Automated Visual Inspection of Railroad Tracks

Esther Resendiz, *Member, IEEE*, John M. Hart, and Narendra Ahuja, *Fellow, IEEE*

***Abstract*—Thousands of miles of railroad track must be inspected twice weekly by a human inspector to maintain safety standards. A computer vision system, consisting of field-acquired video and subsequent analysis, could improve the efficiency of the current methods. Such a system is prototyped, and the following challenges are addressed: the detection, segmentation, and defect assessment of track components whose appearance vary across different tracks and the identification and inspection of special track areas such as track turnouts. An algorithm that utilizes the periodic manner in which track components repeat in an inspection video is developed. Spectral estimation and signal-processing methods are used to provide robust detection of the periodically occurring track components. Results are demonstrated on field-acquired images and video.**

***Index Terms*—Railroad track inspection, spectral estimation.**

I. Introduction

COMPUTER vision has recently been applied to several railroad applications due to its potential to improve the efficiency, objectivity, and accuracy when analyzing large databases of acquired video footage and images. Algorithms can potentially provide a more objective assessment of track conditions than human inspectors. However, it is difficult to create an algorithm that is robust to numerous unforeseen conditions. Spatial templates and other application-specific detection methods can be developed to accomplish specific inspection tasks [1]–[8]. However, there is great value in creating a general method to inspect components without prior knowledge of component appearance. By detecting periodic components without prior knowledge of spatial appearance, a computer vision system may one day perform track inspection over thousands of miles of track with minimal human involvement.

Periodically occurring components are often encountered in infrastructure inspection. For example, a railroad track is composed of many individual ties, and a train is composed of individual railcars. Most repeating components are similar to each other but not identical due to various manufacturing differences and environmental conditions. Railroads are vital to the infrastructure of most countries, but many inspection tasks are performed manually by a human inspector. Computer vision algorithms are useful in several railroad tasks, including track inspection [1]–[8], ballast inspection [9], rail profile measurement [10], safety appliance inspection [11], and monitoring intermodal transport [12]–[14]. Other problems involving infrastructure inspection could also benefit from computer vision.

Manuscript received November 12, 2011; revised March 21, 2012, June 10, 2012, and November 2, 2012; accepted December 6, 2012. Date of publication January 24, 2013; date of current version May 29, 2013. This work was supported in part by the Office of Naval Research under Grant N00014-09-1-001, by the National Science Foundation under Grant IIS 08-12188, by the American Association of Railroads' Technology Scanning Program, and by the NEXTRANS Center. Additionally, the Rail Transportation and Engineering Center of the University of Illinois at Urbana–Champaign provided technology development support. The Associate Editor for this paper was B. Ning.

The authors are with the Department of Electrical and Computer Engineering, University of Illinois at Urbana–Champaign, Champaign, IL 61801 USA (e-mail: esther.i.resendiz@gmail.com; jmhart3@illinois.edu; n-ahuja@illinois.edu).

Digital Object Identifier 10.1109/TITS.2012.2236555

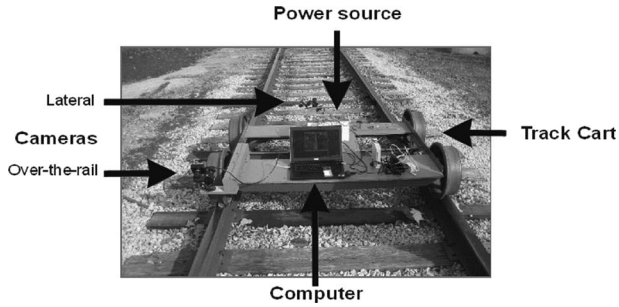


Fig. 1. Track cart.

(a) (b)

Fig. 2. (a) Lateral view of the track. (b) Over-the-rail view of the track.

