending on requirements of ballast design in the area.

4.4.3. Comparison of Results between Teams

A comparison of results was conducted between the University of Alaska Fairbanks (UAF) and the UVM in order to assess the teams' methods of evaluating ballast conditions. For this comparison, portions of the St. Louis Metrolink were selected. The objective was to conduct general analysis of the sections and evaluate the radar scans, particularly looking for possible anomalies. The selected radar scans were of approximately 10 m representations of track. No ground information was provided for the actual conditions of the track. The calibration file indicated that the surface should be about 10 ns. In the following figures (Figure 74 through Figure 91, the results of UVM are presented on the left, and the results of UAF are presented on the right. The span of track represents roughly 1.2 km of track. Some sections were excluded by UVM, and were not treated by UAF. The segments not represented were 1201 to 1501, 1801 to 2100, and 2400 to 2700, out of 1 to 3600 possible segments.

The areas identified by UAF were identified by visually examining the radar scans after filtering and using the GUI to try to identify areas of interest within the scans. The UVM team had its own methods for accomplishing this task. Since this report is being compiled by the UAF team, no commentary will be made on the UVM methods or results, only description of what is highlighted in UVM's radar scans. All editing to the UVM radar scans was performed by that team. UAF applied high-pass and low-pass filtering, KLT background removal, and anti-jittering to all scans. UVM presented a vertical selection between 0 and 25 ns, while UAF presented the length of track between 8 ns and 20 ns vertically, which creates a slight distortion between the groups' images but poses no functional significance.
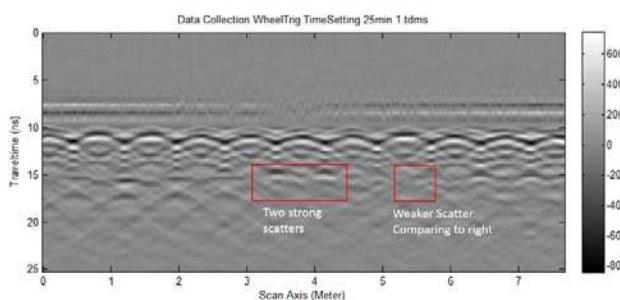
Figure 74 Segments 1 to 300 processed by UVM, identifying two strong scattering regions and one weaker scattering region.
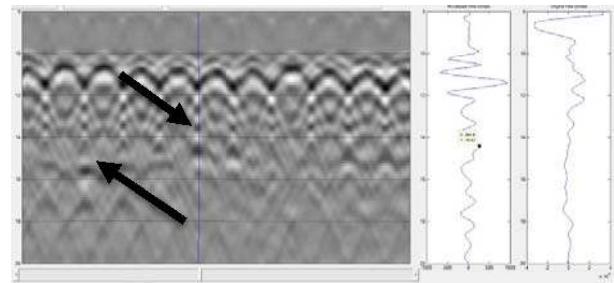
Figure 75 Segments 1 to 300 processed by UAF, identifying two reflections of interest.

61

The first segments evaluated were 1 to 300. In both cases, the prominent features are the railroad ties causing the hyperbolic pattern near the surface. In both cases, post-processing results in a similar scan. On examining the region UAF identified two reflection areas where the waterfall trace indicated a strong enough reflection to be suspicious of the area being anomalous.
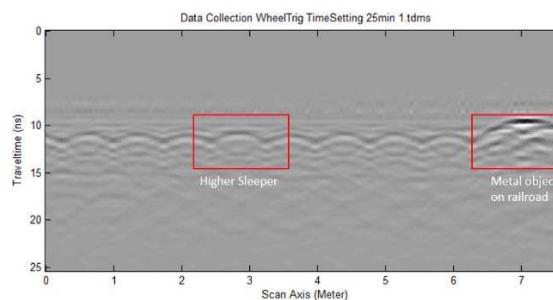
Figure 76 Segments 301 to 600 processed by UVM, identifying a high sleeper and a metallic object on the track.
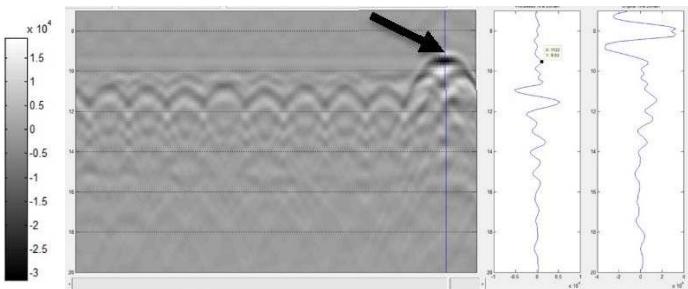
Figure 77 Segments 301 to 600 processed by UAF. The main feature of interest is a very strong reflection appearing just above the 10 ns line.

The radar scans of segments 301 to 600 were largely in agreement between the two groups. The UAF group identified the high magnitude reflection above the sleeper line as an area of interest as well. Interestingly, the peak for this region is not at the same time location in the waterfall trace. The reason for this difference could be a metal object on the track, but the nature of the anomaly is unknown.
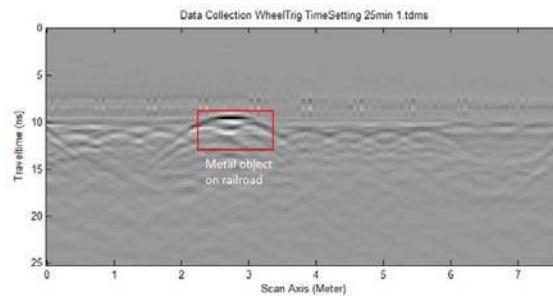
Figure 78 Segments 601 to 900 processed by the UVM group, identifying a metallic object on the railroad.
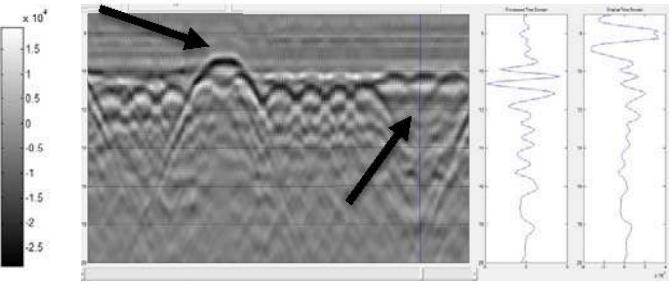
Figure 79 Segments 601 to 900 processed by UAF. An object causing a large reflection is on the right and a dipping region is on the left.

The same general features are present in segments 601 to 1200. Both groups identified a large reflection on the surface region of the track. UAF applied automatic gain control to examine the region on the left, which appears to have a break in the pattern presented by the sleepers along the surface. The nature of this apparent valley is unknown, but appears to vary from the rest of the track section and should be compared against track design to see if it is an area of interest.
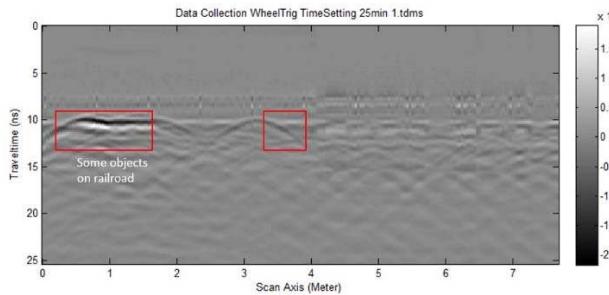
Figure 80 Segments 901 to 1200 processed by UVM, identifying some object on the road and another unidentified region.
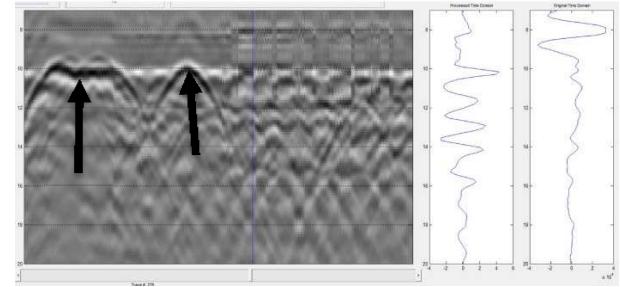
Figure 81 Segments 901 to 1200 as processed by UAF. Two areas of interest are identified, and a high degree of jittering is noted on the right side of the radar scan.

Two strong features are present for both UVM and UAF. The features appear right at the expected boundary layer. UVM identifies the left anomaly as being metal objects on the track, while not identifying the object on the right. UAF identifies both of these areas. The right object presents classically as a metallic object, while the left signature is unknown and may be structural. UAF applied automatic gain control to enhance the image due to the strange patterns on the right half of the image and discovered a high degree of jittering occurring in the data, which makes it too difficult to interpret.
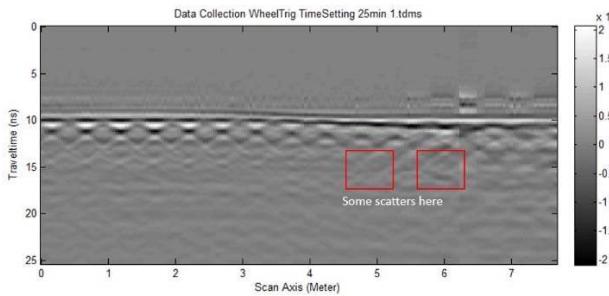
Figure 82 Segments 1501 to 1800 processed by UVM identifying some sub surface scattering.
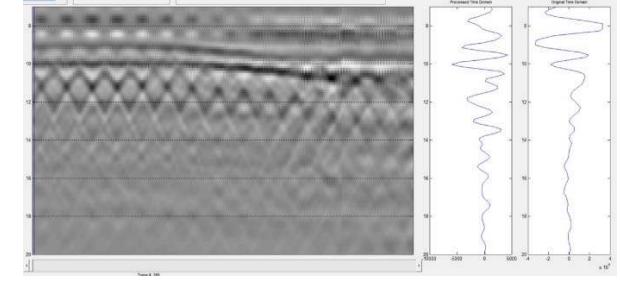
Figure 83 Segments 1501 to 1800 processed by UAF. No areas of interest are identified.

Segments between 1501 and 1800 are tame for both groups. The UVM group identified some scattering patterns in the ballast, while the UAF group did not identify anything of interest sub-ballast. In both cases some above-ballast jittering occurs, as well as what appears to be mirage effects in the data that cannot be filtered out. The surface in both scans appears to gently slope downward, left to right.
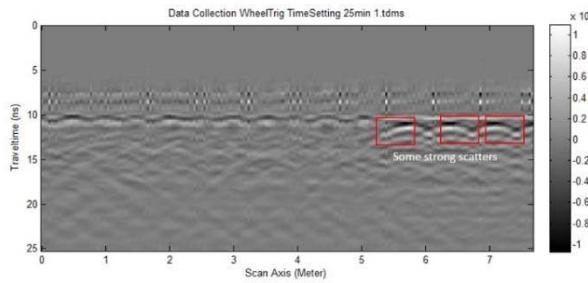
Figure 84 Segments 2101 to 2400 processed by UVM, identifying three strong scatters.
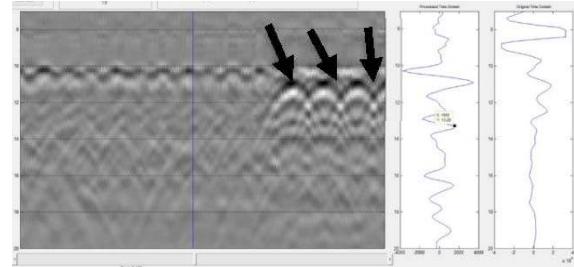
Figure 85 Segments 2101 to 2400 processed by UAF. Three sudden changes in the radar scan are identified near the surface.

Neither UVM nor UAF identified any areas of interest within the ballast for segments 2101 to 2400. Both groups did note a sudden change in the right region of the radar scan near the surface. Insufficient information is available for determining the nature of this disturbance, but it could be depressed sleepers and would warrant further investigation.
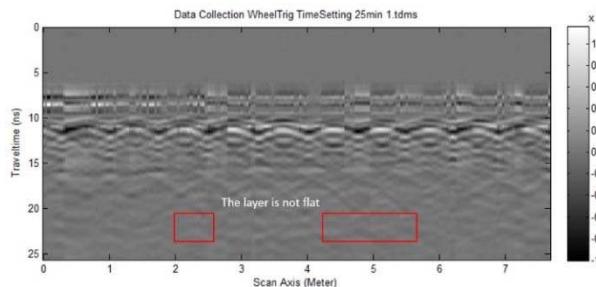
Figure 86 Segments 2701 to 3000 processed by UVM, noting the layer is not flat.
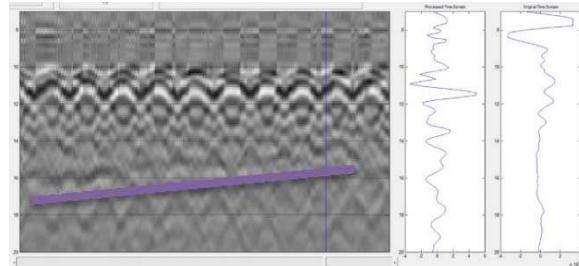
Figure 87 Segments 2701 to 3000 processed by UAF. The team highlighted a sloping region in the region.

Segments between 2701 and 3000 for both teams provided a large amount of jittering above the surface that could not be filtered. Both teams noted an uneven layering pattern in the region. UAF highlighted the signal trace that appears to slope upward from left to right under the surface. No track information is known, but this may be a narrowing of the ballast-soil interface, or may indicate an increasing fouling region of ballast in this span of track. Checking track design would shed light on this.
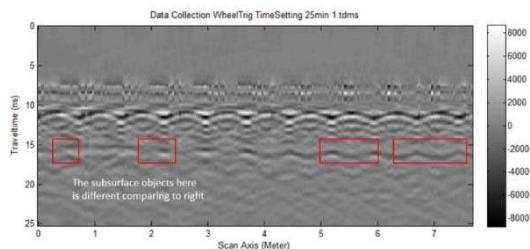
Figure 88 Segments 3001 to 3300 processed by UVM, noting subsurface objects on the left and right.
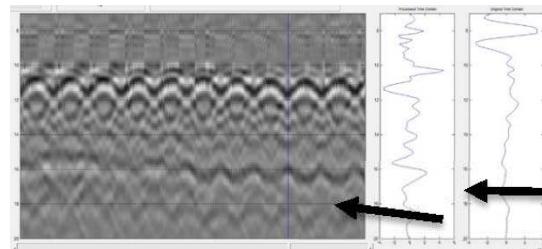
Figure 89 Segments 3001 to 3300 processed by UAF with AGC to see the subsurface trend.

The main feature for both groups in the segments between 3001 and 3300 is the fairly constant pattern along the 15 to 16 ns region. UVM notes that the reflections in this section are different from left to right. UAF identifies this trend in the data. No track information is known, but this is in an acceptable range and an expected pattern for the ballast-soil interface. This interface has been weak to unidentifiable in the radar scans for this track, so a strong reflection now may indicate fouling or at least a very moist region here. Knowing the intended depth of the ballast would allow calculating the depth of this line and help determine its significance.
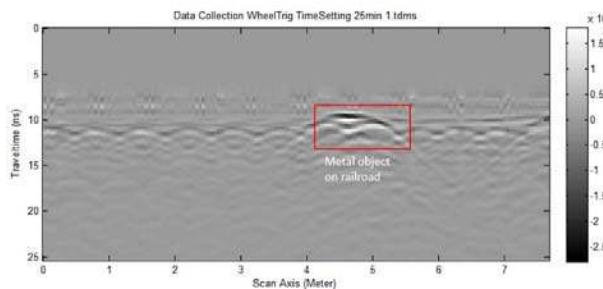
Figure 90 Segments 3301 to 3600 processed by UVM. A metal object on the track is highlighted.
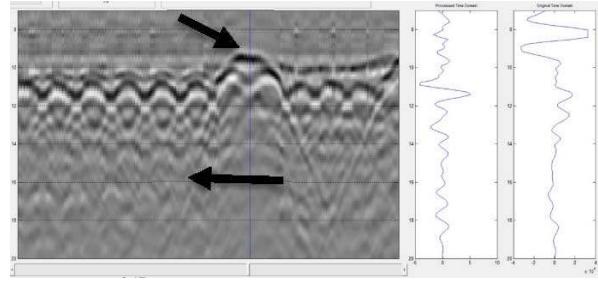
Figure 91 Segments 3301 to 3600 processed by UAF with AGC applied. UAF noted a strong reflection above the surface.

The final segments examined by both groups (3301 to 3600) only revealed what appears as a metallic object on the surface of the track. UAF applied automatic gain control to its scan and identified a pattern that might indicate a ballast-to-soil interface near the 16 ns line, though no information about the track was known.

4.5. Conclusion and Recommendations

Our objective of this study was to construct a GUI that would bring together filtering algorithms and quantitative methods for determining the depth of anomalies in the ballast layer. A basic set of filters were provided to allow band-pass filtering, background noise removal, anti-jittering, and automatic gain control. These filters were combined with a STFT feature to allow spectrogram analysis, as well as a custom tool to allow the capability to calculate depths within the ballast layer. These algorithms were bundled into a graphical user interface, which allows the user to view a radar scan segment and associated waterfall graphs for the individual traces simultaneously. Application of the filter effects can be observed immediately with this system, and data markers can be used to highlight the numerical values of points of interest. This information can then be used with the depth calculator to estimate depths of anomalies within the first layer, or for this application, the ballast layer of railroad track.

The completed system was compared with analysis performed by the UVM team, which was interested in automatically identifying anomalous areas. The comparison showed that the UAF system compared well against the UVM system in terms of providing radar scan images with which a user could use the data to calculate the depth of identified anomalies. When applied to field data, the system was capable of identifying the pertinent features of the ballast, which allows further analysis to be performed.

65

As with any research project, a little more work can always be done. This algorithm suite was implemented through software and designed specifically for the radar developed by UVM. The flexibility of software and the high degree of processing capabilities permit the possibility of modifying the program to be more general. A modification would involve implementing a control panel or a settings feature that would allow a user to specify the operational parameters of the radar system. Since low-pass and high-pass filters are established every time the function is called, this modification would allow greater flexibility with systems compatible with this one. Such a modification would also open the way for development of other filtering algorithms, which were not required for this project but would increase the robustness of the package.

An interesting graphical color representation of ballast was created toward the end of a study by Roberts et al. [47]. Recreating the method reported in that study would require a great amount of work, but would provide a way of viewing ballast conditions in color. Such an effort would need to work in tandem with the ability to obtain ground truth in the areas of interest for which the method is used. The method may add the ability to reconstruct the physical characteristics of the ballast by using change-point detection and analysis statistical methods.

