Chapter 3 - Data Processing Algorithms

3.1 Processing algorithms for evaluating PERES data

Bridge inspectors will need a quick, accurate analysis method to evaluate data collected from concrete bridge decks using the PERES GPR system. Developing this methodology is a difficult task due to the large amounts of data that are generated, (10 megabytes/m² of bridge deck coverage), by the system and the difficulties involved in consistently interpreting the data. Algorithms that use statistical pattern recognition and image processing techniques provide the means to analyze PERES data with these practical realities in mind. A method that incorporates both of these techniques has been developed to perform two primary tasks. First, the reinforcing steel in the concrete bridge decks is detected using an image processing algorithm that can also be labeled as a syntactic pattern recognition algorithm. Second, the locations of potential flaws and sound concrete are detected using a statistical pattern recognition algorithm. The results of both of these algorithms are combined to provide an image of high likelihood distress locations.

3.1.1 Expert knowledge for processing algorithm development

A significant prerequisite for developing image processing and statistical pattern recognition algorithms is expert knowledge about features of interest in data. For PERES data from aging concrete bridge decks these features include responses to distress, (such as delaminations), and reinforcing steel. Based on repeated observations, feature characteristics can be determined that distinguish a feature from its background. These feature characteristics are called “measurements.” Measurements provide input to image processing and pattern recognition algorithms that classify data, (section 1.1.5). The error rates of these techniques depend on how well measurements separate data into
distinct, correct classes in spite of noise and clutter in the data. Selection of appropriate measurements prepares the way for effective pattern recognition techniques. The measurements used for analysis of PERES data are described in Sections 3.3.1 and 3.3.2.

3.2 Pattern recognition and image processing algorithm implementation fundamentals

Developing pattern recognition and image processing algorithms tailored to PERES GPR data analysis required methods developed in the literature and original methods that had not been used before. Template matching and quadratic discriminant analysis were two of the key methods used. Fundamental concepts of these two methods, as they relate to algorithm implementation, are provided in Sections 3.2.1 through 3.2.3. Details of how these methods are used to analyze GPR data are in Section 3.3.

3.2.1 Image processing

Image processing techniques provide important tools that prepare images for effective pattern recognition processing. In some cases the lines between image processing and pattern recognition processing are blurred. For this study, the approach to locating reinforcing steel in PERES bridge deck data could be categorized as image processing or a simple form of syntactic pattern recognition. The approach involves a two dimensional convolution of the image with a template and a subsequent search for local maxima in the resulting image (Section 3.3.1). Local maxima are generated for long, linear features with a particular orientation, which is indicative of a response to reinforcing steel in the PERES data. The template matching concept this technique relies on, (which the pattern recognition measurements for detecting distress locations also depend on), is described in the following section.

3.2.2 Template matching

To effectively locate objects and classify features in image data, a frequent approach has been to use template matching. This approach defines a template which is
representative of the object in the image that needs to be located. This template is compared to candidate objects in the data using a match measure. Objects that closely match the template produce a high value for the match measure, while objects that differ significantly have a low match measure. In one dimension the equations for template matching are provided in equation 3.1 (Jain, 1995).

\[
C_{fg}[i] = \sum_{k=1}^{m} g[k] f[i+k]
\]

\[
M[i] = \frac{C_{fg}[i]}{(\sum_{k=1}^{m} f^2[i+k])^{1/2}}
\]

Here \(C_{fg}\) is the cross-correlation between a template, \(g\), and a signal, \(f\). The match measure \(M\) is calculated by normalizing the cross-correlation using signal values. A simple one-dimensional example of template matching can be demonstrated with the following signals, (Figure 3.1):
Signal 1 displays an abrupt spike at time=4, and has a value of zero everywhere else. The similarity of signals 2 and 3 to signal 1 can be calculated using a match measure. First, $M$ is calculated for signal 1 itself, which produces a value of $M=2.0$. Signals 2 and 3 produce values $M=1.0$ and $M=1.79$ respectively, when match measures are computed.