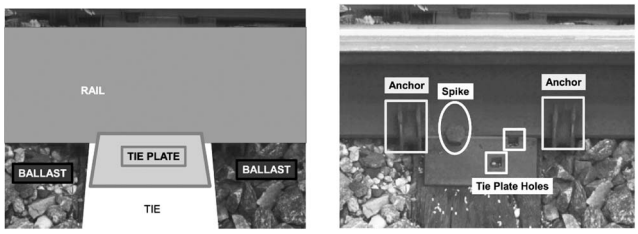
A. Track Inspection

The Federal Railroad Administration (FRA) requires track to be inspected for physical defects at specified time intervals, which may be as often as twice per week [3], [4]. Computer vision could potentially supplement the current manual inspection process due to its ability to objectively process large amounts of video and image data. Recently, a track cart has been developed to acquire track inspection video. This track cart, shown in Fig. 1, captures video of a railroad track with off-the-shelf cameras, and records these data to a laptop. Railroad track inspection algorithms are developed to inspect the image and video data for defective track components.

Fig. 2 shows the two camera viewpoints used for track inspection. Fig. 2(a) shows the lateral viewpoint, where the side of the track is visible. Fig. 2(b) shows the over-the-rail viewpoint, where both sides of the track are visible.

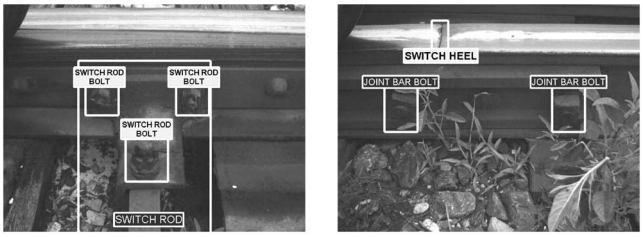
B. Components

Track components are shown in Fig. 3. Fig. 3(a) delineates the largest components. The rail is in the top half of the image.



(a) (b)

Fig. 3. (a) Localization of rail, ballast, tie, and tie plate. (b) Localization of spikes, tie plate holes, and anchor.



(a) (b)

Fig. 4. Turnout components. (a) Switch rod and its bolts. (b) Switch heel and joint bar bolts.

The rail is the part of the track along which the train wheels move. The wooden tie is delineated with a white trapezoid, and the ties are oriented perpendicularly to the rail. The steel tie plate is delineated with a gray trapezoid. Tie plates are placed between the tie and the rail where they intersect and hold the rail to the tie. The ballast, which is labeled on the left and right sides of the tie, is composed of small rocks.

The following objects are localized in Fig. 3(b): one spike (shown here in an ellipse), two tie plate holes (shown in the squares), and two anchors (shown in the rectangles). Spikes are hammered into the tie plate to keep the rail in place and at the proper gauge. Rail anchors secure the rail from moving perpendicular to the tie. The components shown in Fig. 3 are commonly inspected for compliance with FRA regulations. Spatial templates were used to detect these components in [1]–[4], and edge-based methods were used to detect track components in [5]–[8].

1) Turnouts: At certain locations in a track, there is convergence as tracks join each other and divergence as one track forks into two. These areas of the track are known as turnouts or switch areas, and defects in the switch area will frequently result in an accident [1]. Fig. 4 shows some of the components in turnout inspection. These are the switch rod and its bolts [see Fig. 4(a)], and the switch heel and joint bar bolts [see Fig. 4(b)].

C. Related Work in Spectral Estimation

This paper combines Gabor filters with spectral estimation to detect periodically occurring objects. The Gabor filters transform an image into directionally filtered versions [15], and Gabor-filtered images can be used to discriminate various textures, such as those found in outdoor scene classification [16]. In this paper, the filtered images are combined into 1-D signals, and spectral estimation is used. Specifically, the multiple signal classification (MUSIC) algorithm is used for its ability to detect multiple periodicities in the presence of noise. Similar spectral estimation techniques have recently been applied to vehicular congestion classification [17] by analyzing acoustic signals and to symmetry detection in images [18] by analyzing angular correlation. In both [17] and [18], spectral estimation algorithms were effective at detecting signals in noisy data (such as busy traffic scenes and various images).

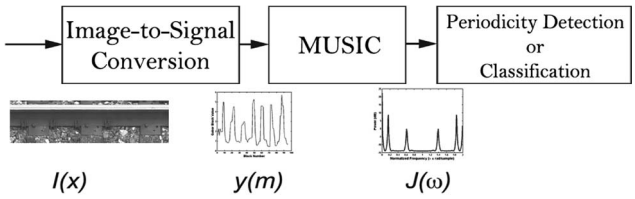


Fig. 5. Algorithm overview.