CHAPTER 5
TRACK DYNAMIC MODEL

5.1. Introduction

Hanging tie (Figure 92Figure 92 Hanging tie []., also known as floating tie) is a railroad track distress when voids have been developed beneath ties due to uneven ballast settlement and improper maintenance practice. It in return will lead to an increase in dynamic impact loading on the top of the ballast and therefore further deteriorate the track structure. At the same time, due to the reduced support stiffness from ballast, the vibration level could increase exponentially and the passenger comfort level cannot be guaranteed. Nevertheless, it is often difficult to identify this problem in an efficient and non-destructive manner. In practice, hanging tie can only be detected by lifting up ties one by one, which requires a long maintenance time period and high maintenance cost.
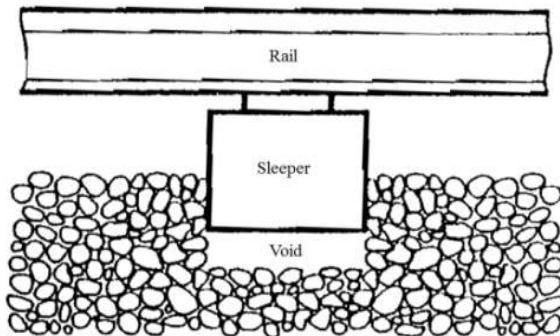
Figure 92 Hanging tie [*49*].
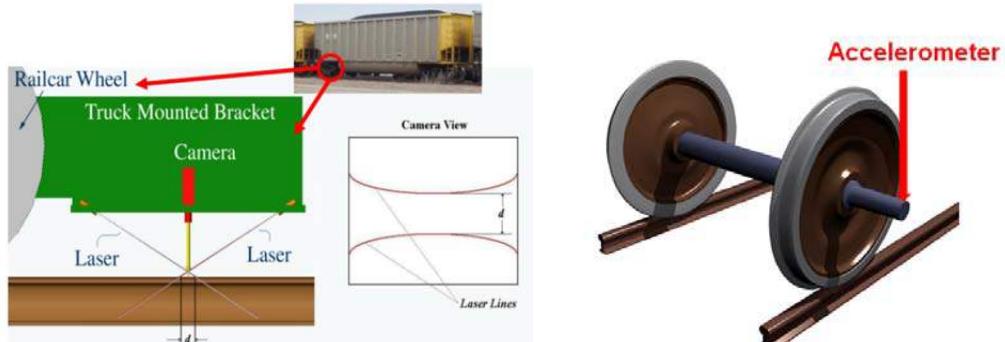
Figure 93 Moving track deflection recording devices.

In this report a fast and non-destructive method to identify hanging-tie problems is proposed. An effort of associating hanging tie identification to "Moving Track Deflection" has been made. "Moving Track Deflection" is defined herein as the vertical movement of the wheel rail contact point when the wheel moves. A homogeneous track should have "Moving Track Deflection" as a horizontal line. However, in reality, due to rail surface roughness, tie spacing, and other changes in track characteristics along the track, "Moving Track Deflection" will

always display fluctuations. Devices have been made to record moving track deflections. For instance, the one proposed by El-Sibaie et al. [*50*] (Figure 93) has two lasers and one camera. By measuring the width of the two laser beams projected on the rail, the moving deflection can be calculated. The other recording device which is implemented in this research is to install accelerometers on the moving wheels. The accelerometers can record the accelerations of all locations along the track as the wheels pass by. Then, one can get "Moving Track Deflection" by using double integration.

The field data has been recorded in Boston and St. Louis metro lines by an accelerometer installed on the steel wheel of a hi-rail SUV (similar to what is illustrated in Figure 93 on the right). The train speed was restricted to a very low range in this study (less than 10 meter/second). Also, the weight of the testing hi-rail SUV needs to be neither too heavy to close the gap between tie and ballast nor not too light to detect the hanging tie. To identify the hanging-tie problem and further determine the necessity of track maintenance, the results of "Moving Track Deflection" are interpreted by spectrum analysis. Due to the irregularities of the track system, it is believed that the spectrum change for "Moving Track Deflection" of a normal track will be different than the spectrum change for that of a track with hanging-tie(s).

To characterize tracks with/without hanging ties, a dynamic track model capable of calculating track deflections under moving load for different track configurations is introduced in this study. One of the first steps on the study of hanging-tie problem is the accurate modeling of tie spacing. Ties modeled as discrete masses are placed in succession in designated spacing. Another important aspect is to model the contacts of rail-tie and tie-ballast. These contacts are modeled as a spring-dashpot system, in which the support stiffness is represented by linear springs. The formulation of this numerical model is introduced in the following section.

5.2. Numerical Model

Figure 94 shows the dynamic track model proposed and implemented in this research. The arrow which denotes a point load describes the train loading. The point load is moving along the rail. The rail is modeled as an Euler beam. The rail pad, tie, and ballast layer are represented by means of a system of mass, spring, and damper with designated spacing. The soil part is modeled as a 3D finite element model. The moving load, track and subgrade soil are coupled in order to attain the responses of the whole system.
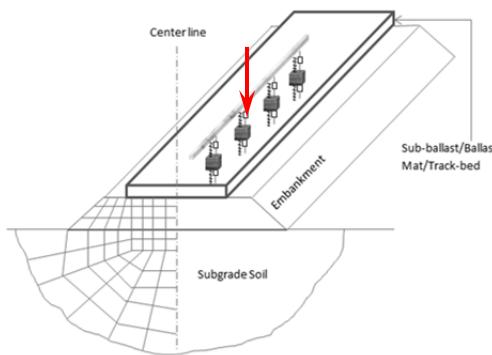
Figure 94 Sandwich track model coupled with subgrade soil.

The vertical train loads applied to the rail with time shifts are combined into one concentrated load $f(x,t)$. This solution is to create a constant point load occurring at different times and locations. If the time interval between the occurrences of two consecutive loads is small enough, a whole sequence of loads can be regarded as a moving load.

$$f(x, t) = f(t)\delta(x - vt) \tag{1}$$

where,

x = position;
v = velocity of trains;
δ = Dirac's delta; and
t = time.

For the rail, the governing equation can be written as:

$$EIU_r(x, t)'''' + \rho U_r(\ddot{x}, t) + \varepsilon U_r(\dot{x}, t) + TU_r(x, t)''$$
$$= f(t)\delta(x - vt) - \sum_m a_m(t)\delta(x - x_m) \tag{2}$$

where,

EI = bending stiffness of rail;
$U_r(x, t)$ = rail deflection as a function of time;
ρ = unit mass of rail;
v = damping of rail itself, which will be set to zero for convenience;
T = rail axial force caused by temperature increase;
$f(t)$ = wheel load function;
δ = delta function;
x_m = location of the mth tie;
v = wheel speed; and
$a_m(t)$ = compression force at the mth tie between rail and support as a function of time.

For the discrete support system,

$$a_m(t) = [U_r(x_m, t) - U_t(x_m, t)]K_p(m) + [\dot{U}_r(x_m, t) - \dot{U}_t(x_m, t)]D_p(m)$$

$$b_m(t) = [U_b(x_m, t) - U_a(x_m, t)]K_b(m) + [\dot{U}_b(x_m, t) - \dot{U}_a(x_m, t)]D_b(m)$$

$$M_t(m)\ddot{U}_t(x_m, t) = a_m(t) - [U_t(x_m, t) - U_b(x_m, t)]K_p(m)$$
$$+ [\dot{U}_t(x_m, t) - \dot{U}_b(x_m, t)]D_p(m)$$

$$M_b(m)\ddot{U}_b(x_m, t) = [U_t(x_m, t) - U_b(x_m, t)]K_p(m)$$
$$+ [\dot{U}_t(x_m, t) - \dot{U}_b(x_m, t)]D_p(m) - b_m(t) \tag{3}$$

where,

$a_m(t)$ = the compression force at the mth tie between rail and support as a function of time;
$b_m(t)$ = the compression force at the mth tie between support and soil foundation as a function of time;
$U_r(x_m, t), U_t(x_m, t), U_b(x_m, t), U_a(x_m, t)$ = are the rail/tie/ballast/soil surface

deflection at the mth tie as a function of time;

$K_p(m)$ and $K_b(m)$ = the stiffness of the mth pad and ballast, respectively;

$D_p(m)$ and $D_b(m)$ = the damping of the mth pad and ballast, respectively;

$M_t(m)$ = the mass of the mth tie; and

$M_b(m)$ = the equivalent mass of ballast underneath the mth tie.

For 3D soil, a conceptual equation can be set up as:

$$\{U_a(x_m, \omega)\} = [GS] * \{b_m(\omega)\} \tag{4}$$

Where $[GS]$ is the green function of soil and can be solved by the following equation:

$$([K_s] - (\omega - \lambda v)^2) * [M] * [\overline{dS}] = [\overline{F}] \tag{5}$$

For any frequency "ω", the soil surface deflection in frequency domain can be obtained by conducting discrete wave number Fourier Transform on Equation (5). The results will form the matrix $[GS]$ in Equation (4).

It can be seen that deflection of rail and soil surface at any position are functions of force vectors $\{a_m\}$ and $\{b_m\}$. Once these two vectors are solved other results such as rail deflection, tie deflection, tie-ballast force, stresses in the soil, etc. can be solved by using Equation (6).

$$\begin{pmatrix} AK_r + I & BK_r \\ GS * C & GS * D - I \end{pmatrix} \begin{Bmatrix} U_r(x_{1...m}, \omega) \\ U_a(x_{1...m}, \omega) \end{Bmatrix} = \begin{Bmatrix} F(x_{1...m}, \omega) \\ 0 \end{Bmatrix} \tag{6}$$

As results solved so far are in the frequency domain, Fast Fourier Transform (FFT) needs to be performed to transform the solutions back to time domain.

5.3. "Moving Track Deflection" Spectrum Study

Table 3 Track components properties

Track Component	Value
Rail bending rigidity	12 MN·m^2
Rail mass	130kg/m
Pad stiffness	280 MN/m
Ballast stiffness	70 MN/m
Pad damping	60 kN·sec/m
Ballast damping	80 kN·sec/m
Tie mass	280 kg
Ballast mass	300 kg

"Moving Track Deflection" is the vertical movement of the wheel rail contact point when the wheel moves. The model in this report contains a detailed track model and a moving load running on the rail surface. The setting of the calculation in this study is the rails resting on 50 ties. The spacing between two ties is 0.508 m, which is a typical value for wood crossties. The

total running length is 22 m. The train load takes place from left to right at a speed of 20 m/s and the running time is 0.30 s. The track model has the following characteristics in Table 3.

In Figure 95, the "Moving Track Deflections" as a function of time for the entire running length is shown. It corresponds to the condition when there is no missing tie. The periodic fluctuation implies the changes in tie support stiffness. The wave crests represent the moment that load is running between ties. The wave troughs represent when load is running on ties.
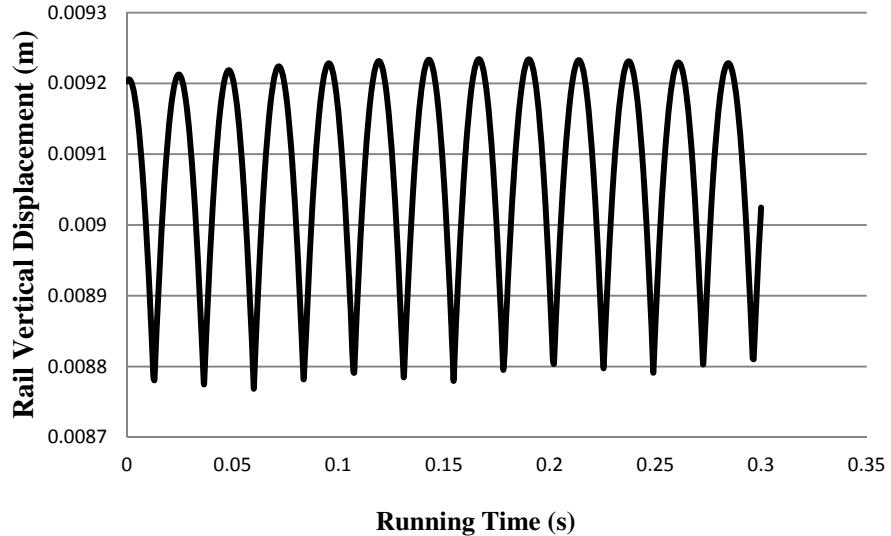
Figure 95 Rail vertical displacement contour for the entire length.

To explore the mechanism of the hanging-tie problem, Fast Fourier Transform (FFT) is performed to obtain the spectrum of the "Moving Track Deflection". Figure 96 shows the spectrum of no hanging ties. The maximum value occurs at the frequency around 40 Hz which corresponds to the tie spacing based on the relationship between train speed and tie spacing ($f = v/\lambda$=20/0.508=39 Hz). The remaining peaks may correspond to other track components.
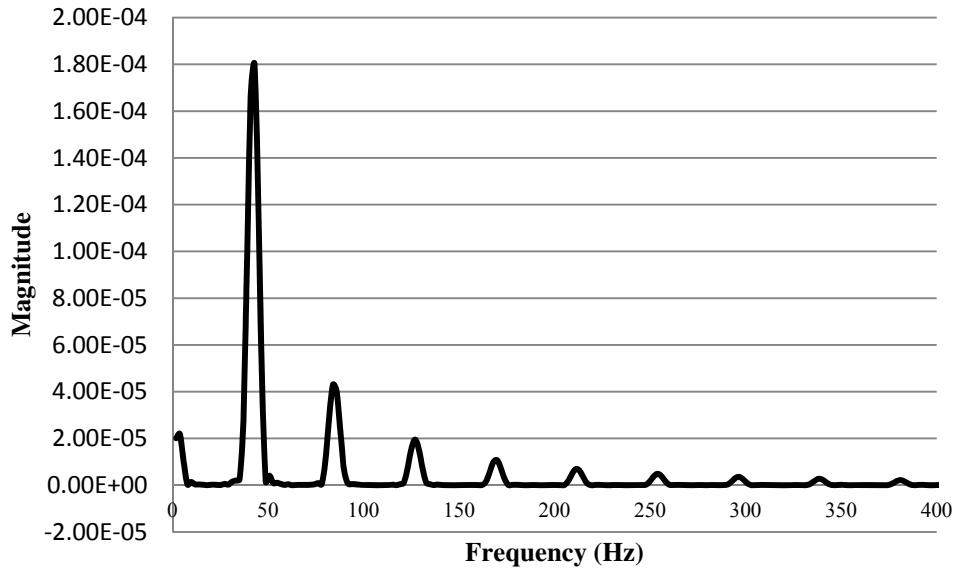
Figure 96 Rail surface deflection spectrum of no hanging ties.

To distinguish between tracks with and without hanging ties, the difference in spectrum for the hanging-tie situation is a pivotal issue in this study. With the definition of hanging ties, the ties are 'floating' on the ballast, resulting in less or even zero support for rails. Therefore, in order to simulate hanging ties, the stiffness of pad and ballast are set to very low values: 280 N/m and 70 N/m, respectively. For no hanging-tie conditions, the two values are 280 MN/m and 70 MN/m, respectively. In Figure 97, the time domain results for both hanging-tie and no hanging-tie situations are plotted. From the simulation results, the periodic responses associated with tie spacing can still be detected. However, because of the existence of hanging tie(s), a significant hump emerges in the plot for the hanging-tie situation. Moreover, the existence of hanging ties also increases the track deflection near the locations of hanging ties. The range of influence by hanging ties is about 5 to 6 tie spacing in this example.
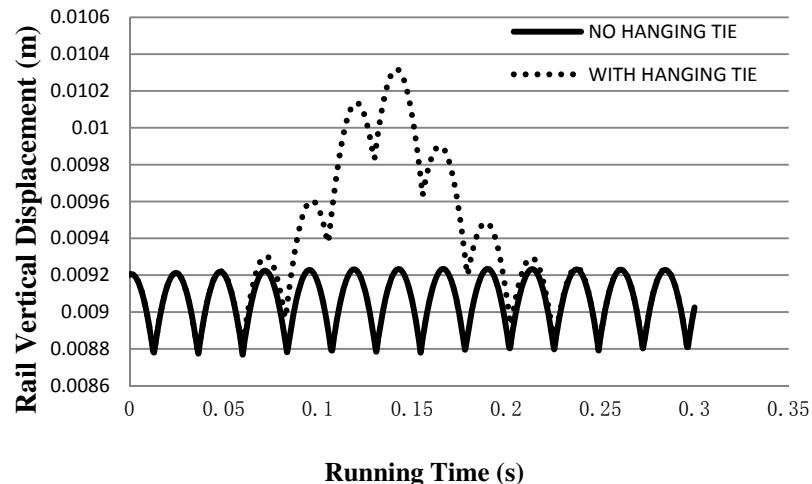
Figure 97 Rail surface deflections of hanging ties and no hanging ties.

72

However, the results in time domain are not able to show the track deterioration level and the interaction of different track components. Hence, FFT is adopted again to transform the rail vertical displacement results into frequency domain as shown in Figure 98. The spectrum generated as a result of hanging ties has a significantly different mode compared to the spectrum of no hanging ties. The highest peak moves to a much lower frequency region and the magnitude of the highest peak increases several times in contrast to the no hanging-tie situation. This result manifests that most of the energy flows to the low frequency due to the existence of hanging ties. This phenomenon indicates that the locations near hanging ties will have a much larger dynamic impact loading which could accelerate the deterioration of track. Meanwhile, the peak associated with tie spacing remains at the same spot (around 40 Hz). The same thing happens to other peaks that can be found in the spectrum of no hanging-tie conditions. The magnitudes of these peaks stay approximately the same under both conditions.
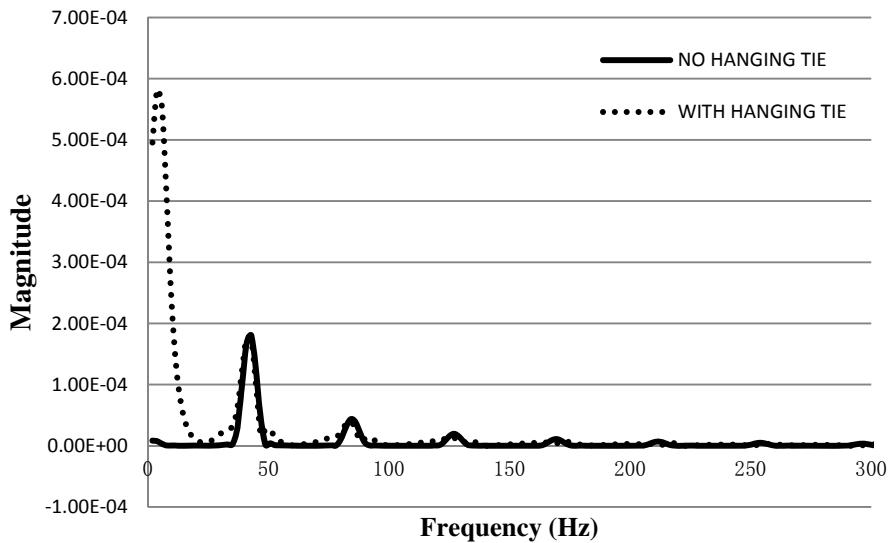
Figure 98 Rail surface deflection spectrum of hanging tie and no hanging ties.