Figure 3.1 Template matching example signals.
Signal 3, which has a match measure that is close to 2.0, is similar to signal 1 because it has an abrupt spike near time=4. The similarity between the two signals is reflected in the close match measures. Signal 2 is a poor match to signal 1, with high magnitude values that form a plateau rather than a spike. This is indicated by a low match measure. If appropriate templates can be defined as expected signals, this example illustrates that match measures can provide a means of determining the similarity of an unknown signal to a reference template. This template matching operation can be performed on two dimensional images in a similar way to this one dimensional example. The one dimensional equations are used to calculate measurements for input to the pattern recognition algorithm that determines distress locations (Section 3.3.2), while the two dimensional equations are used for locating the reinforcing steel (Section 3.3.1).

3.2.3 Quadratic discriminant analysis

Quadratic discriminant analysis is a pattern recognition technique which classifies data based on statistical distributions of measurement data. Ideally, the technique is used on measurement data that is normally distributed, (which makes it the minimum error rate classifier), but it can also be used on data that approximates a normal distribution. The effectiveness of the method depends on the distribution of the data relative to the normal model and the distances between the means of class measurements. The quadratic discriminant analysis classifier calculates the locations of decision boundaries in measurement space based on the locations of distribution intersections in the measurement data, (using a normal model). The locations of these intersections can be defined using the following equation (Duda, 1973):

\[
g_i(x) = -\frac{1}{2} (x - \mu_i)^t \sum_i^{-1} (x - \mu_i) - \frac{1}{2} \log |\sum_i| + \log P(\omega_i) \tag{3.2}
\]

Here \( x \) is the vector of measurement values, \( \mu \) is the mean value of the \( i \)th measurement class, \( P(\omega_i) \) is the a priori probability of that nature is in the state \( \omega \), and \( \Sigma \) is defined as the matrix of standard deviations with elements defined by the expected value of the class measurement values and class means as follows:
\[ \sigma_j = E[(x_i - \mu_j)(x_j - \mu_j)] \] (3.3)

The decision surfaces defined in equation 3.2 are hyperquadrics, which can take the form of hyperplanes, hyperspheres, hyperellipsoids, hyperparaboloids and hyperhyperboloids (Duda, 1973). A representative example of a hyperelliptical decision surface for the bivariate normal case is shown as a projection in Figure 3.2. In this figure, the green and red hyperellipsoids are contours of constant probability density for each of two classes, shown from above. The hyperelliptical decision surface, (shown in blue), indicates the location of the intersection of the two probability densities. The diagonal line fill inside the blue hyperellipsoid is on one side of the decision surface and signifies one decision region, while the dot filled area is on the other side of the decision surface signifies the second decision region. Other hyperquadric decision surfaces can be similarly represented.

For many pattern recognition classification problems, quadratic discriminant analysis parameterizes decision boundaries that are highly effective. In some cases, however, the distribution of the data does not fit the normal model. In these cases the optimum decision surface can be more complicated than a hyperquadric. Neural network methods and other sophisticated techniques can sometimes be used to parameterize these more complicated decision surfaces.
3.3 Algorithm description and implementation

The implementation of a processing method for analysis of PERES data from concrete bridge decks was accomplished using a collection of algorithms written in C, MatLab and the Partek software package. This collection of algorithms was designed to determine the locations of distress, reinforcing steel and sound concrete in the data. The C algorithms were primarily used to prepare data for analysis in MatLab and Partek. Partek was used to implement a quadratic discriminant analysis of the data, which determined high likelihood locations of distress in the data. Matlab was used to implement an image processing technique for detecting locations of reinforcing steel in the data and to subsequently integrate and display the distress and reinforcing steel detection results.

3.3.1 Reinforcing steel detection algorithm

The motivation for detecting and locating reinforcing steel, (particularly by automated means), in bridge decks is well established for condition assessment purposes, (Section 1.1.1). Corrosion damage initiates at the reinforcing steel and frequently results in delamination cracking. Reinforcing steel is therefore an important feature of most concrete bridge decks and knowledge about its location is useful.

PERES is capable of detecting reinforcing steel in many concrete bridge decks due to the differences in dielectric properties between concrete and steel (Section 1.3.1). For these bridge decks, PERES data can be used to produce images of the reinforcing steel locations inside the concrete deck. Automating the classification of PERES responses to reinforcing steel in concrete is desirable for distinguishing between responses to distress, reinforcing steel and sound concrete.