This paper is organized as follows. Section II presents an overview for all three algorithms for automated inspection of railroad components. MUSIC is also introduced. Section III describes the first of the algorithms, i.e., component detection for railroad track. In Section IV, turnout area detection is demonstrated. Section V describes tie detection in inspection video using custom filters based on Gabor textures. In Section VI, experiments are conducted on a data set collected from the track cart for all three algorithms. Section VII concludes and provides future directions for this paper.

II. Overview

This paper presents three algorithms: one for component detection in railroad track (see Section III), one for turnout detection (see Section IV), and one for tie detection (see Section V). These algorithms all follow the three steps outlined in Fig. 5.

The specific image-to-signal conversion and the final periodicity detection or classification varies for each algorithm. $I(x)$ is the input image or video to the image-to-signal conversion, $y(m)$ is the 1-D signal that will be used for MUSIC, and $J(\omega)$ is the power spectrum of periodicities detected, where the peaks correspond to strong evidence of periodicity at that frequency ω (where $\omega = 2\pi/T$ for a period T). In the final step of periodicity detection or classification, the algorithms will do one or more of the following: Detect that periodicity occurs based on the strength of $J(\omega)$ (as in Sections III and IV), or use the periodicity estimate T to spatially localize components (as in Sections III and V).

A. MUSIC

The MUSIC algorithm is used to detect periodicity in a 1-D signal. In MUSIC, a received signal $\mathbf{y}$ is

$$\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{v} \tag{1}$$

where $\mathbf{A}$ is the signal subspace, $\mathbf{s}$ is the vector of signal amplitudes with respect to that subspace, and $\mathbf{v}$ is a noise vector. First, the covariance matrix $\mathbf{R_y}$ is computed

$$\mathbf{R_y} = E\{\mathbf{y}\mathbf{y}^H\} = \mathbf{Y}\mathbf{Y}^H = \mathbf{A}\mathbf{R_s}\mathbf{A}^H + \sigma^2\mathbf{I} \tag{2}$$

Fig. 6. Rail track inspection panorama.

where $\mathbf{Y}$ is a rectangular Toeplitz matrix such that $\mathbf{Y}\mathbf{Y}^H$ is a biased estimate of the autocorrelation matrix for signal $\mathbf{y}$ [20]. Matrix $\mathbf{Y}$ is defined as

$$\mathbf{Y} = \begin{bmatrix} y(h+1) & \cdots & y(1) \\ \vdots & \ddots & \vdots \\ y(M-h) & \cdots & y(h+1) \\ \vdots & \ddots & \vdots \\ y(M) & \cdots & y(M-h) \\ y^*(1) & \cdots & y^*(h+1) \\ \vdots & \ddots & \vdots \\ y^*(h+1) & \cdots & y^*(M-h) \\ \vdots & \ddots & \vdots \\ y^*(M-h) & \cdots & y^*(M) \end{bmatrix} \tag{3}$$

where M is the length of the signal $\mathbf{y}$, $y^*(M)$ is the complex conjugate of $y(M)$, and h is an index $1 \leq h \leq M$.

B. Gabor-Transformed Image

The panoramic image is transformed into several Gabor representations, producing different versions of the image, each transformed by one particular Gabor filter. This is done in a blockwise manner [21] and then analyzed in a rowwise manner. Each Gabor dimension d is used to create a signal for each row r so that $\mathbf{y}_d = \mathcal{G}_d(I_r(x))$, where $I_r(x)$ is the image at row r. Each signal $\mathbf{y}_d$ can produce an estimate of periodicity using the MUSIC algorithm.