To obtain the deterioration level of hanging ties, three different cases of tie support stiffness are set up in this study. In Figure 99, the blue line represents tracks without hanging ties; the red line with a distinct peak in low frequency region is for tracks with moderate hanging-tie issues; and the black line represents tracks with severe hanging-tie problems. As shown in Figure 99, tracks with a hanging tie all have a significant peak in low frequency region regardless of their degrees of deterioration, but the magnitude under the severe hanging-tie situation is twice larger than that under the moderate hanging-tie situation. In other words, the severity level of the hanging tie can be determined by the magnitude of the highest peak in low frequency region. A larger magnitude usually means the track has a severer hanging-tie problem.

73
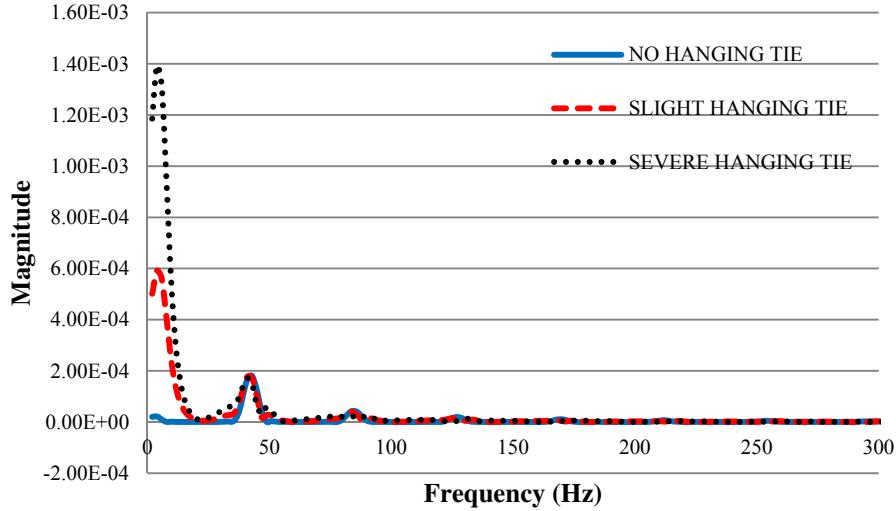
Figure 99 Different cases of tie support stiffness.

The previous simulation results (Figure 99) are only based on one hanging tie. In the real world, there could be multiple hanging ties spread over a track segment. Also, there can be many different combinations for the distribution of hanging ties. It is unrealistic to simulate all combinations. Therefore, for the purpose of characterizing different hanging-tie situations, five typical scenarios are considered here: (1) no hanging ties; (2) one hanging tie; (3) two adjacent hanging ties; (4) three consecutive hanging ties; and (5) two non-adjacent hanging ties. The first two are the footstones of this research and are already shown in Figure 99. As concluded previously, the vicinity of a hanging tie has a relatively large impact loading. In return, the large impact loading could further damage the track system near the hanging tie. Due to this, consecutive hanging ties are often found in the field. Therefore, here we focus on two adjacent hanging ties and three consecutive hanging ties.

The following figures show the moving track deflections of these five situations. Figure 100 shows the results of the first four scenarios. The "Moving Track Deflection" has the largest displacement for the scenario with three consecutive hanging ties. The scenario with two adjacent hanging ties results in the second largest displacement. The curve with no significant humps is for the scenario without hanging ties.
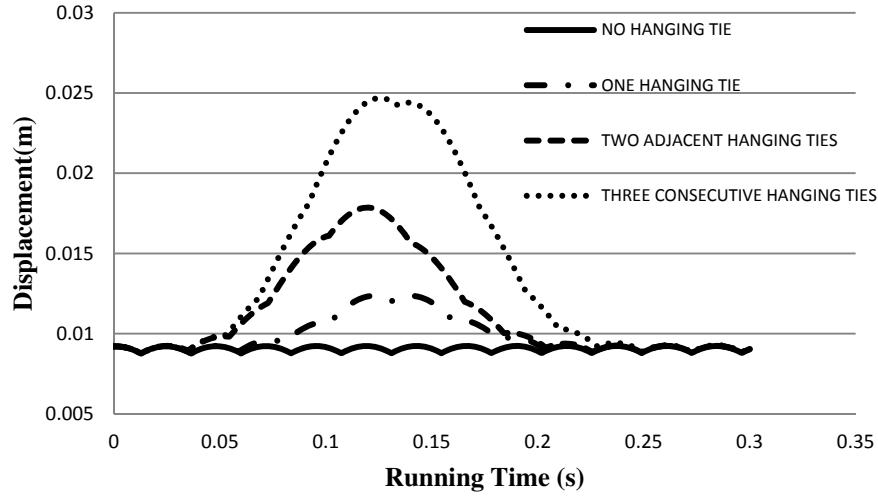
Figure 100 "Moving track deflection" contours for different cases of hanging ties.

Figure 101 shows the "Moving Track Deflection" spectrums of three scenarios of hanging ties. The spectrums of all three scenarios share a common characteristic: a significant response in low frequency region (around 4-10 Hz). The difference among them is the magnitude of the peak response. For one hanging tie, the magnitude of the peak response is the smallest; for the scenario with three consecutive hanging ties, its magnitude is the largest. Larger magnitudes suggest more energy (or larger dynamic impact) goes into area(s) near the hanging tie(s), causing further damage. Also, more hanging ties mean weaker support to the rail and lead to larger downward displacements. When this happens, passengers' comfort level can be significantly impacted.
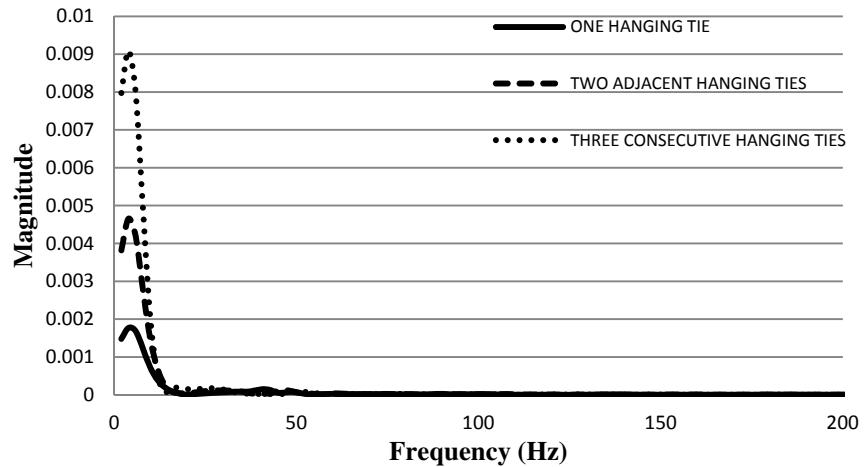
Figure 101 Rail surface deflection spectrums of different cases of hanging ties.

Hanging ties located separately are also common in the field. Figure 102 shows a simulated case for two non-adjacent hanging ties. The spacing between the two hanging ties is 2.5 m, which is equivalent to the space for around five ties. As shown in Figure 102, this case results in two humps in the moving track deflection contour.
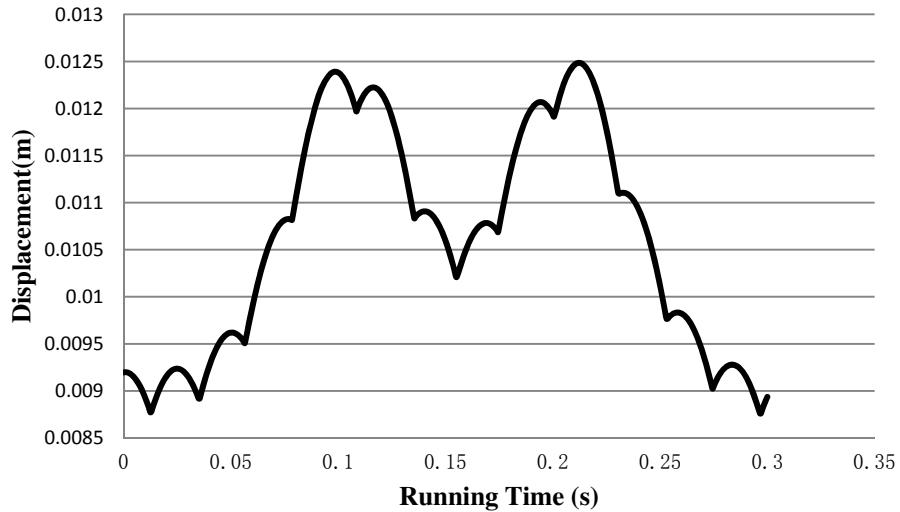
Figure 102 "Moving track deflection" contour for two non-adjacent hanging ties.

Using FFT again, the spectrum of the two non-adjacent hanging ties case is generated and plotted in Figure 103. Compared to the result for one hanging tie, this spectrum has two peaks in the low frequency region.
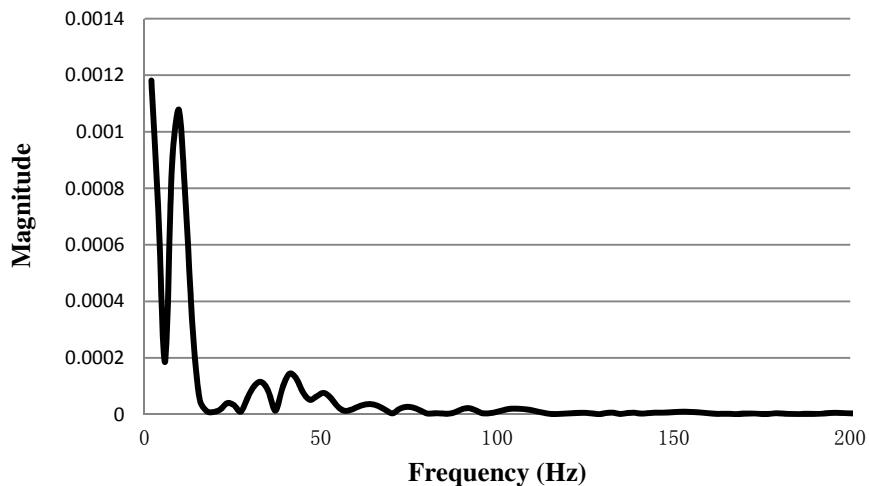
Figure 103 Rail surface deflection spectrums of two non-adjacent hanging ties.

According to the previous spectrum analysis, it is obvious that the spectrums for scenarios with hanging ties are significantly different from the one without hanging ties. Tracks without hanging ties only have small peaks associated with track components (in relatively high frequency regions), such as tie spacing. However, tracks with hanging ties have more significant peaks in lower frequency regions (around 5-10 Hz). In general, the severity level of hanging-tie problem can be determined based on the magnitude of the highest peak in spectrum analysis. The larger the magnitude, the more severe the hanging-tie problem is. Also, the magnitudes of the peaks in spectrum analysis seem to increase as the number of hanging ties increases.

5.4. Field Test Results and Case Study

In this section, field data sets collected from Boston and St. Louis are analyzed. In the field, an accelerometer was installed to record the vertical acceleration of the center of a steel wheel when the vehicle moves. Figure 104 shows a segment of the acceleration data collected in the field. The X-axis is time and the Y-axis is the steel wheel's center acceleration. The corresponding displacement plot can be calculated by implementing double integration of time.
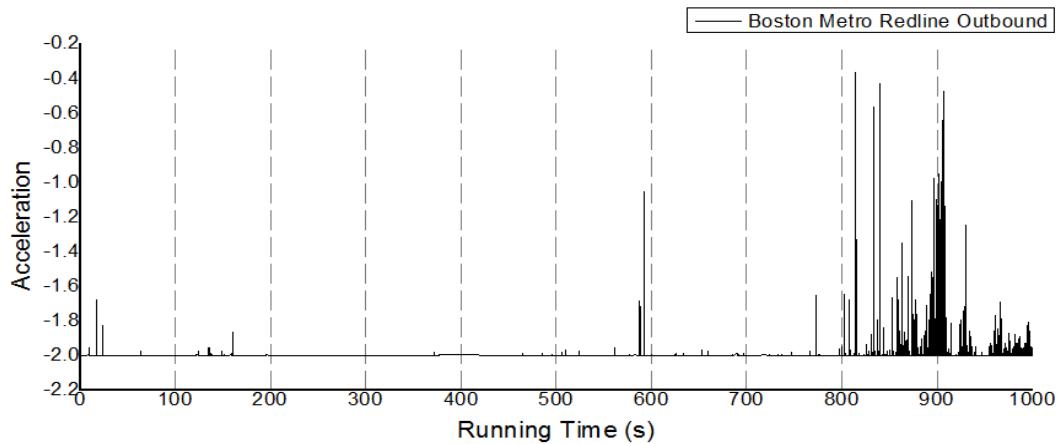
Figure 104 Acceleration of rail surface at Boston Redline outbound.
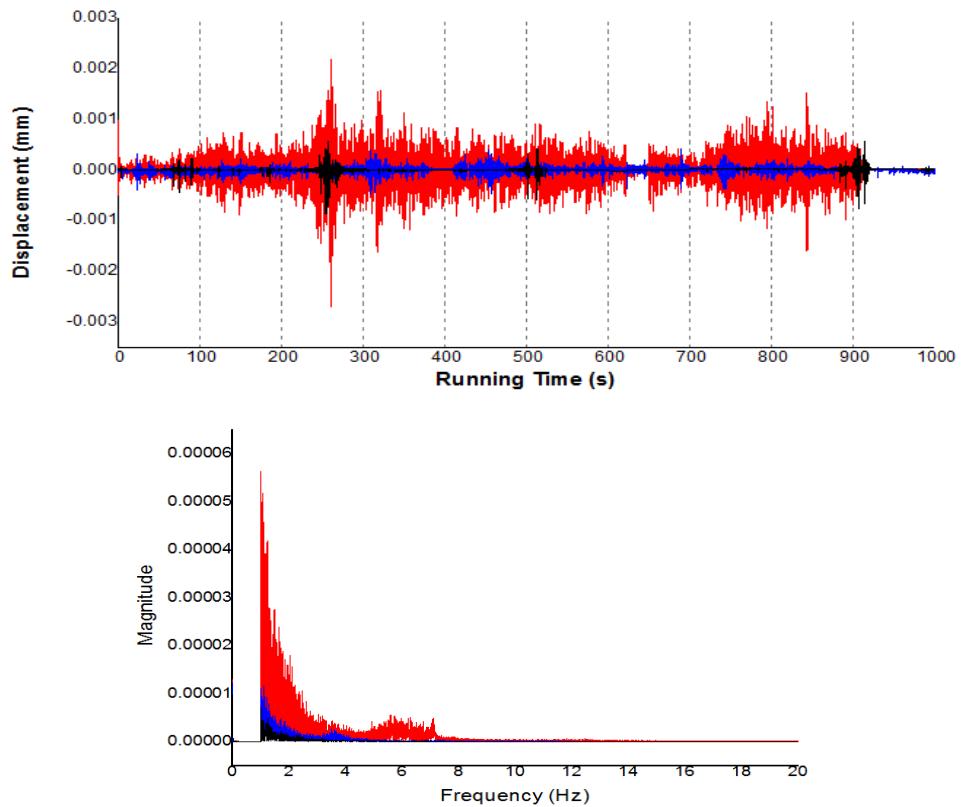
Figure 105 Spectrum analysis of the field data.

Figure 105 shows three typical scenarios of hanging-tie problems in time and frequency domains. The signal in red is for the Scott Air Force Base Westbound in St. Louis; the signal in blue is for Cross-County Westbound in St. Louis; and the signal in black is for MBTA (Massachusetts Bay Transportation Authority) Red Line Inbound in Boston. These three lines demonstrate different levels of hanging-tie problems. The top plot in Figure 105 shows the displacement of rail surface in time domain and the bottom plot is the corresponding result in frequency domain.

The purpose of this task is to establish a fast and non-destructive detection method to locate hanging ties. According to Figure 105 and the conclusions in the previous section, the red signal (Scott Air Force Base Westbound) has the worst hanging-tie problem, while the black signal has the best track condition among the three scenarios. However, in practice such conclusions must be drawn with caution due to the following reasons. First, the track stiffness during the simulation is considered homogeneous and does not change along the track, while in the field test the track is not a homogeneous structure. The amplitude is not only affected by the train speed and hanging-tie conditions, but also affected by the subgrade, depth of ballast and sub-ballast and so on. A track without hanging ties but with low stiffness could also have large amplitudes in low frequency region during spectrum analysis. This is because more energy is dissipated into track structure. Therefore, large amplitudes in spectrum alone cannot be used to determine hanging ties. Second, the field data could have much more uncertainty (i.e., noise) than the simulation data. Third, a key to successfully detecting hanging ties is to identify the responses associated with train speed and tie spacing in spectrum analysis. If such responses demonstrate a clear and regular pattern, the track is usually in good condition. If the track has hanging-tie problems, the spectrum curve can change drastically and make it difficult to see the pattern for the responses associated with train speed and tie spacing.