A series of processing steps is used to accomplish the classification of responses to reinforcing steel in PERES data collected from bridge decks. The first two steps locate the depth of the dominant responses to reinforcing steel in the three-dimensional tomographic data. The third step analyzes two-dimensional images to determine the location and orientation of the reinforcing steel in the PERES data. Data from bridge
deck R7 will be presented along with the description of the algorithm to illustrate the results that the technique produces.

The first step in locating reinforcing steel in the radar data measures the contrast in all of the plan view layers of tomographic data from PERES. A three-dimensional illustration of a section of PERES data collected from bridge deck R7 is presented in Figure 3.3 to clarify the description of the analysis. In the illustration, the three-dimensional structure of the data can be observed and indications of the reinforcing steel can be distinguished as predominantly yellow areas. The variance of the magnitude response in each plan view tomographic layer of the data is calculated, where each layer is at a given depth as shown in Figure 3.4, (example data layer shown in red). Results from the variance calculation performed on PERES data from bridge deck R7 are
Figure 3.4 Representation of example data layer sampled at constant depth (red).

Figure 3.5 Variance values calculated for each 1 mm thick tomographic layer, (concrete surface at 80 mm).
presented in Figure 3.5. The three local maxima between layers at 100 mm and 200 mm correspond to the locations of the dominant responses to rebar. These three local maxima in the variance plot indicate the high contrast in these image layers relative to neighboring layers. Of the local maxima in the plot, these three local maxima can be identified based on phase information in the radar data.

Reconstructed PERES data consists of complex numbers, which carry both magnitude and phase information. Thus far in the discussion of reconstructed PERES data, the magnitude information has been used to visualize and describe PERES responses. The phase of the PERES response is also useful because a phase shift of 180 degrees is observed in the response when a radar pulse is reflected at an interface between a dielectric and a metal (Shadowitz, 1975). In bridge decks, these interfaces occur between concrete, (a dielectric), and the reinforcing steel, (a metal). Figure 3.6 shows a Hilbert transformed image of the phase response at a concrete depth of 3.9 cm, (corresponding to the first of the three peaks in the variance plot where the responses to reinforcing steel are dominant). The Hilbert transform converts phase to a binary quantity, where the result is $-1$ for phase from 0 to $\pi$ and $+1$ for phase from $\pi$ to $2\pi$. The white regions in Figure 3.6 correspond to values of $-1$ and the black regions correspond to

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure36.png}
\caption{Hilbert transform image of the phase response at depth=3.9cm in bridge deck R7 data.}
\end{figure}
values of 1. This image and the other phase images selected based on local maxima in the variance plot can be used to locate the three layers that have dominant responses to reinforcing steel. These three layers provide representations of reinforcing steel oriented top to bottom in Figure 3.6, left to right in Figure 3.7, and a combination of the two orientations corresponding to the third local maximum. All three of these selected local maxima are responses to the top mat of reinforcing steel in the deck, which contains reinforcing steel oriented both parallel and transverse to the direction of traffic flow over the deck. (Appendix C, Figure 7). The alternating changes of phase in these images are counted in both the parallel and transverse direction to locate the reinforcing steel layers. The frequent changes of phase in the layers containing dominant reinforcing steel responses allow this to be used as a criterion for locating these layers. The three layers with the largest number of alternations in the parallel and transverse directions respectively are stored for further analysis.

Next, a template matching operation is performed in two dimensions using variations on the filter shown in Figure 3.8. This filter was designed by tailoring parameters of a Gaussian distribution based on observed reinforcing steel responses in data and subsequent testing of the filter shape on PERES data. The filter mimics the
profile of the reconstructed magnitude response PERES produces in the plane of the reinforcing steel. For typical reconstructed data taken with PERES over reinforcing steel in concrete, a high magnitude response is produced by the reflection from the top surface of the reinforcing steel. The magnitude of the response tapers where the reconstructed reflections originate from the oblique surfaces of the reinforcing steel. This profile is observed along the length of the reinforcing steel, as the filter suggests, but is generally interrupted by orthogonal crossing reinforcing steel and noise in the actual data. The characteristic reinforcing steel profile is most clearly represented in the constant depth tomographic layers of PERES data corresponding to the locations of reinforcing steel.