To produce a cumulative estimate of periodicity for a particular row, rather than combining all estimates for each signal $\mathbf{y}_d$, all signals $\mathbf{y}_d$ are combined, and MUSIC is applied to the cumulative sum. To accomplish this, matrices from each of the dimensions $\mathbf{Y}_d$ are accumulated as follows:

$$\mathbf{Y} = \sum_d \mathbf{Y}_d. \tag{4}$$

The number of periodically repeating components in each row is estimated initially as L, and $\mathbf{R}_y$ is decomposed into D eigenvectors, which are denoted as $\mathbf{e}_j$, where $j \in \{1, \ldots, D\}$. The eigenvectors $\mathbf{e}_j$, where $j \in \{(L+1), \ldots, D\}$, span the noise subspace. This is used to find the maximum values of the following function:

$$J(\omega) = 20 \log_{10} \left(\frac{1}{\sum_{j=L+1}^{D} |\mathbf{a}^H(\omega)\mathbf{e}_j|^2} \right). \tag{5}$$

By using a signal space $\mathbf{a}^H(\omega)$ of Fourier transform basis functions, $J(\omega)$ can be computed as a sum of the fast Fourier transform (FFT) of all noise eigenvectors e_j, where $j \in \{(L+1), \ldots, D\}$. Intuitively, the eigenvectors in the noise subspace are orthogonal to the eigenvectors in the signal subspace. The eigenvectors in the signal subspace will contain a signal at the periodicity that is dominant; therefore, this same periodicity will be absent from the noise subspace. Therefore, the FFT of the noise eigenvectors for that frequency ω will have a close-to-zero value, producing a high value for $J(\omega)$. From this, up to L periodicities that are present along $I_r(x)$ can be detected (since the noise subspace will be orthogonal to L periodicities), and these correspond to L peaks in $J(\omega)$.

III. Detecting and Segmenting Periodically Occurring Components

In railroad inspection, periodically occurring components are often encountered. An example track panorama is shown in Fig. 6, where each wooden tie occurs at four approximately equidistant locations. MUSIC detects both dominant and less-dominant periodicities, which is useful since many large components are themselves made of smaller periodically repeating components.

For simplicity, the algorithm is applied only to images with repeating components in the horizontal direction, but it could also be applied along other orientations. Each image $I(x)$ is decomposed in a blockwise manner. Each blockwise row is referred to as r, and there are R such rows.

This algorithm, which is shown in Fig. 7, builds off of the basic structure in Fig. 5. An additional fourth step of component localization has been added once a strong periodicity is detected in each of the panorama rows. The entire algorithm is shown in Fig. 8, where the part that stems in Fig. 5 is located on the left and the part that is unique to the component localization of this algorithm is on the right.

In the initial step (see top of Fig. 8), the image row $I_r(x)$ is decomposed into M blocks $[b_0(x)b_1(x), \ldots, b_M(x)]$. Gabor features are extracted for each block $\mathcal{G}_d(b_m(x))$, where d is the dimension of the Gabor feature and $\mathcal{G}()$ is the transform. The Gabor features from [21] were used for their discriminative abilities. The application of the dth dimension of the Gabor transform to all blocks in the image row $I_r(x)$ is represented as $\mathcal{G}_d(I_r(x))$.

The output of the MUSIC algorithm provides an estimate for dominant periodicity of each image row T_r but no component localization. The phase offset of the repeating component υ_r is defined with respect to the leftmost coordinate, which is at $x = 0$ (the leftmost position of $I_r(x)$).

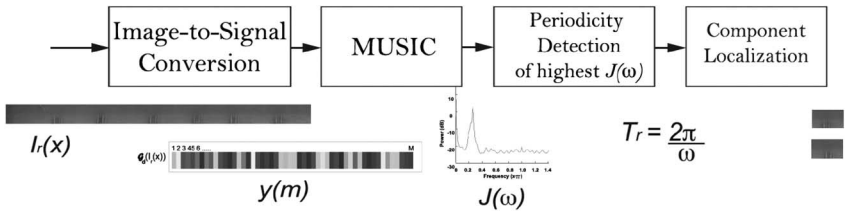


Fig. 7. Component localization algorithm as it relates to the main algorithm in Fig. 5.