The proposed track dynamic model and the spectrum analysis are applied to additional data collected from Scott Air Force Base Westbound (SAFBW), Cross-County Westbound (CCW), and MBTA Redline Inbound (RI). The data from each line is segmented into nine segments. Each segment represents approximately 100s of running time. It was difficult to maintain a constant speed during the data collection. For the field data collections in Boston and St. Louis, the average hi-rail vehicle speeds were between 3.5 and 5.75 mph. Therefore, the responses associated with train speed and tie spacing should be around 5 Hz. The spectrum analysis results for the three rail lines are shown in Figure 106 through Figure 108.
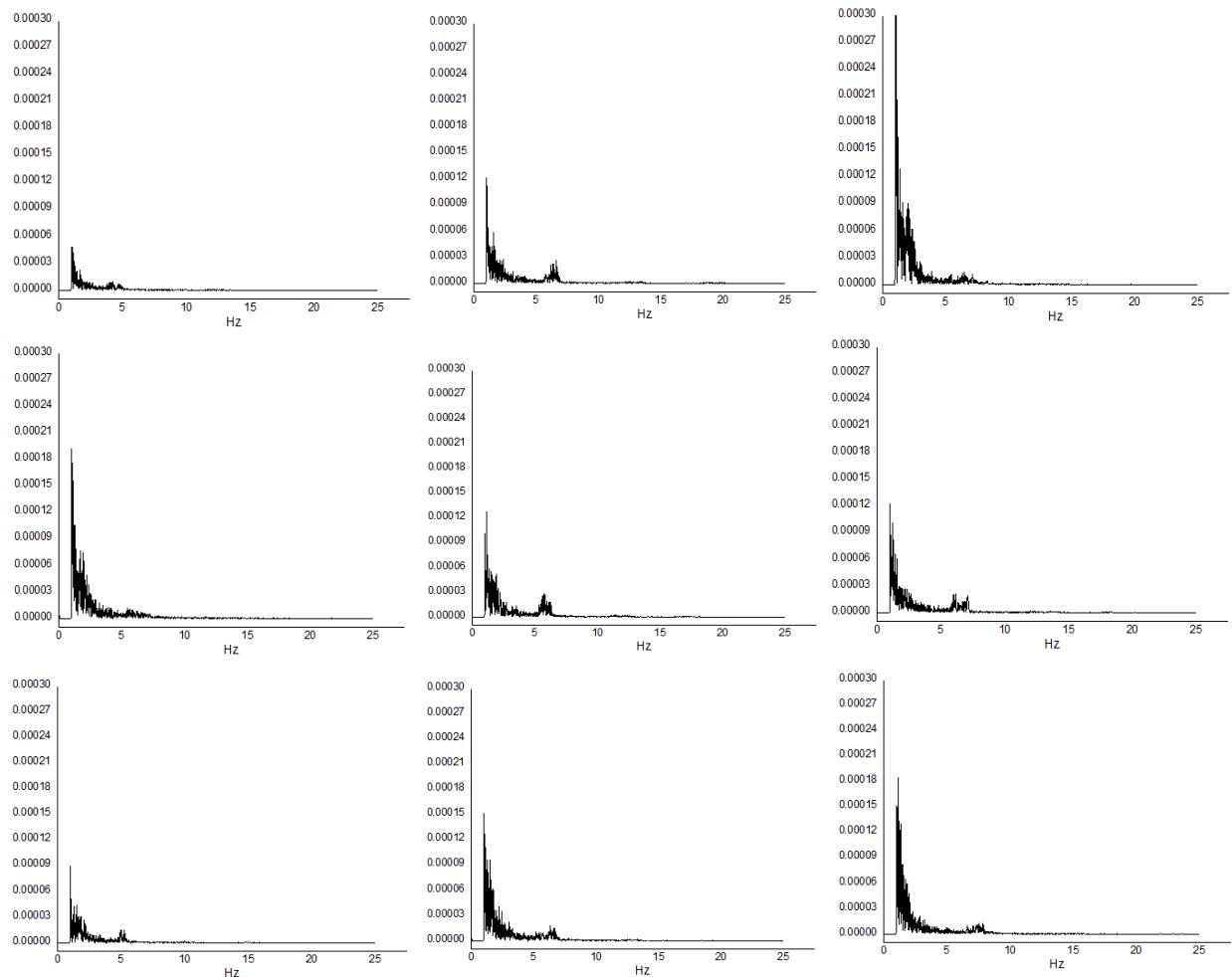
Figure 106 Spectrum analysis results for Scott Air Force Base Westbound.
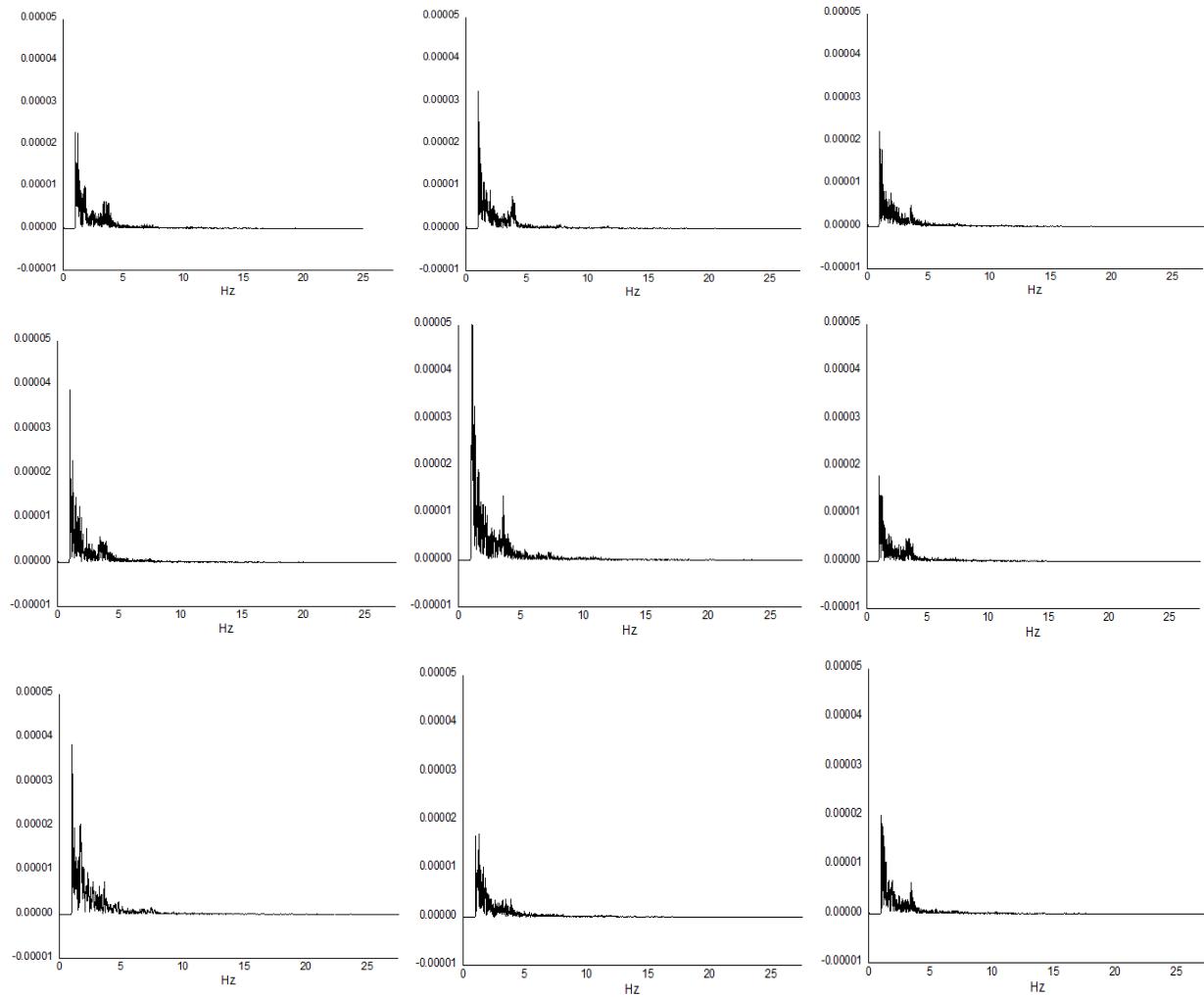
Figure 107 Spectrum analysis results for Cross-County Westbound.
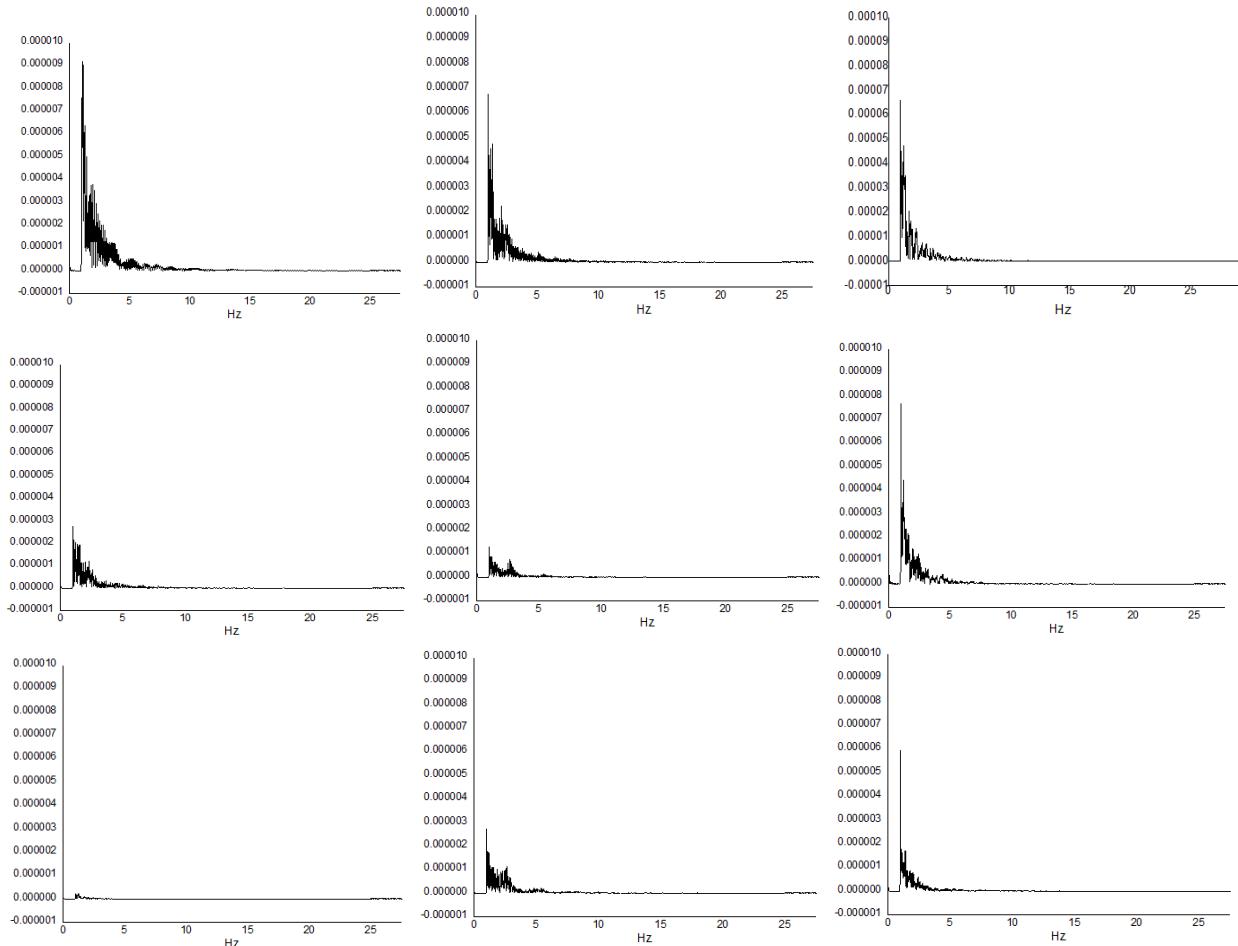
Figure 108 Spectrum analysis results for MBTA Red Line Inbound.

From the spectrums above, the responses associated with train speed and tie spacing can be detected for most segments of SAFBW and CCW. Only one or two segments do not have significant responses associated with train speed and tie spacing. However, it is hard to detect the responses in segments from MBTA RI. This suggests that RI has the severest hanging-tie problem. In the spectrums of SAFBW and CCW, one can still detect major responses in low frequency regions. These could be caused by noise in the field. Although the noise is not significant enough to blur the peak response generated by tie spacing, it can lead to the wrong conclusion of the existence of hanging ties. However, the spectrums for field data with hanging ties should have the shapes shown in Figure 108, in which the curves are smooth and it is difficult to detect the peak responses generated by tie spacing.

To better compare the spectrum analysis results for the three rail lines, the values corresponding to their largest peaks of the nine segments are plotted in Figure 109. Even though the spectrums show that SAFBW has the largest amplitudes while RI has the smallest ones, RI actually has the worst hanging-tie problem.
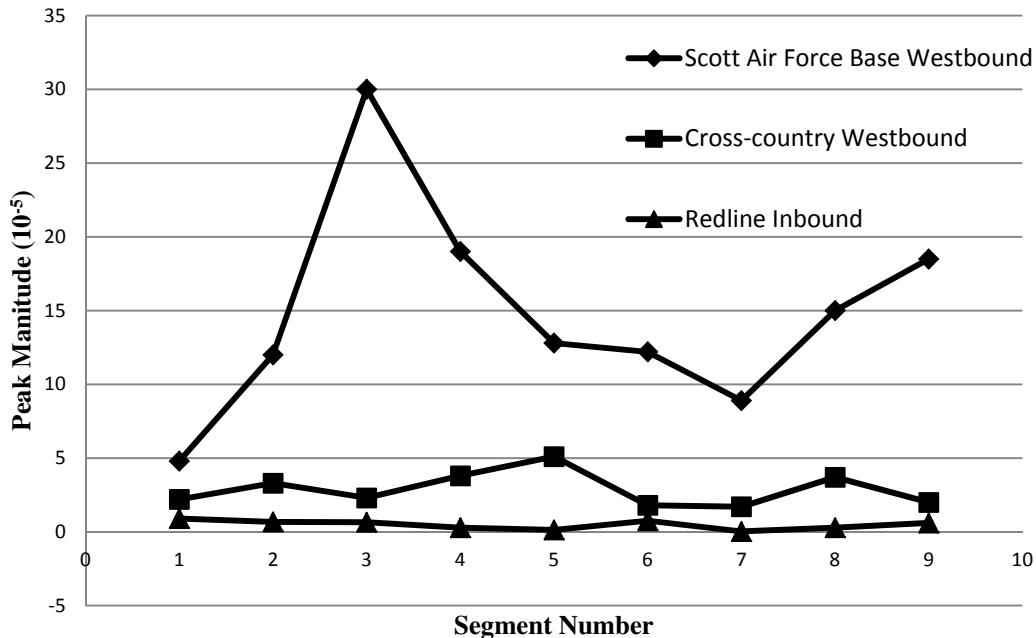
Figure 109 Comparison of deterioration level.

Additionally, according to Figure 109, Segment 3 of SAFBW has no significant responses associated with train speed and tie spacing, which implies that this segment may have the worst track condition among all segments of this line. To identify the locations of hanging ties for this specific segment, it is further divided into ten sections. Each section represents about 10s of vehicle running time. Following the same analysis procedure, the spectrums of these ten sections are plotted in Figure 110. The largest peaks for each section are plotted in Figure 111. From the results in Figure 110, it is obvious that Sections 7, 8, and 9 have significant fluctuations and it is difficult to identify the patterns for the responses associated with train speed and tie spacing, suggesting that these track sections may have severe hanging-tie problems.
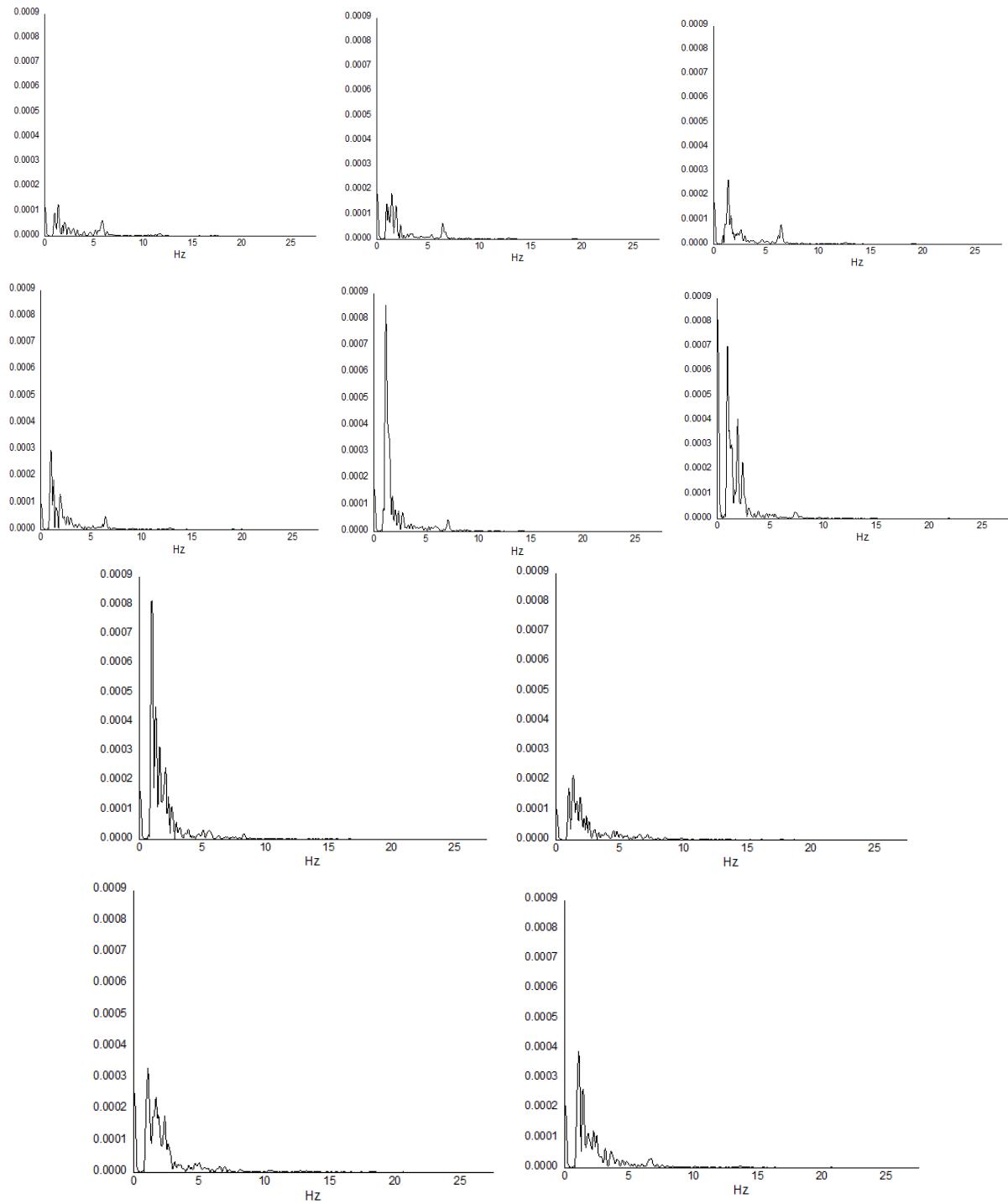
Figure 110 Spectrum analysis of Segment 3 from Scott Air Force Base Westbound.
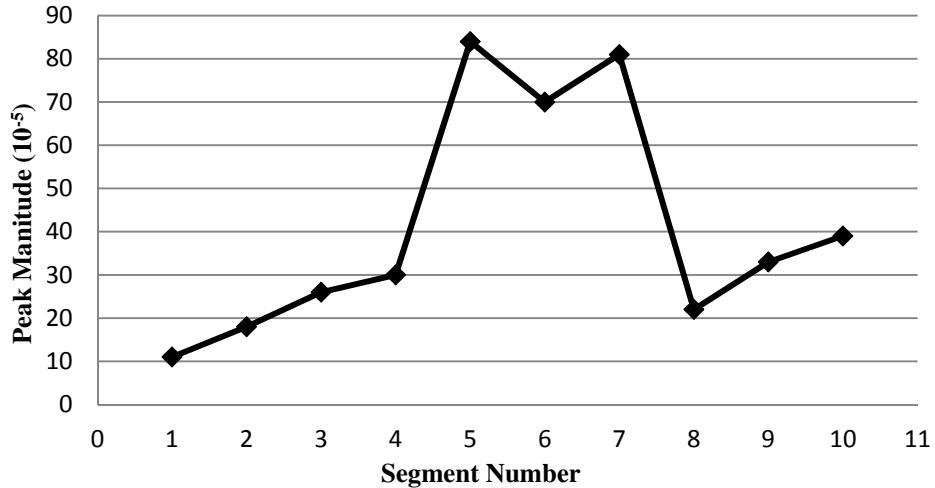
Figure 111 Comparison on deterioration level for SAFBW Segment 3.