The filter in Figure 3.8 is first used to select the best two candidate layers in the data for further analysis. One layer is selected for locating reinforcing steel in the traffic direction and the other layer is selected for locating reinforcing steel in the transverse direction. These layers are selected by calculating the responses to the full width filter in each of the two directions, (traffic direction and transverse), using template matching in two dimensions. The width of the filter is the same as the width dimension of the data when the filter is used to detect reinforcing steel in the traffic direction. The width of the filter is identical to the length dimension of the data when the filter is used to detect reinforcing steel in the transverse direction. The tomographic layers filtered using this

![Figure 3.8 Filter for reinforcing steel template matching.](image)
process are the three candidate layers previously selected using phase information. The images that produce the largest average amplitude in this filtering process are selected for locating reinforcing steel in the traffic direction and the transverse direction.

The appropriate constant depth layers have been selected for use in calculating reinforcing steel locations in PERES data, based on the procedure that has been described thus far. The final step required to parameterize the location of the reinforcing steel uses the template matching filter from Figure 3.8 again. A description of the calculations for locating reinforcing steel in the traffic direction follows.

First, a template matching filter with half the width of the complete image is convolved, from left to right across the top half of the image. Results from this convolution are normalized based on the template matching formulation, (Equation 3.1), and local maxima in the result correspond to template matches, (Figure 3.9). An

![Figure 3.9 Template matching result from the top half of the image.](image)
Figure 3.10 Template matching result from the bottom half of the image.

Figure 3.11 Detected reinforcing steel locations shown in red, (traffic direction).
identical operation is performed on the lower half of the image, which produces the result in Figure 3.10. The local maxima from the upper and lower filters are computed, which provides two points along the line each piece of reinforcing steel lies on. The points that are nearest neighbors in the corresponding filters are assigned to identify each piece of reinforcing steel. Detected reinforcing steel locations can then be plotted (Figure 3.11). In Figure 3.11, the locations of detected reinforcing steel are plotted as red lines over an original image of the layer located at 3.9 cm in depth, (beneath the concrete surface), from reconstructed PERES data collected over bridge deck R7. The color scaling in the figure ranges from black for small magnitude responses and ranges up in gray scale values to white for large magnitude responses.

A similar procedure is repeated to detect reinforcing steel in the transverse direction, but this time the filter is convolved with an image from top to bottom. The layer 7.4 cm below the concrete surface was selected from the reconstructed PERES data collected over bridge deck R7 based on the procedure previously described. The results in Figure 3.12 are obtained. The performance of the reinforcing steel detection algorithm is quite good, based on these two example illustrations. All of the reinforcing steel in the top mat, (in both the traffic and transverse directions) are accounted for based on the detection algorithm. The orientation of two pieces of reinforcing steel are
incorrect in the traffic direction (as observed in Figure 3.11), but this is due to ambiguities in the reconstructed data. Two of the detected pieces of reinforcing steel in Figure 3.12 are false detections. One of these false detections is entirely incorrect, (the third red line from the top), but the second is due to artifacts in the data introduced by metal supports for a wooden structure used to position PERES over the bridge deck, (bottom red line). The performance of the reinforcing steel detection algorithm will be discussed in more depth in the following sections and in Chapter 4, Results.

3.3.2 Distress detection algorithm

Bridge inspectors need accurate, efficient methods for evaluating PERES data from bridge decks. In particular, a fast way of locating indications of distress in the data makes it much more usable. Developing an algorithm to perform this task requires a combination of knowledge about bridge deck distress modes, the expected responses to the distress modes in the PERES data and a method for selectively classifying these responses. A pattern recognition classifier designed to address these issues for this study is described in this section. A discussion of error rates for the classifier and the types of data the classifier can process is included in Chapter 4.