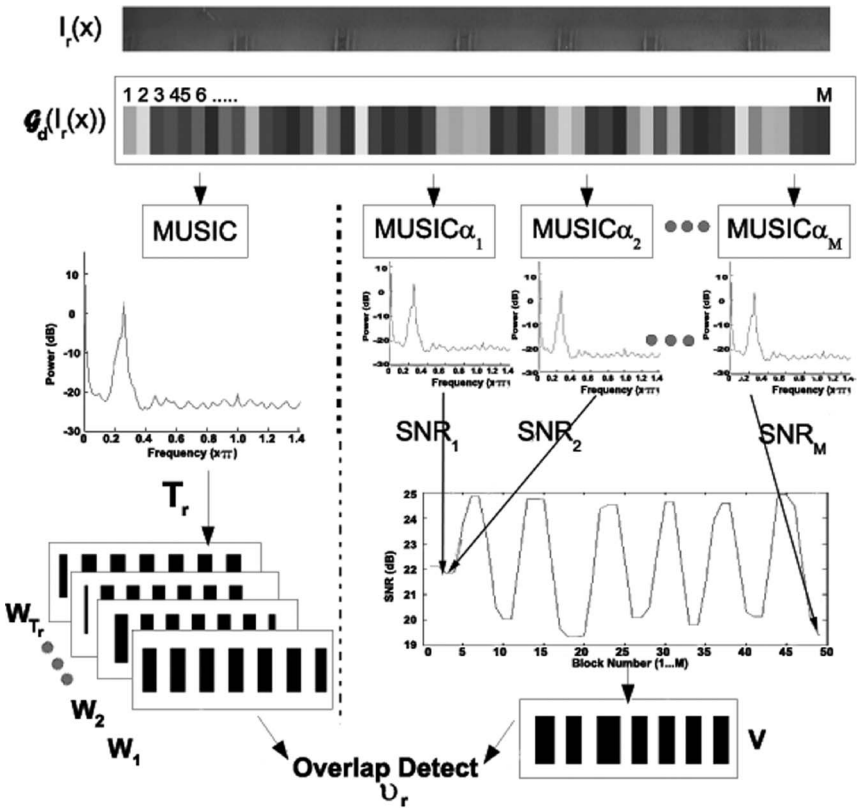


Fig. 8. Component localization algorithm.

The value υ_r is computed for each row, and this is accomplished by identifying which of the m blocks in a particular row $I_r(x)$ repeats with periodicity T_r. Each m block has a d-dimensional Gabor response defined as $\vec{\alpha}_m = \|\mathcal{G}(b_m(x))\|$. These Gabor responses are used to create a weighted version of (4), where each matrix $\mathbf{Y}_d$ is weighted with the Gabor response of that dimension d. This is shown in (6) for one particular candidate block m. Separate MUSIC equations are solved for each m, and the SNR of each of these M equations' resulting T_r is examined

$$\mathbf{Y_m} = \sum_d \alpha_{m,d} \mathbf{Y}_d. \tag{6}$$

This is demonstrated in Fig. 8, and the resulting SNR for the dominant peak is plotted in Fig. 8. This produces a sinusoidal SNR, with respect to m, with a period of T_r. A high SNR for block m indicates that there is a repeating object in that block. The SNR signal, which is a function of m, is thresholded to produce a binary mask $\mathbf{V}$ (on the right side of Fig. 8).

The computed T_r is then used to create candidate binary masks $\mathbf{W}_\upsilon$ (on the left side of Fig. 8). Each of the T_r binary candidate masks consist of repeating blocks with a duty cycle of $T_r/2$ and a phase offset of $\upsilon = \{1, \ldots, T_r\}$. The candidate mask that has the maximum overlap with $\mathbf{V}$ is determined, and this determines the phase offset υ_r. After υ_r is determined, all components that occur at every T_r pixels, beginning at location υ_r, are segmented. A duty cycle of 50% is assumed; therefore, each component is $T_r/2$ pixels wide.

A. Nondominant Periodicity Detection

The MUSIC algorithm detects multiple periodicities. In this application, the dominant frequency will produce two peaks in $J(\omega)$ since input $\mathbf{y}$ is real. The methodology in Section II can be performed on nondominant frequencies by setting $L > 2$ and detecting a third peak in $J(\omega)$. After isolating the component with primary periodicity (with $L = 2$), the algorithm is rerun to detect the less-dominant periodicities. Section III-B contains an example with multiple periodicities.