5.5. Conclusions

In this chapter, the vertical displacement of rail induced by a moving point load is investigated. A track dynamic model is proposed for this purpose. By varying the stiffness of pad and ballast, rail vertical displacements with and without hanging ties are simulated by this model. Using the Fast Fourier Transform, spectrum analyses are conducted on the simulated data. According to the spectrum analysis results, it is found that the spectrums for track segments with hanging ties have shapes that are significantly different from those for segments without hanging ties. From the spectrums of the simulation data, tracks without hanging ties only have frequency responses associated with track components, such as tie spacing, while tracks with hanging ties have significantly larger responses in low frequency regions.

Rail vertical displacement data was also collected in the field from Boston and St. Louis. For the field data, it was found that the amplitudes of frequency responses alone are not enough to identify hanging ties. This is because noise in the field data can also generate large frequency responses in low frequency regions. Hence, an improved method was proposed to identify hanging ties. Based on this method, a track segment is considered to have hanging tie(s) if no responses associated with train speed and tie spacing can be detected in its spectrum of vertical displacement data. After analyzing the field data collected from Boston and St. Louis, the MBTA RI line was found to have the severest hanging-tie problem.

5.6. Future Research

In this research, the spectrums of field and simulated vertical displacement data for track segments with hanging ties in general match well. The spectrums for segments with hanging ties are significantly different those for segments without hanging ties. However, during the field data collection, important inputs to the track dynamic model such as soil properties and geometry were not collected. Therefore, we can only conclude that there is a significant difference in spectrums of tracks with and without hanging ties. An important next step is to revisit the data collection sites and verify the identified hanging-tie problems. Also, additional field data (e.g., soil properties and geometry) needs to be obtained to calibrate the track dynamic model.

CHAPTER 6

SYSTEM INTEGRATION AND FIELD IMPLEMENTATION

6.1. System Integration

The developed subsystems have been integrated and installed on a hi-rail vehicle. Since the subsystems were developed by different groups in the team, they were designed to be able to work relatively independently. In this way, the testing of one subsystem will not require the presence of another subsystem. The entire system consists of four major components: laser subsystem, GPR subsystem, track dynamic subsystem, and decision support subsystem (discussed in Chapter 6). The first three subsystems all involve both hardware and software development. Since each of these three subsystems is going to generate a significant amount of data, they are connected to separate computers for data collection, processing, and storage. This is to ensure that all data can be processed and logged without delay or being missed. If a more powerful computer with a large-volume high-speed hard drive (e.g., solid state drive) is available, the three subsystems can be all connected to one computer. Their control and data acquisition software can also be integrated.

Although the laser and GPR subsystems have their control and data acquisition software installed on different computers, they do share the input of a wheel encoder. The encoder was installed on a steel wheel of the hi-rail vehicle we rented as shown in Figure 112 below. The signal generated by the encoder triggers both GPR and laser scans. Also, the distance information generated by the encoder is used to tag the collected GPR data. The collected laser data is geocoded using both GPS and the wheel encoder. Each laser file has a GPS coordinate and is about 2 meters long. Given the distance information, the collected laser and GPR data can be correlated. In this way, defects or suspicious features found in one data set can be compared against features identified in another data set.

Figure 112 Installation of wheel encoder and accelerometer.

The only hardware for the track dynamic system is a 3-axial accelerometer, which is meant to measure track deflection. Since the accelerometer keeps track of the hi-rail vehicle's longitudinal acceleration, our plan is to calculate distance travelled by performing a double

integral on the longitudinal acceleration data. In this way, the collected track deflection data can be compared with the laser and GPR data. Additionally, the time stamps of accelerometer and laser data can be compared to confirm the locations of different accelerometer data points. In future studies, the encoder signal can be fed into the track dynamic system for geocoding the accelerometer data. This additional redundancy can help precisely match laser, GPR, and accelerometer (track deflection) data.

6.2. Field Implementation

The team rented a hi-rail SUV from Colvins Truck Rental in Waltham, Massachusetts. This study requires us to make a lot of changes to the hi-rail vehicle to install mounts for the GPR, laser, and track dynamic subsystems. The PI is extremely grateful to Colvins Truck Rental. Without their permission and assistance to modify the SUV, it would be impossible for us to complete the field data collection as scheduled.

Once the hi-rail vehicle was identified, the PI and two graduate students measured the dimension of the vehicles and developed different schemes to mount various subsystems. After numerous discussions among the project team, a final design in the figure below was selected.
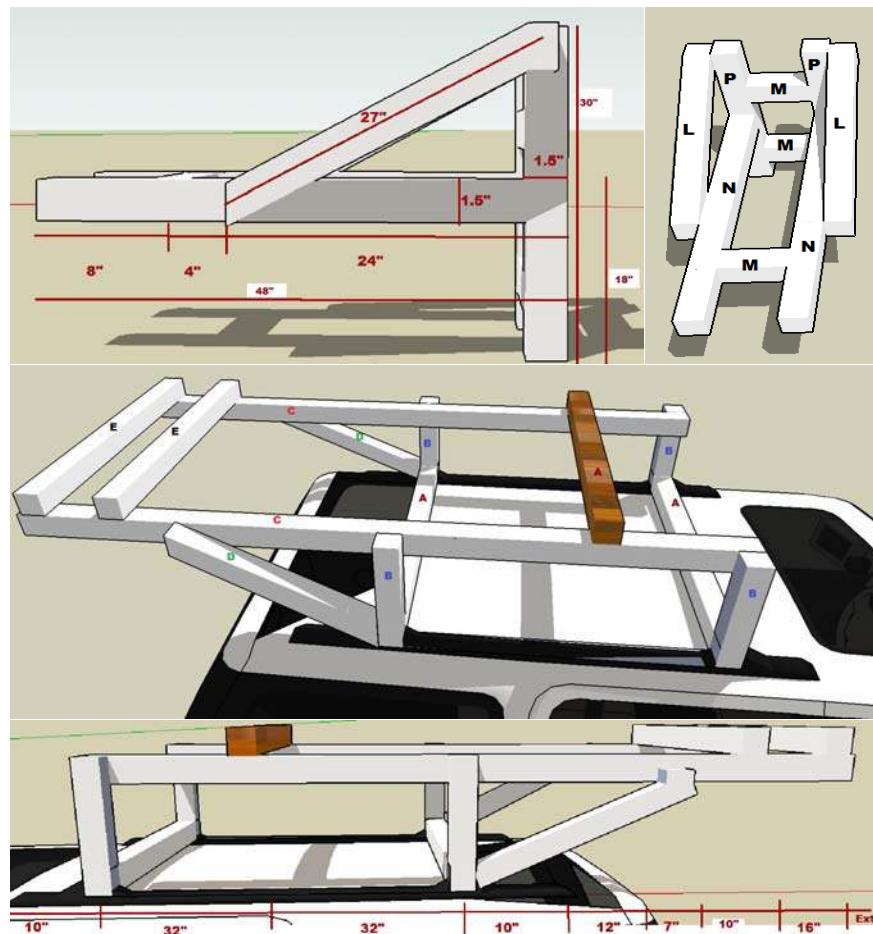
Figure 113 Mounts for laser and GPR subsystems.

86

In August, 2013, CodeRed coordinated the Right-of-Way (ROW) Safety Training with MBTA staff for the project team. Project team members designated to collect the laser and GPR data along the MBTA Green and Red Lines attended the ROW Safety Training at Cabot Yard Training Facility. On August 15th, 2013, the project team members along with some MBTA staff arrived at the Blandford grade crossing (see Figure 114) on Commonwealth Avenue in Boston for the first data collection. CodeRed coordinated with MBTA staff the ROW access and walked the segment of westbound track to conduct track gauge measurements approximately every 100 ft. The hi-rail vehicle accessed the ROW to begin collecting the laser and GPR data from Blandford grade crossing to Summit Street for approximately 6 miles traveling westbound. On August 16th, 2013, the project team and MBTA staff arrived at Central Ave (see Figure 115) grade crossing along the Mattapan High Speed Line. The team collected laser and GPR data from Mattapan Station to Cedar Grove Station for approximately 3 miles traveling northbound. CodeRed assisted the project team in conducting the track gauge measurements every 100 ft and the collection of sample ballast for approximately 1.5 miles from Central Ave Station to Valley Road Station. To avoid interrupting revenue services, all data collections were conducted between 12AM and 4AM.

Figure 114 Data collection on MBTA Green line at Blandford grade crossing.

Figure 115 Data collection on MBTA Red line at Central Ave station.

In August 2013, CodeRed coordinated the Right-of-Way (ROW) Safety Training with St. Louis Metro staff for the project team. On August 19th, 2013, the project team members designated to collect laser and GPR data attended the ROW Safety Training at the Ewing Maintenance Facility. From August 19th to August 20th, CodeRed assisted the project team with collecting laser and GPR data along Metro's ROW from Shiloh-Scott Station to Belleville Station for approximately 6 miles. Also, CodeRed coordinated with Metro staff the ROW access. Metro St. Louis provided a separate hi-rail vehicle for the team to ride and conduct track gauge measurements every 100 ft. On August 21st, 2013, CodeRed assisted the project team with collecting laser and GPR data from Forest Park Station to Sunnen Station along the Cross County extension for approximately 6 miles. The westbound data collection ended near Big Bend Station due to the test hi-rail vehicle's loss of battery power that supported the equipment.

Figure 116 Data collection in St. Louis from Shiloh-Scott Station to Belleville Station.

Figure 117 Data collection in St. Louis Forest Park Station to Sunnen Station.

Overall, the data collections were successful. In total, we collected about 300 GB GPR data and 200 GB laser data. Since we used 2 desktops and 1 laptop for data acquisition, we had to draw power from both the engine and two batteries. Although we ran out of power during the second data collection in St. Louis, these collected data sets are already enough for us to test the developed laser, GPR, and track dynamic models.

One flaw with the data collection plan was the measurement of rail gauge for validating the laser system. We measured rail gauge at certain intervals and was hoping to compare these data sets with the laser results. However, it turned out that finding the corresponding rail gauge measurement locations in the laser data precisely is very difficult. Therefore, it is difficult to compare the field gauge measurements with the laser system outputs. To address this issue, the team borrowed two rail segments from MBTA and conducted rail gauge measurements in a lab. The lab measurements were compared with the laser system results to validate its accuracy. Our lab test results suggested that the laser gauge measurement error was less than 1mm. Since the laser system's accuracy is usually better when it is in motion, the actual field gauge measurement error should be smaller than 1mm.

CHAPTER 7

DECISION SUPPORT SYSTEMS

7.1. Introduction and Scope

This chapter describes the development of a web-based Geographic Information System (GIS) used to map the collected laser and GPR data. Its main objective is to visualize and enable users to query the collected raw GPR and laser data. Algorithms developed for processing the GPR, laser, and track dynamic models will be used to identify missing fasteners, broken ties, wide gauge, fouled ballast, suspicious underground features, and hanging ties. These processed data will be loaded into the decision support system as images. Quantitative information will be provided so that the locations of these defects can be easily queried. This system will also display passenger rail lines and their stations.

With this GIS-based Decision Support System (DSS), different departments of a rail transit agency can have easy access to the conditions of their rail assets. Senior managers can directly generate reports on their asset conditions without going through different offices. Combining data from various sources (i.e., laser, GPR, and track dynamic model) can allow users to find potential correlations among different defects. Additionally, data generated by different inspection scans can be compared to see if there are major changes to the track over time.

We have also built an accompanying mobile application for the DSS, which is a handy tool for field inspection trips. Some companies have developed handheld devices for track inspections and maintenance activity scheduling. For example, ENSCO has developed a *Digital Track Notebook*. We believe that a mobile App is a better choice than those handheld devices. A mobile App can be installed on different mobile devices (e.g., smartphones and tablets). Most mobile devices are equipped with high-resolution cameras, GPS, and communication capabilities. Inspectors in the field can take pictures of track defects, complete and upload inspection reports, pull out historical inspection reports, and locate nearby defects. These features have all been implemented in the developed mobile App. In the future, advanced features can be added to the mobile App, for example, generating inspection trip schedules and adding them to inspectors' smartphone calendars automatically.

For handheld inspection devices, companies need to maintain both the hardware and software. Mobile App users do not need to worry about hardware development or upgrade and can simply focus on improving the mobile App software. Also, most people today always carry a smartphone and this can help lower the inspection cost for purchasing hardware. By using a mobile App, we can synchronize its data across multiple devices rather than a special handheld device. Also, the entire inspection report can be completed in the field and uploaded to a central database in real time. Another interesting feature to add is to save the collected data on the mobile device and upload them to the central database when a secure Wi-Fi network is available. Similarly, inspectors can download the historical records of sites to be inspected onto the mobile device and review them in the field. These two ideas have not been fully implemented in the developed mobile App yet due to time and resource constraints. However, they may significantly reduce an agency's data communication costs.

We demonstrated this mobile App idea to several rail transit agencies (including MBTA and Metro St. Louis). They showed strong interest in it. The PI is currently working with CodeRed on securing a contract with Metro St. Louis to develop a customized mobile App for them and to integrate the mobile App with their IT system. As part of this effort, the team will expand the GIS-based DSS for Metro St. Louis and add features to facilitate the scheduling of their maintenance and capital project activities.

7.2. Technologies

The development of the GIS-based DSS involved the following tools and technologies: Microsoft SQL Server, ArcGIS for Desktop, ArcGIS for Server, ArcGIS API for JavaScript, and PhoneGap. The applications of these tools and technologies are described in detail below.

7.2.1. Microsoft SQL Server

Microsoft SQL Server was used as the main back-end to store all the project data, which mainly includes spatial data (e.g., train stations, rail lines, and locations of maintenance activities), attribute data (e.g., asset conditions, built year, and historical inspection records), and images (e.g., raw B-scan images, processed B-scan images, laser images, and output images of track dynamic model). There are several reasons for choosing Microsoft SQL Server for this project. First of all, it works well with the ArcSDE (Spatial Database Engine) technology, which is a critical component of ArcGIS for Server. ArcSDE serves as a gateway between GIS clients and Relational Database Management Systems (RDBMS); secondly, it allows us to store all project related data in a single database. This makes is easier to manage and query the data; thirdly, it allows multiple user to access the database simultaneously. This is a basic but important requirement that the developed DSS must satisfy; and last but not least, this project generated a huge amount of data. Regular geodatabases such as personal geodatabase cannot handle such a large data set.

7.2.2. ArcGIS for Desktop

Table 4 Example attribute table for a rail station feature class

BlueLine

OBJECTID *	LayerName	StationFrom	StationTo	BuildYear	SHAPE *	SHAPE.STLength()
1	Blue Line	Fairview Heights Station	Washington Park Station	1994	Polyline	4814.849623
2	Blue Line	Washington Park Station	Jackie Joyner-Kersee Stn	1994	Polyline	3290.090119
3	Blue Line	Jackie Joyner-Kersee Stn	Emerson Park Station	1994	Polyline	1477.020372
4	Blue Line	Emerson Park Station	5th & Missouri Station	1994	Polyline	2594.179585
5	Blue Line	5th & Missouri Station	East Riverfront Station	1994	Polyline	1944.648932

Adding spatial attributes and Structured Query Language (SQL) support for them to Microsoft SQL Server alone is not enough to support GIS. Thus, we also used ArcGIS for Desktop in this project. Using ArcGIS for Desktop, spatial data is usually organized as feature classes. Only features of the same type can be included in one feature class. For example, point and polyline features cannot be included in the same feature class. Each feature class is often

associated with an attribute table as shown in Table 4. Each row of this table represents one feature. The contents of this table can be accessed through SQL.

ArcGIS for Desktop was employed to manage the spatial and attribute data stored in the Microsoft SQL Server database. For example, we used it to create layers from the Microsoft SQL Server data sets, visualize data on a map, assign appropriate symbols to features, and to manipulate/edit features. Using ArcGIS for Desktop and ArcGIS online imagery services, we were able to digitize the rail lines and train stations for Metro St. Louis and saved them in two separate feature classes.

Based on the initial rail line feature class, several additional line feature classes were created. In these new line feature classes, line features were divided into equal segments that are either 2-meter or 10-meter long. This is because the laser subsystem generates output files that represent 2-meter rail segments, while the GPR subsystem generates output files that represent 10-meter rail segments. By matching feature lengths with the generated output files, it would be easier to attach the raw and processed data to features in the rail line feature classes and visualize the inspection results. It is also possible to explore other ways to visualize the data. For example, we may use linear referencing or dynamic segmentation to associate the laser and GPR data with the rail line features. However, this would require additional effort to process the laser and GPR data. For example, we may need to stitch some homogeneous laser or GPR data segments together. For this particular project, we chose to divide features into equal segments, which is simple and straightforward.