A crucial step toward developing a pattern recognition classifier for detecting distress in PERES data collected from concrete bridge decks is to gain and use expert knowledge about how the distress modes initiate and propagate. Delamination cracking due to corrosion of reinforcing steel is the dominant distress mode in bridge decks (Section 1.1.1). This damage initiates at the corroding reinforcing steel and propagates through the concrete preferentially in the plane of the reinforcing steel mesh, below the concrete surface. This is valuable information because it provides a linkage between the location of distress damage and a consistent feature, (almost all concrete bridge decks are designed and built with reinforcing steel), in PERES data. Furthermore, reflections from reinforcing steel in concrete are generally very strong in PERES data due to the interface between a dielectric and a metal at the concrete, reinforcing steel interface (Section 1.3.1). The proximity and orientation of delamination cracking relative to reinforcing steel in concrete forms the basis for the design of the pattern recognition algorithm for distress detection used in this study.
An algorithm for determining the location of reinforcing steel in PERES data has been developed, (Section 3.3.1) which can provide significant information to a distress detection algorithm, because a relationship between reinforcing steel and a common distress mode, (corrosion-induced delamination) has been established. This relationship can be utilized by examining the individual waveforms that make up reconstructed PERES data. Waveforms that pass through a significant feature, such as delamination distress or reinforcing steel display a strong response in the reconstructed data (Figure 3.13). The distress waveform in Figure 3.13 was obtained from reconstructed PERES data by averaging one hundred individual waveforms that passed through a simulated delamination distress, (the simulated distress was a 2.54 cm thick foam insert in bridge deck R13). Indications of the distress in the waveform include the strong peak at approximately 12 cm, (6 cm below the concrete surface), and a second peak at approximately 23 cm. The first peak corresponds to the location of the distress itself, while the second peak is clutter that results from internal reflections between the simulated distress and the surface of the bridge deck. A representative response to sound
concrete was also produced by averaging one hundred waveforms from deck locations corresponding to sound concrete in the reconstructed PERES data, (Figure 3.13). Very small variations in the sound concrete response are observed relative to the response to distress. These results indicate that a clear distinction can be made between reconstructed waveforms corresponding to distress and concrete. Results from other reconstructed PERES data including a 0.6 cm thick synthetic distress and a delamination, (confirmed using the chain drag test for bridge deck R12), show that similar responses are obtained. For each response a high magnitude peak, (that corresponds to the distress), is followed by a second peak that is caused by reflections of the radar pulse, (between the distress and the antenna). These results were all obtained for concrete decks without an asphalt overlay. Comparable results have been obtained in reconstructed GPR data by Mast (1993).

The consistency of the PERES responses to distress, along with data collected by Mast (1993), suggests that a method which detects key features from these responses will have potential. Based on the relationship between the location of reinforcing steel in bridge decks and delamination distress, the PERES response to distress can be predicted qualitatively. A template has been developed which matches the profile of an expected response to distress in reconstructed data (Figure 3.14). In practice, the location of the
first peak response, (going from shallow to deeper depths), in the template is adjusted based on the position of the top mat of reinforcing steel in the data, as determined by the reinforcing steel detection algorithm. The second peak, corresponding to ringing between the antenna and distress in the deck, is positioned at twice the depth of the first peak. The sound concrete template is defined by a constant magnitude of 5 for all depths. The magnitudes of template values are selected based on average responses to simulated distress and sound concrete in bridge deck R13, and are tuned based on pattern recognition results from bridge deck R13. Relative magnitudes of template values determine the response to the template.

Two match measurements are calculated for each waveform in a PERES data set using Equation 3.1. These two match measurements are based on the distress and sound concrete templates, respectively. Match measurements are used as input to a statistical pattern recognition algorithm. This algorithm is implemented in the Partek software package and uses a quadratic discriminant analysis classifier to categorize input data. Training data for the classifier was obtained from bridge deck R13. 1826 waveforms were selected for the training data, where 1029 were known to come from locations corresponding to sound concrete and 797 were known to come from locations.
corresponding to distress. Figure 3.15 presents a plot of the training data in measurement space after the classifier was implemented. In the figure, a vector is drawn between each class mean and the categorized data points. Two distinct clusters can be observed in the measurement space corresponding to the distress and sound concrete waveforms respectively. Errors the classifier makes near the boundary of the decision surface are indicated in blue. A priori probabilities were 0.5 for both classifications. A Summary of the results from the classifier applied to the training data is presented in Table 3.1.