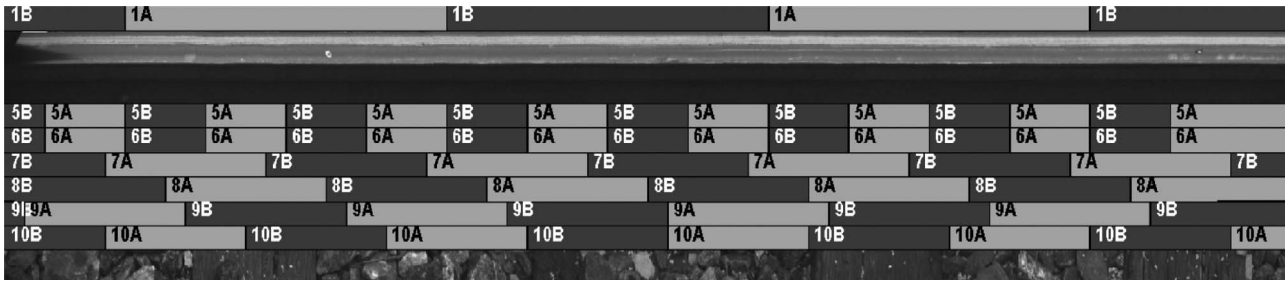


Fig. 9. Periodicity detection of the track panorama in Fig. 6. Each row is labeled using the detected period T_r and phase shift υ_r.

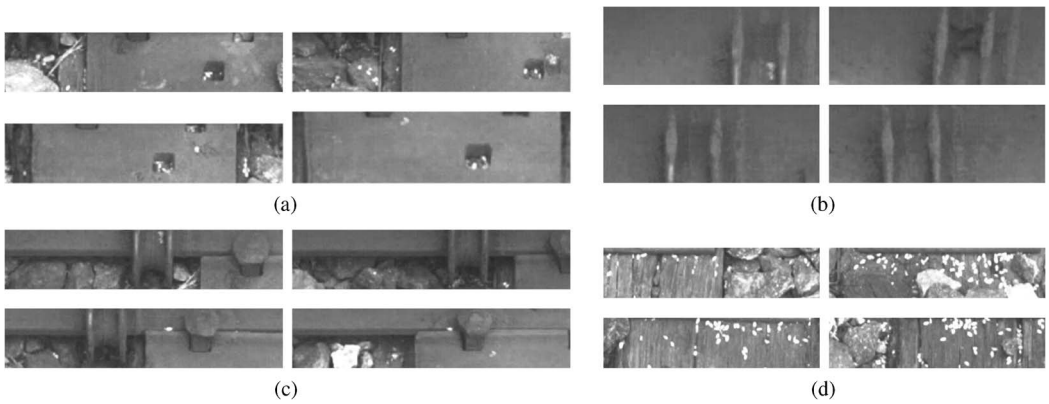


Fig. 10. Detected primary components. (a) Object 8A. (b) Object 8A. (c) Object 7A. (d) Object 10B.

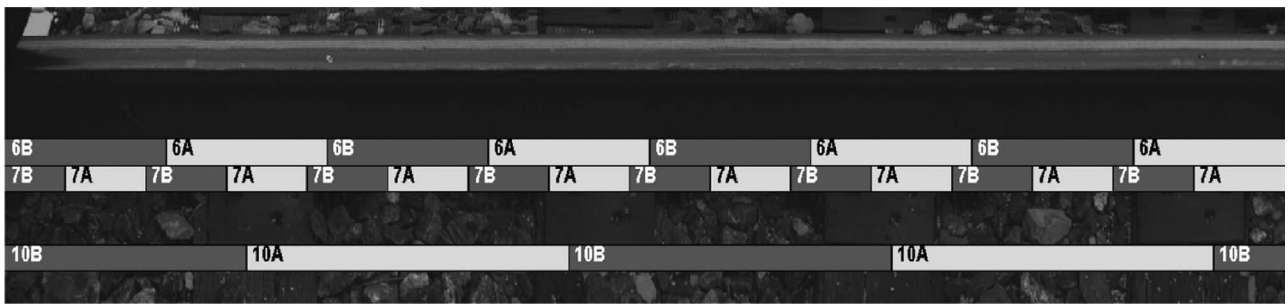


Fig. 11. $I_r(x)$ with strongest secondary periodicities ($r = 6$, 7, and 10).