7.2.3. ArcGIS for Server
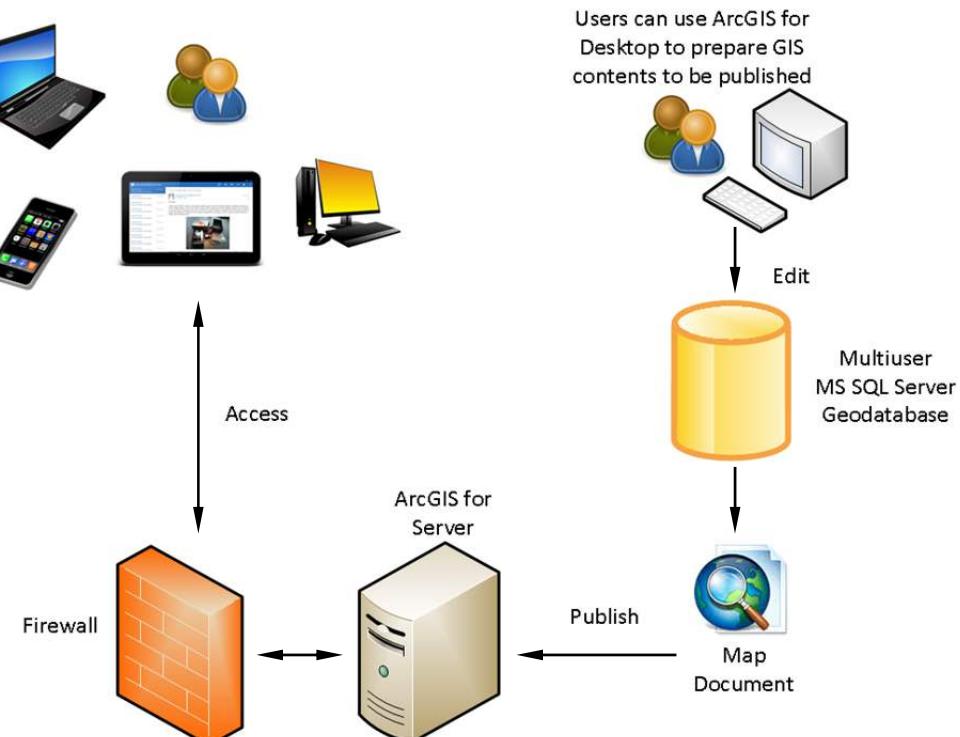
Figure 118 DSS architecture.

ArcGIS for Server is a product developed by Environmental Systems Research Institute (ESRI) for publishing GIS products online. The published contents can be accessed via computers connected to the internet, various mobile devices (e.g., tablets and smartphones), and other devices that can connect to web services. For this project, ArcGIS for Server, ArcGIS for Desktop, and Microsoft SQL Server were all installed on a workstation purchased for laser data acquisition. All computers connected to the UMass Lowell campus network can have access to the published GIS contents. For full system implementation, these software tools can also be deployed on virtual machines or commercial cloud platforms.

ArcGIS for Desktop was used to edit the GIS contents to be published. It is also seamlessly integrated with ArcGIS for Server. ArcGIS for Desktop includes a straightforward wizard to publish GIS contents that include both data and services. Thus, ArcGIS for Desktop serves as client and also a tool that we use to create the GIS resources to be hosted on our GIS server. The developed DSS architecture is illustrated in Figure 118. During the publishing process, service providers have the option to enable capabilities that will define the various ways users can access the online GIS services. Once the GIS services are up and running, they can be used in many applications, devices, or APIs that can communicate through HTTP (Hyper Text Transfer Protocol).

ArcGIS for Server also includes a component called Manager, which is essentially a web application that provides an intuitive point-and-click interface for administering the server. For this project, this Manager was used to view the server logs, stop and start GIS services, publish service definitions, create users and define their roles for security, and to perform other similar tasks.

7.2.4. ArcGIS API for JavaScript

ArcGIS API for JavaScript is another key technology used in this project. It allows us to develop custom applications that utilize the full breadth of the published web services within a web interface we designed. It is a lightweight way to embed maps and tasks in web applications. Based on the maps and data on our own GIS Server, we were able to use ArcGIS API for JavaScript to build the web-based DSS that allows users to view the maps and data, query them, generate reports, and to display widgets such as map legend, overview map, and measurement widget.

The ESRI developers who created the ArcGIS API for JavaScript used Dojo to simplify their development process and to ensure that the developed applications behave the same in different web browsers. For example, the map zoom in/out and pan animations both use Dojo. The graphics layer also depends on Dojo. Additionally, the zoom level slider and info window are Dojo widgets.

JavaScript is a language that runs within a web browser. It is the basis for the developed web-based DSS. Dojo is an open source toolkit that helps users write robust and efficient JavaScript code. Toolkits like Dojo provide functions that allow users to do things in easier or more efficient ways. Using libraries from the Dojo toolkit can reduce the lines of code written and make JavaScript applications quicker and more stable. Also, Dojo fully supports creating classes and modules using Asynchronous Module Definition (AMD) style code.

7.2.5. Technologies for Mobile Application

As discussed before, the project team also developed a mobile application for passenger rail inspection. This mobile application was developed using **PhoneGap**, which is a free and open source framework that allows developers to create mobile applications using standard web APIs. Most mobile applications are developed using platform-specific APIs such as those for iOS or Android devices. However, PhoneGap enables software developers to build applications for mobile devices using JavaScript, HTML5, and CSS3. The resulting mobile applications are considered hybrid, meaning that they are neither truly native mobile applications nor purely web-based applications. They are essentially web views packaged as mobile applications and have access to native platform-specific APIs.

A major disadvantage of mobile Apps built using PhoneGap is their potential lack of performance. These Apps are typically slower compared to native Apps developed using platform-specific APIs. However, an important benefit of using PhoneGap is that web developers can develop a mobile App quickly using JavaScript, HTML5, and CSS3. They do not need to learn a new programming language such as Java for Android and Xcode for iOS. This is particular important for this project, since the web-based DSS was developed using JavaScript, HTML, and CSS. Choosing PhoneGap makes it much easier and quicker to develop a mobile App based on the web-based DSS. Hence, we used PhoneGap for this project to create an accompanying mobile App for the web-based DSS. Although its speed is not as fast as native Apps, we felt that it is sufficient for a prototype system demonstrating the idea of using mobile App for passenger rail inspection. In the future, this prototype App can be improved or redeveloped as a native mobile App.

7.2.6. Other Technologies

HyperText Markup Language (HTML) is a standard markup language used to create web pages. It is used in this study to generate the basic framework that gives the web-based DSS its structure.

Cascading Style Sheets (CSS) is used to define the layout of text and other materials for the project website. It is critical for designing the appearance of the project web pages.

JavaScript (JS) is a popular and dynamic computer programming language. It is commonly used as part of web browsers. JavaScript allows client-side scripts to interact with users, control the web browser, communicate asynchronously, and to alter the document content that is displayed.

jQuery is a fast, small, and feature-rich JavaScript library. It makes things like HTML document traversal and manipulation, event handling, animation, and Ajax much simpler with an easy-to-use API that works across a multitude of browsers.

jQuery UI is a collection of GUI widgets, animated visual effects, and themes implemented with jQuery, CSS, and HTML. In this project, we utilized some of its widgets, for example, the dialog widget.

PHP is a popular server-side scripting language for developing dynamic web applications. It can handle a variety of tasks. It can evaluate form data sent from a browser, build custom web content and send it to the browser, fetch data from a database, and even send and receive cookies. In this project, we mainly used PHP to interact with data in Microsoft SQL Server.

Structured Query Language (SQL) is a standard programming language designed for managing and accessing data in a relational database management system (RDBMS). SQL consists of data definition and manipulation languages. It can be used to insert, query, update, and delete data. It is also used for schema creation and modification and data access control.

7.3. System Functionalities

In this section, the main functionalities of the web-based DSS are described. These functionalities are grouped into 1) common GIS tools, 2) query and reporting tools, and 3) measurement tools. All these functionalities are also available in the mobile App.

7.3.1. Common GIS Tools

The developed DSS provides a set of commonly used GIS tools to allow users easily navigate among different key features. These tools include:

1. 📍 **Identify** – Upon clicking on the map, this tool displays the information about all features located at that point from different layers (see Figure 119 below). Also, upon clicking on any train station features, it will execute the 'Search by Station' query (see 4.iii in Section 7.3.2.) and display more detailed information about a station's subcomponents and inspection history. This tool remains active until deactivated.
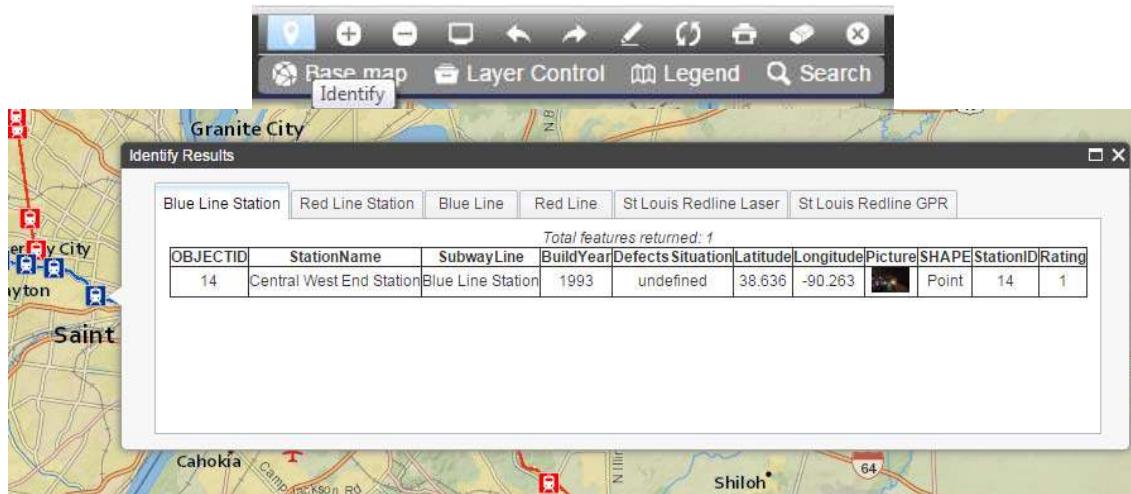
Figure 119 Result of applying the identify tool.

2. ➕ **Zoom In** – Single left click to activate this tool. Click and drag simultaneously to draw a rectangle to zoom into that particular area. This tool remains active until deactivated.

3. ➖ **Zoom Out** – Single left click to activate this tool. Click and drag simultaneously to draw a rectangle to zoom out. This tool remains active until deactivated.

4. 🖵 **Full Extent** – Click on it to zoom to the initial extent of the map.

5. ↖ **Previous Extent** – Click to go back to the previous extent of the map.

6. ➡ **Next Extent** – Click to go to the next extent of the map.

7. ✎ **Edit Features** – By clicking on this tool, users will be directed to a new web page (see Figure 120) wherein they can examine and update the attribute information of any stations or line segments. They can also delete and add features. From the left window, they can select the type of feature they would like to add. To add a feature, users can simply click on the map where they want to place it. It is also possible to modify an existing feature's position by dragging it.

    Click on 'go back to homepage' will allow users to go back to the main page (previous page). Upon clicking on 'Maintenance date filter', a new tab will be opened (highlighted in the red rectangle in Figure 120). This tool will allow users to display selected features that satisfy certain criteria. It can help transit agencies visualize their maintenance and capital project activities and help them better schedule those activities.

8. ⟳ **Refresh** – Click to refresh/reload the map.

9. 🖨 **Print Map** – Click to print the map.

10. ◈ **Clear Selection** – Click to deselect any graphics or features that are currently being selected on the map.

11. ⊗ **Deactivate** – Click to deactivate the currently active tool.
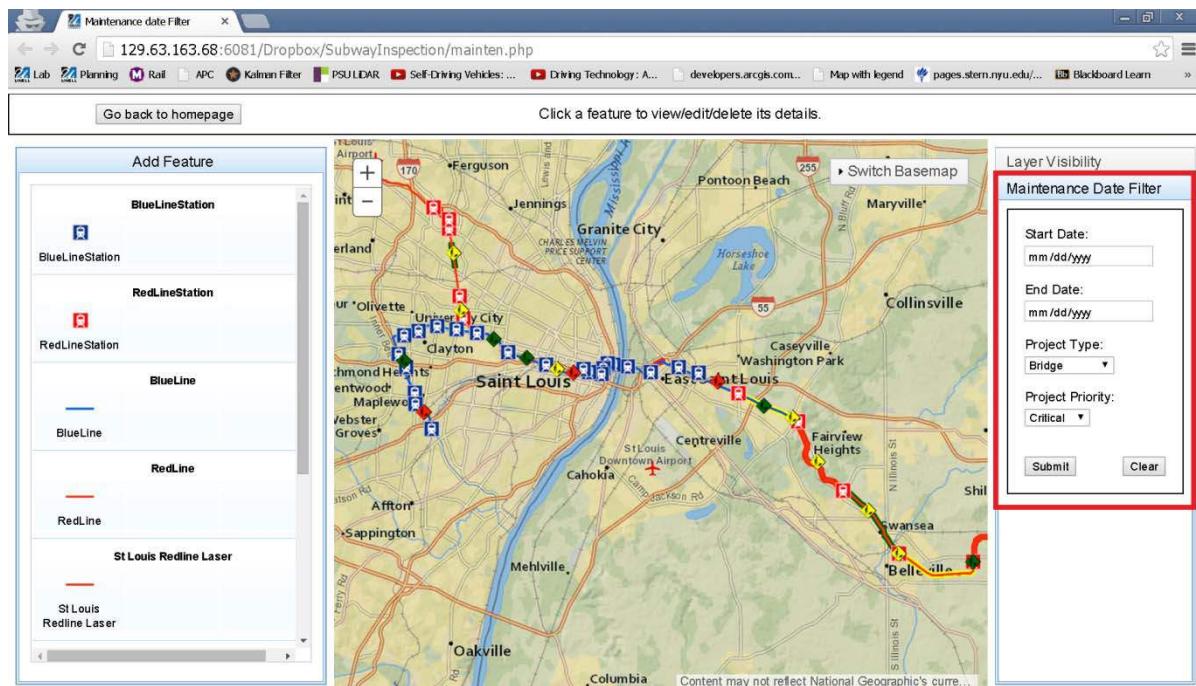
Figure 120 New webpage for updating and editing features.

12. **Overview Map at the Bottom Right Corner** – It turns on/off an overview window of the map.

7.3.2. Query and Reporting Tools

In this section, tools related to visualize/display the data, layer control, queries, and reporting are described.

1. **Base Map** – Hover on it and choose among 12 different base maps.

2. **Layer Control** – Click to open the layer control dialog box. Click on each layer name to make it visible or hide it. Users can also click and drag a layer up or down to change its drawing order.

    This layer control tool is synchronized with the '*Legend*', '*Selected Layers*' (see Figure 121), '*Identify*', and '*Search by Shapes*' tool. When a layer is checked off (i.e., made invisible), it will not appear nor will be queried in the above mentioned tools.
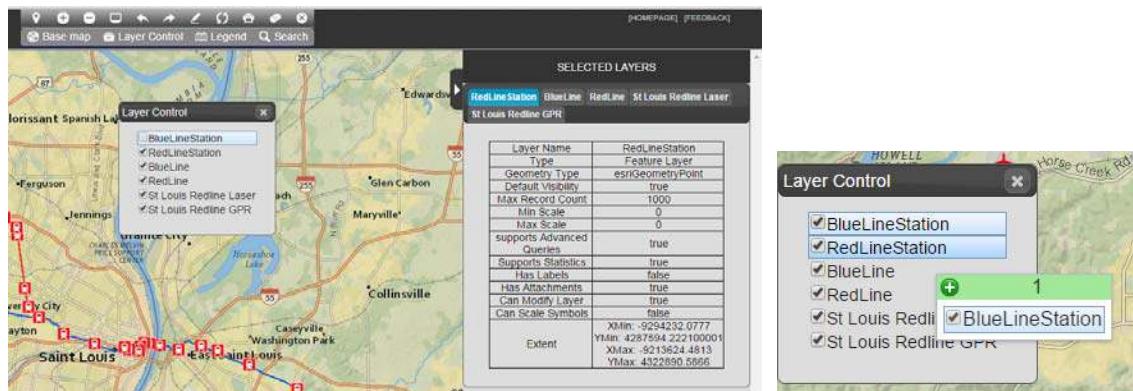
Figure 121 Layer control tool.

3. **Legend** – Click to view the legend for features in visible layers.

4. **Search** – Hover and select from different types of queries.

    i. *Search by Shapes* – This tool allows users to select features fall within a particular extent drawn by them. There are six options available, which are 'rectangle', 'circle', 'ellipse', 'triangle', freehand polyline', and 'freehand polygon'. The selection result will be displayed in the bottom slider with each layer in a different tab. Users can click on any record to zoom into the corresponding feature. If a record contains an image, they can click on the thumbnail to open the full-size image in a new window.

97

ii. ***Search by Address*** – Click on the text field to enter a location. Based on the input text, some recommended locations may prompt up. Press 'enter' or select one entry from the recommended list.

iii. ***Search by Station*** – This tool shows the detailed information about the rail station of interest. In addition to the information (e.g., attributes and inspection history) about the station itself, this tool also shows the subsystems of each station and the various components of each subsystem, including their attribute information and inspection history as shown in Figure 122.

Select any station from the dropdown list and press search. The subsystems for that station will be displayed in the bottom slider. Clicking on a subsystem name will further display its components.

By clicking on the 'Switch to History' link, the historical inspection data of the given station or subsystem will be displayed. Users can also add new inspection records and attach pictures to them.