After training, the classifier can be applied to an entire data set to determine where probable locations of distress are. These results must be combined with information from the reinforcing steel detection algorithm to make an effective analysis, which will be discussed in section 3.3.3. The initial results from the pattern recognition classifier are presented as an image in Figure 3.16 for bridge deck R13, (as an example).
Table 3.1 Pattern recognition classification for training data (using full leave one out method).
This binary image displays a plan view of deck areas classified as distress in white and areas classified as sound concrete in black. A significant amount of noise can be observed in this image, which can be removed using a size filter. This filter removes four connected (Jain, 1995) regions corresponding to detected distress from the image if they have fewer than fifteen connected picture elements, (each picture element is \(1 \text{ cm}^2\)). After removing these noisy areas in the image, Figure 3.17 results.

In this image, the two locations where PERES detected simulated distress
(Figure 1.12) each display significant responses based on the pattern recognition classifier. Other areas in the data which have been initially classified as distress are actually responses to reinforcing steel, (which will be accounted for in Section 3.3.3 based on the reinforcing steel detection algorithm), noise and clutter. Results from other bridge deck sections and field data will be presented in Chapter 4. The method for combining the results of the reinforcing steel detection algorithm and the distress detection algorithm are discussed in the following section.

3.3.3 Combining results

The results from algorithms for detecting reinforcing steel and distress in bridge decks are useful individually, but combining the output from both is important for interpreting the results appropriately. Before describing how the output from these two algorithms are combined, the reinforcing steel detection results from bridge deck R13 are presented here in Figure 3.18 and 3.19. Overall, the reinforcing steel detection results are reasonable and, in particular, all but one piece of reinforcing steel is detected in the traffic direction. However, there are also falsely detected pieces of reinforcing steel in the

![Image](image_url)

*Figure 3.18 Reinforcing steel detection in the traffic direction.*
traffic direction. This can be attributed in part to noise, but is also influenced by the distress indications in the data. The reinforcing steel filter is designed to reject these localized distress indications, but there is an influence on algorithm performance. Similar results for detecting reinforcing steel in the transverse direction show that the algorithm performance is effective, but is influenced by noise and distress indications. The effects of these influences will be discussed in Chapter 4, Results.

The output from the reinforcing steel detection algorithm and the distress detection algorithm can be assessed together if the influence of the reinforcing steel on the reconstructed data is recognized. The reinforcing steel in the bridge deck produces a response in the reconstructed PERES data, which is evident in the location of the reinforcing steel and a region around it. Based on observations of the response to reinforcing steel in reconstructed data from bridge decks R7 through R13, the region of influence was established as 3.5 cm in both plan view directions orthogonal to the length axis of the reinforcing steel. Beyond this region of influence, average peak amplitudes in the data are less than 2/5 of the maximum peak amplitude caused by adjacent reinforcing steel. Inside the region of influence, the average peak amplitudes are above this threshold. To identify waveforms corresponding to distress accurately, waveforms

Figure 3.19 Reinforcing steel detection in the transverse direction.
within the region of influence of the reinforcing steel are excluded from consideration as areas corresponding to distress. This prevents reinforcing steel, (which can produce similar responses to distress and simulated distress), from being inappropriately categorized as distress.

Once these areas have been excluded from consideration as distress, a size filter is applied to locate the largest areas where distress is detected. The three largest areas where distress is detected, after areas in the region of influence of the reinforcing steel are excluded from consideration, are shown in Figure 3.20. Two of these areas correspond to the location of a simulated distress, (in the red circle), as shown in Figure 1.7. The third area in Figure 3.20 is a result of noise and clutter in the data. Using a lower threshold on the size filter, the thirty largest areas in the analyzed R13 data, (shown in Figure 3.21), include a third response to the simulated distress detected in the left portion of the deck. Figure 3.21 also includes three areas that correspond to the simulated distress in the lower middle part of the deck section, circled in red and a single response to the simulated distress in the upper middle portion of the deck, also circled in red. A discussion of distress detection and problems associated with noise and clutter in PERES data will be presented in Chapter 4, Results.
Figure 3.20  Three largest detected distress areas (deck R13).

Figure 3.21  Thirty largest detected distress areas (deck R13).