Upon clicking on a previous inspection record, a window will prompt up and display a list of files attached to it. Clicking on a file name will open that file. Clicking on 'View Gallery' will open up the image gallery to display all images (if any) that have been attached to that inspection record. Clicking on the export icon at the upper right corner will allow users to download the displayed results in an Excel file.
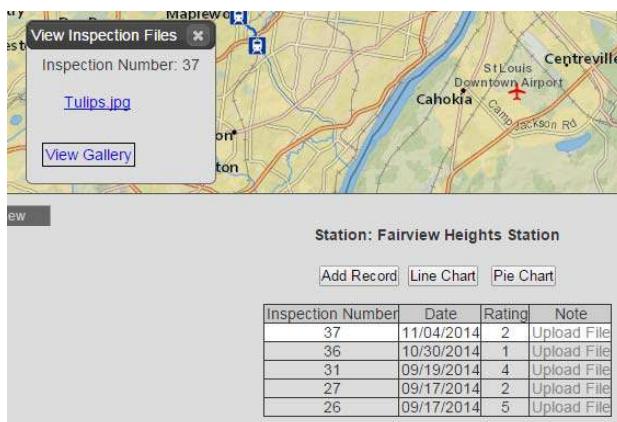
Figure 122 Search by station tool.

iv. ***Search by Layer*** – This tool allows users to search features from a chosen layer and narrows them down by an optional keyword. For the displayed results, clicking on a record will zoom to the corresponding feature on the map. Clicking on an image link will open the image in a new window. Users can also download the search results as an Excel file.

v. ***Search by Attributes*** – This tool is similar to the above '*Search by Layer*' except that it allows users to define the search criterion more clearly.

5. **Right Sidebar** – It displays layer information in different tabs (see 'SELECTED LAYERS' in Figure 121).

6. **Bottom Slider** – It displays results from the various search tools (see the bottom half of Figure 122).

7.3.3. Measurement Tools

1. **Area** – Click to activate this tool. Once activated, click to draw a polygon encompassing the area to be measured and double click to complete the drawing. This tool remains active until deactivated.

2. **Distance** – Click to activate this tool. Once activated, click to draw lines and double click to complete the drawing. This tool remains active until deactivated.

3. **Location** – Hover anywhere on the map to view its latitude and longitude. Click on map to drop a flag and display its latitude and longitude.

7.4. Data Dictionary

To explain the developed web-based DSS, it is important to describe the structures of the database and the tables in it. Using the data obtained from Metro St. Louis as an example, the following tables explain how the database was developed. Table 5 and Table 6 are for rail stations and lines, respectively. For the rail station table, an attribute field "StationID" was included. This "StationID" field allows us to link other dynamic information (e.g., inspection history) with each rail station. Table 6 was meant to show the static information associated with each rail segment. However, we can certainly add an additional attribute field such as "RailID" to this table so that it can be linked to other tables with dynamic information for each rail segment.

Table 5 Attribute table for Blue/Red Line stations

FIELD NAME	DATATYPE	FIELD SIZE	KEY	NULLABLE
OBJECTID	Int		Primary Key	No
StationName	Nvarchar	50		Yes
SubwayLine	Nvarchar	50		Yes
BuildYear	Nvarchar	50		Yes
DefectsSituation	nvarchar	50		Yes
Latitude	numeric	38,8		Yes
Longitude	numeric	38,8		Yes
Picture	nvarchar	50		Yes
SHAPE	geometry			Yes
StationID	int			Yes
Rating	int			Yes

Table 6 Attribute table for Blue/Red Line

FIELD NAME	DATATYPE	FIELD SIZE	KEY	NULLABLE
OBJECTID	int		Primary Key	No
LayerName	nvarchar	50		Yes
StationFrom	nvarchar	50		Yes
StationTo	nvarchar	50		Yes
BuildYear	nvarchar	50		Yes
OperatedBy	nvarchar	50		Yes
SHAPE	geometry			Yes
Image	int			Yes
Rating	int			Yes

Table 7, Table 8, and Table 9 are for publishing the processed GPR, laser, and track dynamic model results. To accommodate the insertion of the generated images, the entire rail lines were divided into equal segments. For this project, we used one line feature to represent both westbound and eastbound lines, because the westbound and eastbound lines in St. Louis were parallel to each other. For other rail transit agencies with nonparallel lines, these tables can be modified easily to represent the two directions separately.

For each of the three tables, attribute fields were created to store the images generated by the GPR, laser, and track dynamic models. Intuitively, one cannot query those images. To make it easier for system users to identify problematic or suspicious segments, numeric attribute fields were also created for each of the three tables, representing the track conditions of different segments. Using Table 8 as an example, attribute fields were created to store gauge, tie conditions, and number of missing fasteners. With such information, users can easily locate segments with surface defects. With the identify tool mentioned earlier, the GPR and track dynamic model results of the same segments can be reviewed.

Table 7 Attribute table for Blue/Red Line GPR data

FIELD NAME	DATATYPE	FIELD SIZE	KEY	NULLABLE
OBJECTID	int		Primary Key	No
LayerName	nvarchar	50		Yes
StationFro	nvarchar	50		Yes
StationTo	nvarchar	50		Yes
BuildYear	nvarchar	50		Yes
OperatedBy	nvarchar	50		Yes
SHAPE_Leng	numeric	38,8		Yes
SHAPE	geometry			Yes
AK_W	varchar			Yes
AK_E	varchar			Yes
UVM_W	varchar			Yes
UVM_E	varchar			Yes
Rating_W	float			Yes
Rating_E	float			Yes

Table 8 Attribute table for Blue/Red Line laser data

FIELD NAME	DATATYPE	FIELD SIZE	KEY	NULLABLE
OBJECTID	int		Primary Key	No
LayerName	nvarchar	50		Yes
StationFro	nvarchar	50		Yes
StationTo	nvarchar	50		Yes
BuildYear	nvarchar	50		Yes
OperatedBy	nvarchar	50		Yes
SHAPE_Leng	numeric	38,8		Yes
SHAPE	geometry			Yes
pic_W	varchar			Yes
pic_E	varchar			Yes
Gauge_W	float			Yes
Gauge_E	float			Yes
Tie_W	float			Yes
Tie_E	float			Yes
FT_W	float			Yes
FT_E	float			Yes
GaugeR_W	float			Yes
GaugeR_E	float			Yes
TieR_W	float			Yes
TieR_E	float			Yes
FTR_W	float			Yes
FTR_E	float			Yes

Table 9 Attribute table for Blue/Red Line track dynamic model data

FIELD NAME	DATATYPE	FIELD SIZE	KEY	NULLABLE
OBJECTID	int		Primary Key	No
LayerName	nvarchar	50		Yes
StationFro	nvarchar	50		Yes
StationTo	nvarchar	50		Yes
BuildYear	nvarchar	50		Yes
OperatedBy	nvarchar	50		Yes
SHAPE_Leng	numeric	38,8		Yes
SHAPE	geometry			Yes
pic_W	varchar			Yes
pic_E	varchar			Yes
Rating_W	float			Yes
Rating_E	float			Yes

Based on our discussions with some rail transit agencies, they all rely on manual methods (e.g., paper reports or Excel spreadsheets) to schedule maintenance activities and capital projects. As part of this project, we introduced a GIS tool to visualize planned or proposed maintenance

activities and capital projects on a map. Different divisions of a rail transit agency will have a designated employee to add/update their requests for right of way (ROW) using this GIS system. These requests are stored using Table 10. Such requests will be visible to other divisions and the Chief of Rail Operation & Maintenance. Such an information-sharing platform will allow different divisions to collaborate with each other and resolve conflicts. The online request submission system makes it easier for the Chief of Rail Operation & Maintenance to gather information. The visualization of maintenance requests on a map can better show their temporal and spatial distributions, which is critical for making informed decisions.

Table 10 Attribute table for maintenance activities

FIELD NAME	DATATYPE	FIELD SIZE	KEY	NULLABLE
OBJECTID	int		Primary Key	No
StationName	nvarchar	50		Yes
SubwayLine	nvarchar	50		Yes
StartMilePost	nvarchar	50		Yes
EndMilePost	nvarchar	50		Yes
SHAPE	geometry			Yes
StartDate	datetime2(7)			Yes
EndDate	datetime2(7)			Yes
ProjectType	nvarchar	50		Yes
ContactInfo	nvarchar	50		Yes

A lot of rail assets have subsystems and each system also consists of several components. For example, a train station includes major subsystems such as escalators, ticket boxes, and platform. An escalator also consists of step chain, structural truss, etc. Each subsystem or component may need to be inspected at a specific frequency. It would be interesting to develop a set of tables to capture such a hierarchical structure and all the historical inspection results.

The following Table 11 through Table 17 are introduced specifically for this purpose. Table 11, Table 13, and Table 15 are for storing static information for train station, subsystem, and component, respectively. The hierarchical relationship among them is defined by StationID, SystemID, and ComponentID. Table 12, Table 14, and Table 16 are for inspection records. Each record has a unique identification number called InspID. Since each station / subsystem / component can have many inspection records. Intuitively, separating static information from inspection records can help reduce information redundancy. Different inspections may generate various numbers of images or other attachments. For reducing information redundancy purpose, it would be better to save these images / attachments in a separate table. Table 17 was introduced for this purpose. Since each inspection has a unique ID, which is InspID. These records in Table 17 can be easily linked with a station, a subsystem, or a component.

Table 11 Attribute table for static train station information

FIELD NAME	DATATYPE	FIELD SIZE	KEY	NULLABLE
StationID	int			No
name	nvarchar	50		No
pic	nvarchar	50		Yes

Table 12 Attribute table for train station inspection records

FIELD NAME	DATATYPE	FIELD SIZE	KEY	NULLABLE
StationID	int			No
InspID	int			No
date	datetime			No
rating	int			Yes

Table 13 Attribute table for static subsystem information

FIELD NAME	DATATYPE	FIELD SIZE	KEY	NULLABLE
SystemID	int			No
StationID	int			No
name	nvarchar	50		No
pic	nvarchar	50		Yes

Table 14 Attribute table for subsystem inspection records

FIELD NAME	DATATYPE	FIELD SIZE	KEY	NULLABLE
SystemID	int			No
StationID	int			No
InspID	int			No
date	datetime			No
rating	int			Yes

Table 15 Attribute table for static component information

FIELD NAME	DATATYPE	FIELD SIZE	KEY	NULLABLE
ComponentID	int			No
SystemID	int			No
StationID	int			No
name	nvarchar	50		No
pic	nchar	10		Yes

Table 16 Attribute table for component inspection records

FIELD NAME	DATATYPE	FIELD SIZE	KEY	NULLABLE
ComponentID	int			No
SystemID	int			No
StationID	int			No
InspID	int			No
date	datetime			No
rating	int			Yes

Table 17 Table for inspection attachments

FIELD NAME	DATATYPE	FIELD SIZE	KEY	NULLABLE
name	nvarchar			Yes
type	nvarchar			Yes
size	int			Yes
content	varchar			Yes
InspID	int			Yes
DateInserted	datetime			No

7.5. Future Work

The developed DSS demonstrated how GIS can be used for rail asset and inspection data management. Several innovative ideas have been proposed and some of them have been implemented. Given the limited resources, the developed DSS still needs many improvements. In the future, the developed DSS can be further enhanced from the following perspectives:

1. The system can be improved by developing native mobile applications for data collection in the field. Also, we can add tools such as offline data editing, saving all data on local database while in the field, and uploading all the data and images taken to the enterprise database on Wi-Fi. Further we can implement a complete synchronization between mobile apps and the Web-GIS system to eliminate any redundancies.

2. For the editing tool, we can implement more restrictions to control who can add, delete, or modify features. For example, we might allow some users to view features but not to add, modify, or delete them. Another strategy can be ownership-based access control which allows users limit access to features if they were created by others.

3. To ensure data integrity, we can add logs to track changes to the data (e.g., who created or updated which features at what time). This is especially important for a multiuser GIS system.

CHAPTER 8

OUTREACH

The team tried many venues to publicize and promote the developed system, which includes meetings with various rail transit agencies and stakeholders, presentations and exhibitions at regional, national, and international conferences, and commercialization of the developed products.

8.1. Meetings with Rail Transit Agencies and Stakeholders

The team reached out to the following transit agencies: Metro St. Louis, MBTA, Cleveland Metro, CTA, and New York Metro. Specifically,

- On December 16, 2014, the team held a meeting with a group of senior leaders from Metro St. Louis to present the developed system, especially the GIS-based decision support system. Attendees of this meeting include Ray Friem (Chief Operation Officer), Scott Grott (Chief Rail Operation & Maintenance), Tracy Beidleman (Director of Programs & Grants), Brain Sellers, Director of Track Maintenance, David Still (Maintenance of Way Supervision), and David Beal (Programs & Grants).
- The team contacted MBTA in December 2014 and scheduled a meeting on January 22, 2015. At this meeting, the team summarized the entire research and discussed future research directions, including cross-tie deformation detection. Attendees of this meeting include Mr. Pascal Baran (Chief Engineering Officer of Keoliscs Commuter Services).
- CodeRed Business Solutions (one of the team members) reached out to Greater Cleveland Regional Transit Authority (RTA) in December 2014. In January 2015, CodeRed presented the project to Richard Newell (Director of Service Quality) and Dr. Floun'say Caver (Director of Bus Operations & Maintenance Hayden District).
- In November 2014, CodeRed Business Solutions reached out to Michal Salvato (Program Manager, Enterprise Asset Management) at New York City Transit (NYCT) & MTA. Our points of contact at both Cleveland RTA and NYCT & MTA stated that the developed tool would provide a capability seriously needed by the rail transit industry and would significantly increase productivity of maintenance personnel and reduce ridership complaints.
- On September 24, 2014, three researchers (Dr. Hailing Yu, Dr. Theodore R. Sussmann, and Dr. Robert Wilson) from the USDOT Volpe Transportation Center and Mr. Gary Carr (Chief of the Track Division of Federal Railroad Administration) visited UMass Lowell. The PI demonstrated the laser system to them. Dr. Theodore R. Sussmann and Mr. Gary Carr also serve on the TAC of this project. Mr. Carr showed strong interest in the laser subsystem of this project. He suggested that the laser system may be used to detect cross-tie deformation due to heavy load or weak track foundation.
- The PI worked with CodeRed Business Solutions and reached out to Jacobs Engineering.

- The PI was contacted by a company called RailPod (http://www.rail-pod.com/). The PI followed up with this company and discussed potential collaboration opportunities.
- The PI recently attended the 2014 Railway Engineering Education Symposium (REES) organized by the American Railway Engineering and Maintenance-of-Way Association (AREMA). At this symposium, the PI shared this research with some potential stakeholders there, including Mr. David Davis from the Transportation Technology Center Inc. (TTCI).

Since this proposed integrated system is very comprehensive, adopting this entire system requires the support from different divisions of a rail transit agency. Based on the team members' experience with various rail transit agencies, we do not expect this to happen soon. However, we have been successful in convincing Metro St. Louis to adopted part of the GIS-based decision support system. As part of this project, the team developed a prototype mobile App that can be used for rail asset tracking and condition monitoring. This system was later improved by CodeRed Business Solutions and is currently being used by Metro St. Louis to inventory their assets.

8.2. Conferences

Since the beginning of this project, the team has presented this research many times at various levels of conferences, including

- Automatic Track Inspection Using 3D Laser Profilers to Improve Rail Transit Asset Condition Assessment and State of Good Repair – A Preliminary Study. This paper has been presented at the 93rd TRB Annual Meeting. The paper has also been included in the 2014 TRB Meeting CD-ROM.
- Development of a High-Speed Ultra Wideband (UWB) Ground Penetrating Radar for Railroad Ballast Inspection. This was a PowerPoint presentation at the 93rd TRB Annual Meeting.
- Spectrum Analysis of Railroad Track with Hanging Tie Problem. This paper has been presented at the 2014 Joint Rail Conference, April 2-4, Colorado Springs, CO. The paper has also been included in the conference CD-ROM.
- Data Analysis Technique to Leverage GPR Ballast Inspection Performance. This paper has been accepted for presentation at the 2014 IEEE Radar Conference, May 19-23, 2014, Cincinnati, OH.
- An Automated System for Rail Transit Infrastructure Inspection. This was an invited PowerPoint presentation at the New Hampshire Society of Professional Engineers (NHSPE) 63rd Annual Engineers Week Awards Banquet & Exhibition.
- An Automated System for Rail Transit Infrastructure Inspection. This was an invited PowerPoint presentation at the November CESTiCC Webinar of the Tier 1 Center for Environmentally Sustainable Transportation in Cold Climates (CESTiCC).
- Hanging Tie Study Using "Moving Deflection Spectrum". This was presented at the 94th TRB Annual Meeting. The paper has also been included in the 2015 TRB Meeting CD-ROM

8.3. Commercialization

In addition, one of the team members (Pavemetrics Inc.) has been working diligently to commercialize and promote the developed laser hardware and software products. The developed algorithms have been integrated into a software package called LRAIL, which is being promoted on their company's website at http://www.pavemetrics.com/en/lrail.html (see Figure 123).
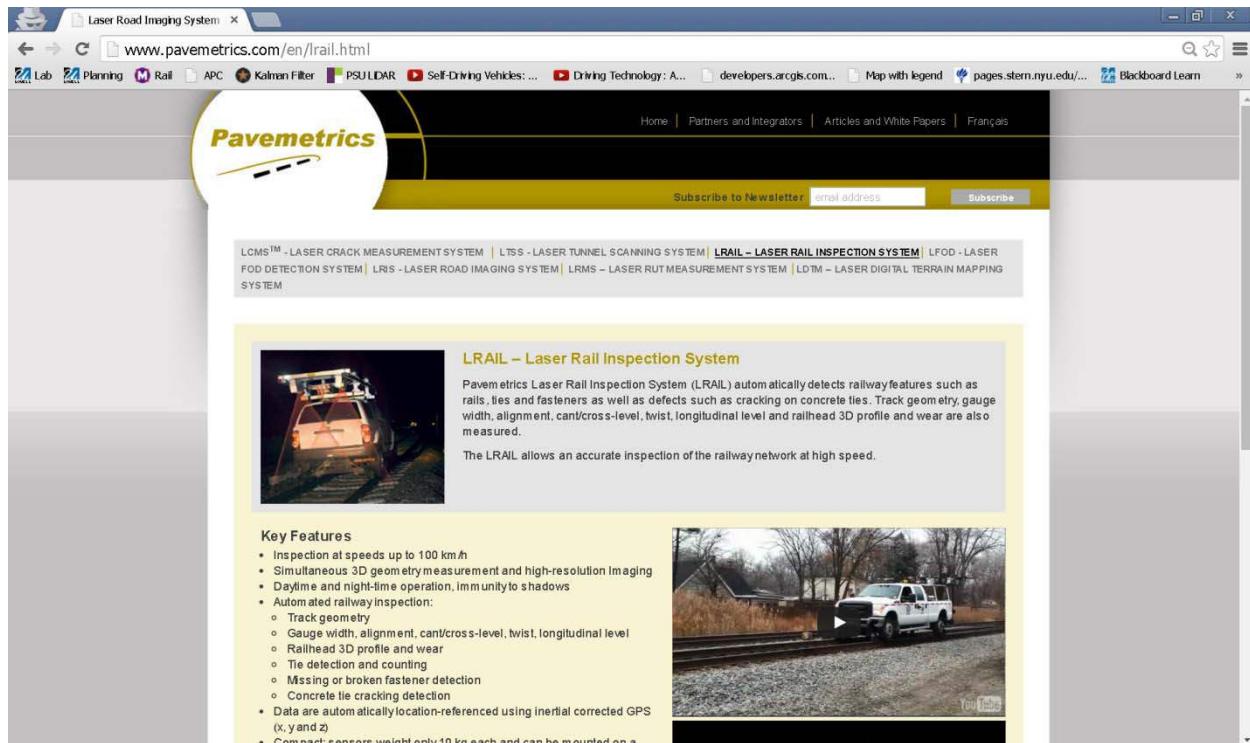
Figure 123 LRAIL product in Pavemetrics website.

In addition to their website, Pavemetrics has been promoting the developed LRAIL system at various national and international conferences. The following table is a list of conferences where LRAIL has been demonstrated.

Table 18 Conferences where LRAIL has been demonstrated.

Transportation Research Board (TRB)	Washington D.C.	Jan. 2014 and Jan. 2015
2014 Transport Research Arena	Paris, France	April 2014
SPAR International	Colorado Springs, CO	April 2014
American Public Transit Association (APTA) Rail Conference	Montreal, QC	June 2014
American Public Works Association	Toronto, ON	August 2014
American Public Transit Association (APTA) Annual Meeting	Houston, TX	October 2014
Bentley YII (Year in Infrastructure Road and Rail)	London, UK	November 2014

CHAPTER 9
CONCLUSIONS AND DISCUSSION

9.1. Conclusions

This project developed an integrated system for passenger rail inspections. This integrated system consisted of four major components: *GPR subsystem, laser subsystem, track dynamic model subsystem*, and *a web-GIS based Decision Support System (DSS)*. The GPR subsystem was designed to identify track subsurface problems such as fouled ballast and suspicious underground objects; the laser subsystem was for detecting surface track defects such as missing fasteners and cross ties, large cracks in cross ties, and wide rail gauge; the track dynamic model was developed to identify a critical track structure problem called hanging tie using spectrum analysis; and the web-GIS based DSS was developed to manage and visualize the results generated by the previous three subsystems. The first three subsystems were integrated and mounted on a hi-rail SUV. This automated and operational (i.e., the equipped SUV) system was then tested at both MBTA and Metro St. Louis. The team also developed algorithms for processing the GPR, laser, and track dynamic model data. The data collected from MBTA and Metro St. Louis was processed by these algorithms and fed into the DSS. The main objective of this project was to provide a cost-effective solution to rail transit agencies for their weekly track inspections, which are currently performed by inspectors walking along the track at most rail transit agencies.

The team followed the proposed team project activities (Attachment #1 submitted to OST-R) and project deliverables (Attachment #2) and have successfully carried out all the proposed tasks. The progresses of these tasks have been summarized in this final report and previously submitted quarterly reports. The following is a brief summary of the four main components. Also, during the course of this project, the team identified additional interesting and innovative ideas. Some of these ideas have been implemented in addition to completing the work promised in the initial proposal. These ideas are also discussed below.

- The team successfully developed a high-speed GPR system with up to 1 cm horizontal resolution at 100 km/h survey speed. Its penetrating capability can reach 1 meter. Such a high-speed and high-resolution GPR system can well satisfy the subsurface inspection needs of passenger rail transit agencies. This GPR system has been tested extensively in our lab at UMass Lowell and University of Vermont and in the field. The lab test results suggest that the developed system was able to successfully detect underground pipes and fouled ballast. The team also developed a suite of algorithms for analyzing the GPR data. Specifically, a 2D entropy algorithm was developed to automatically mark suspicious features and regions in B-scan images. Such a feature is particularly important for processing large data sets generated by GPR scans.

- The team adopted a commercial laser product (i.e., LCMS, which was initially developed for highway pavement inspection) by Pavemetrics for rail surface inspection. Additionally, the team developed algorithms for processing the rail laser

data. The quality of the collected data and the results generated by the developed algorithms suggest that it is a viable and very promising solution to use LCMS for detecting rail surface defects. Pavemetrics (a member of the project team) has already commercialized the developed algorithms and named it as LRAIL. Several rail agencies have expressed interest in this product, although their names cannot be disclosed at this time. Our idea of using LCMS for rail inspection has also drawn attention from some well-known Civil Engineering companies. One of them (Mandli) even volunteered their equipment and employees' time and participated in the first laser data collection in St. Louis. Mr. Gary Carr (Chief of the Track Division of Federal Railroad Administration) also showed interest in this technology and suggested that it may potentially be used for measuring cross tie deformation.

- The team installed a 3-axial accelerometer on a steel wheel of the hi-rail SUV to measure track deflection. Spectrum analysis was used to analyze the frequency of the collected data. The accelerometer data was compared with the track deflection data generated by a track dynamic model. From the comparison results, we were able to identify some interesting matches between the two sets of data and unique features for track segments with hanging ties. Such findings suggest that using low-cost accelerometer sensors can have great potential in identifying critical track defects.

- The team successfully developed a web-GIS based DSS for managing and visualizing the collected data. In addition to satisfying this basic requirement, the team also developed a mobile App, designed the database to handle hierarchical relationships among rail assets, and developed a tool that facilitates the scheduling or maintenance activities and capital projects. These tools have great potential to significantly improve the current asset inspection, record keeping, and maintenance activity scheduling practices at many rail transit agencies in the United States. These tools have been demonstrated to several transit agencies. The team is currently negotiating with several agencies and trying to move these ideas into the implementation phase.

Overall, this project generated many interesting results and interest among stakeholders. Some of them volunteered to serve on the project technical advisory committee. Getting the developed system fully implemented at a rail transit agency requires the support and coordination among different divisions such as IT, maintenance, grants, and operation. Based on the team's prior experience, this usually takes a lot of time and varies from agency to agency. The team felt that a more viable means to commercialize the developed system is to start with the private sector. Two members of the team, Pavemetrics and CodeRed Business Solutions, are working diligently on this and have made promising progresses. The progresses they have made in such a short time period confirmed the importance and potential of the developed system.

9.2. Discussion

While conducting the project, the team identified the following areas that may deserve further investigations:

- A set of home-made UWB horn antennas (see Figure 29) was used in this study. With the recently development in 3D printing, it would be very interesting to see how 3D printing can improve the performance of the UWB antennas.

- As suggested by Mr. Gary Carr, it would be interesting to use the adopted laser system for measuring tie deformation. Multiple measurements can be taken at different vehicle loads. The tie deformation results under different vehicle loads can be very important for assessing track health conditions.

- Digital Image Correlation (DIC) has recently attracted much attention and has been widely used for structure health monitoring. Such technology may also be used for detecting track surface problems. It would be interesting to compare the performance of DIC with the laser technology adopted in this project.

- Due to safety restrictions, it is difficult to take subsurface samples from the field to verify features (e.g., suspicious feature or regions) detected by the GPR subsystem (Fortunately, for this project we were able to test the GPR subsystem in our labs). In future studies, algorithms can be developed to compare GPR data from multiple scans. If significant changes have been identified between different scans, there is a good reason to send inspectors to check out that spot and even take subsurface samples to find out the reason.

In future studies, it would be better to also coordinate with the Transportation Technology Center Inc. (TTCI) to carry out additional field tests. TTCI has 48 miles of track for testing different track components. The developed system can be tested under a controlled setting to better assess and validate the performance of each subsystem.

REFERENCES

1   U.S. Federal Transit Administration. *Transit State of Good Repair: Beginning the Dialogue*. Oct. 2008.

2   *Transportation Safety Board of Canada. Statistical Summary, Railway Occurrences 2012*. Available online at http://www.tsb.gc.ca/eng/stats/rail/2012/ss12.asp, Accessed July 20, 2013.

3   Saarenketo, T., and T. Scullion. *Ground Penetrating Radar Applications on Roads and Highways*. Research Report *1923-2F*. Texas Transportation Institute, the Texas A&M University System, College Station, TX, 1994.

4   Maser, K. R. *Pavement Characterization Using GPR: State of the Art and Current Practice*. Special Technical Publication 1375, Nondestructive Testing of Pavements and Backcalculation of Moduli. American Society for Testing and Materials, West Conshohocken, PA, 2000.

5   Maser, K. R. Condition Assessment of Transportation Infrastructure Using Ground-penetrating Radar. In *Journal of Infrastructure Systems*, Vol. 2, No. 2, 1996, pp. 94-101.

6   Stephen, S., C. Estakhri, T. Scullion, and W. Liu. *New Technology for Evaluating Flexible Pavement Construction: Year 1 Report*. Research Report 0-4774-1. Texas Transportation Institute, the Texas A&M University System, College Station, TX, 2006.

7   Maser, K. R., T. J. Holland, R. Roberts, and J. Popovics. NDE Methods for Quality Assurance of New Pavement Thickness. In *International Journal of Pavement Engineering*, Vol. 7, No. 1, 2006, pp. 1-10.

8   Scullion, T. *Using Rolling Deflectometer and Ground Penetrating Radar Technologies for Full Coverage Testing of Jointed Concrete Pavements*. Research Report 0-4517-2. Texas Transportation Institute, the Texas A&M University System, College Station, TX, 2006.

9   Uddin, W. *Ground Penetrating Radar Study – Phase I Technology Review and Evaluation*. Final Report MS-DOT-RD-06-182. University of Mississippi, Center for Advanced Infrastructure Technology, University, MS, 2006.

10 Schlorholtz, S., B. Dawson, and M. Scott. *Development of In-situ Detection Methods for Material-related Distress (MRD) in Concrete Pavements: Phase 1 Report*. Iowa State University, Ames, IA, 2003.

11 Yen, K.S., B. Ravani, and T.A. Lasky. *LiDAR for Data Efficiency*. Final Report WA-RD 778.1. Washington State Department of Transportation, Sep., 2012.

12 Veneziano, D., S. Hallmark, and R. Souleyrette. *Comparison of LiDAR and Conventional Mapping Methods for Highway Corridor Studies*. Final Report, Center for Transportation Research and Education, Iowa State University, 2002.

13 Sawadisavi, S.V., J.R., Edwards, E., Resendiz, and J.M, Hart. Machine-Vision Inspection of Railroad Track. In *Proceedings of the 88th Transportation Research Board Annual Meeting*, Washington, D.C., Jan. 11–15, 2009.

14 BeenaVision. *WheelView: Accurate, Rugged, Reliable, Automatic Wheel Profile Measurement System*. Available online at http://www.beenavision.com/products_wheelview.html, Accessed Mar. 20, 2012.

15 ENSCO. *Rail Surface and Wear Condition Measurement Systems*. Available online at http://www.ensco.com/index.cfm?page=415, Accessed Mar. 20, 2012.

16 Li, Y., C. Otto, N. Haas, Y. Fujiki, and S. Pankanti. Component-Based Track Inspection Using Machine-Vision Technology. In *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*, ACM New York, NY, USA 2011.

17 Shah, M. *Automated Visual Inspection/Detection of Railroad Track*. Final Project Report BD550-08. Computer Vision Lab, University of Central Florida, Orlando, Florida, 2010.

18 Camargo, L. F. M., J.R. Edwards, and C.P.L. Barkan. Emerging Condition Monitoring Technologies for Railway Track Components and Special Trackwork. In *Proceedings of the 2011 ASME/ASCE/IEEE Joint Rail Conference*, Pueblo, Colorado, USA, Mar. 16-18, 2011.

19 Williams, K., M. J. Olsen, G.V. Roe, C. Glennie, Synthesis of Transportation Applications Of Mobile LiDAR. In *Journal of Remote Sensing*. Vol. 5, Issue 9, 2013, pp. 4652-4692.

20 Nelson, R.. How Did We Get Here? An Early History of Forestry Lidar[1]. In *Canadian Journal of Remote Sensing*. Vol. 39, No. S1, 2013, pp. S6-S17.

21 Bergen, K. M., S. J. Goetz, R. O. Dubayah, G. M. Henebry, C. T. Hunsakar, M. L. Imhoff, R. F. Nelson, G. G. Parker and V.C. Radeloff . Remote Sensing of Vegetation 3-D Structure for Biodiversity and Habitat: Review and Implications for Lidar and Radar Spaceborne Missions. In *Journal of Geophysical Research: Biogeosciences*. Vol. 114, Issue G2, 2009.

22 Barber, D., J. Mills, and S. Smith-Voysey, Geometric Validation of a Ground-Based Mobile Laser Scanning System. In *ISPRS Journal of Photogrammetry and Remote Sensing*. Vol. 63, Issue 1, 2008, pp. 128-141.

23 Heritage, G. and A. Large, *Laser Scanning for the Environmental Sciences*. John Wiley & Sons, Inc., Newyork, 2009.

24 Lu, L., C. Ordonez, E. M. DuPont. Terrain Surface Classification for Autonomous Ground Vehicles Using a 2D Laser Stripe-Based Structured Light Sensor. In *International Conference Intelligent Robots and Systems*. (IEEE/RSJ), 2009.

25 FENG, J.-Y., Q.-B. Feng, and C.-F. KUANG, Present Status of High Precision Laser Displacement Sensor Based on Triangulation. In *Journal of Applied Optics*. Vol 3, 2004.

26 Dorsch, R.G., G. Häusler, and J.M. Herrmann, Laser Triangulation: Fundamental Uncertainty in Distance Measurement. In *Journal of Applied Optics*, Vol. 33, No. 7, 1994, pp. 1306-1314.

27 Shah, M., *Automated Visual Inspection/Detection of Railroad Track*. Final report, FDOT Project No. BD550-08, July, 2010.

28 Mertz, C., J. Kozar, J. R. Miller and C. Thorpe. Eye-Safe Laser Line Striper for Outside Use. In *Intelligent Vehicle Symposium*, Vol. 2, Jun. 2002, pp. 507-512.

29 Laurent, J., D. Lefebvre, and E. Samson. Development of a New 3D Transverse Laser Profiling System for the Automatic Measurement of Road Cracks. In *6th Symposium on Pavement Surface Characteristics*, Portoroz, Slovenia. Oct. 2008.

30 Laurent, J. and J.-F. Hébert. High Performance 3D Sensors for the Characterization of Road Surface Defects. In *IAPR Workshop on Machine Vision Applications*. Dec., 2002.

31 Laurent, J., J. F. Hebert, D. Lefebvre, Y. Savard. Using 3D Laser Profiling Sensors for the Automated Measurement of Road Surface Conditions. In *7th RILEM International Conference on Cracking in Pavements*. Vol. 4, 2012, pp. 157-167.

32 Tsai, Y.-C., C. Jiang, and Y. Huang, Multiscale Crack Fundamental Element Model for Real-World Pavement Crack Classification. In *Journal of Computing in Civil Engineering*, Vol.28, No. 4, 2012.

33 Hartley, R. and A. Zisserman, *Multiple View Geometry in Computer Vision*, Second Edition, Cambridge University Press, Mar. 2004.

34 Niblack, W., *An Introduction to Digital Image Processing*. Strandberg Publishing Company, Birkeroed, Denmark, 1985.

35 Umbaugh, S.E., *Computer Imaging: Digital Image Analysis and Processing*. CRC Press, 2005.

36 Dougherty, E.R. and R.A. Lotufo, *Hands-On Morphological Image Processing*. SPIE Publications, 2003.

37 Canny, J., A Computational Approach to Edge Detection. Pattern Analysis and Machine Intelligence, In *Pattern Analysis and Machine Intelligence*, IEEE Transactions on Vol. PAMI-8, Isuue 6, 1986, p. 679-698.

38 Jähne, B., H. Scharr, and S. Körkel, *Principles of Filter Design. Handbook of Computer Vision and Applications*, Academic Press, 1999, pp. 125-151.

39 Jol, H. M., *Ground Penetrating Radar: Theory and Applications*, Oxford: Elsevier Science, 2009.

40 *Data Analysis and Measurement for GPR and Roadway Instrumentation Systems*, Publication Electrical & Computer Engineering Department, University of Florida, Gainesville, 2000.

41 *Ground Penetrating Radar Signal Processing Enhancements for Geophysical Anomaly Identification*. Electronic Communications Laboratory, University of Florida, Gainesville, FL, 1999.

42 Leng, Z. and I. L. Al-Qadi, Railroad Ballast Evaluation Using Ground-Penetrating Radar: Laboratory Investigation and Field Validation. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2159, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 110-117.

43 Loulizi, A., I. L. Al-Qadi and S. Lahouar, Ground-Penetrating Radar Signal Modeling to Assess Concrete Structures, In *ACI Materials Journal*, Vol. 99 Issue 3, May2002, pp. 282 – 291.

44 Maser, K. R. *Ground Penetrating Radar Survery of Pavement Thickness on MN/Road Sections*. Static Highway Research Program, National Council, INFRASENSE, INC, 1994

45 Liu, L. , *GPR for Fast Pavement Assessment*, Publication of University of Connecticut, Storrs, CT, May, 2007.

46 Al-Qadi, I. L. , S. Lahouar and A. Loulizi, *Ground-Penetrating Radar Calibration at the Virginia Smart Road and Signal Analysis to Improve Prediction of Flexible Pavement Layer*

Thickness. VCTIR Report NTIS, Springfield, VA, 2005.

47 Roberts, R., I. L. Al-Qadi, E. Tutumluer and J. Boyle. *Subsurface Evaluation of Railway Track Using Ground Penetrating Radar*. National Technical Information Service, Springfield, VA, 2008.

48 Carlson, A. B. and P. B. Crilly, *Communication Systems: An Introduction to Signals and Noise in Electrical Communication*, McGraw Hill Inc., New York, 2010, pp. 412-422, (Appendix A).

49 Waters, J., E. Selig. *Track Geotechnology and Substructure Management*. Thomas Telford Ltd, 1994.

50 El-Sibaie, M., W. GeMeiner, D. Clark, L. Al-Nazer, R. Arnold, S. Farritor, M. Fateh, S. Lu, and G. Carr. Measurement of Vertical Track Modulus: Field Testing, Repeatability, and Effects of Track Geometry. In *IEEE/ASME/ASCE 2008 Joint Rail Conference*, American Society of Mechanical Engineers, 2008, pp. 151-158