B. Component Detection Example

Components are detected in panoramic images, which are formed from video data that were acquired from the hand-pushed cart in Section I. Panoramic images were created by stitching the video frames together using the computed displacement. The resulting panoramas were $N_y = 360$ pixels in height and between $N_x = 1000$ and $N_x = 2000$ pixels wide. The image was decomposed into overlapping blocks of size $N_b = 64$ in each dimension, and these blocks were then Gabor transformed. The blocks overlap with their neighboring blocks by $(N/2)$ pixels; therefore, for every row, $M = N_x/(N_b/2)$.

Periodic components are detected in the panorama in Fig. 6. In this panorama, certain rows contain both primary and secondary periodicities. The primary periodicities are noted as T_r. T_r and υ_r are computed according to the algorithm in Fig. 8. The detected components' T_r and υ_r are shown in Fig. 9.

Detected components are shown in Fig. 10. The tie plates are located and labeled Object 8A in Fig. 10(a). Note that, in the original image (see Fig. 6), the anchors are missing from the rightmost tie. The algorithm worked well despite this, and in components such as 6B and 7A [see Fig. 10(b) and (c)], it is evident that the missing anchor causes the components to not look identical. Object 10B is shown in Fig. 10(d). Ballast texture is successfully separated from tie texture, despite the presence of unknown granular material on the tie.

The panorama in Fig. 6 is then processed to detect secondary periodicities. The MUSIC algorithm is used, as described in Section III-A. The results are shown in Fig. 11, where Rows 6, 7, and 10 all contain strong secondary periodicities. Note that Row 7, which contained a lower frequency tie plate component in Fig. 10(c), now produces a high-frequency component corresponding to the anchor in Fig. 12(a). A lower frequency

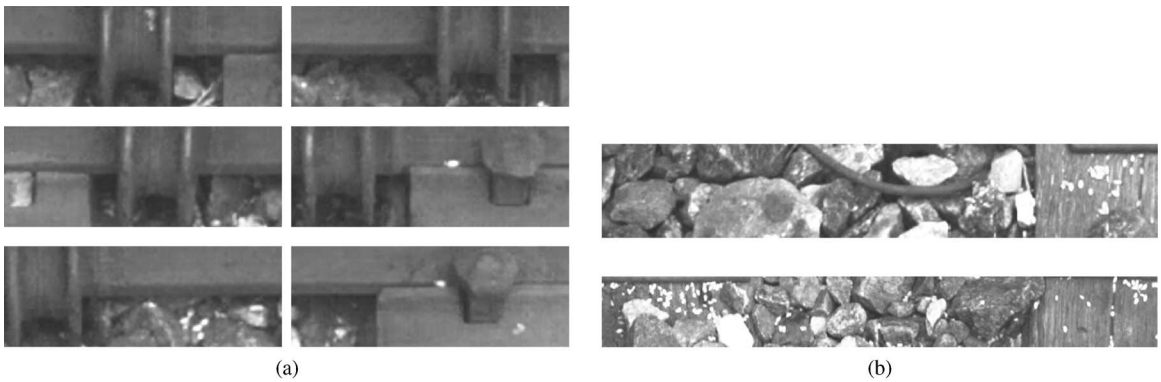


Fig. 12. Detected secondary components. (a) Object 7 B. (b) Object 10 B.

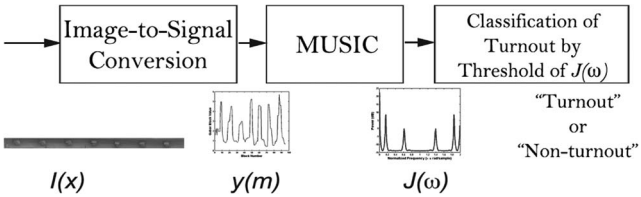


Fig. 13. Turnout detection as it relates to the main algorithm in Fig. 5.

is detected in Fig. 12(b), which contains both the tie and the neighboring ballast as one unit.

The component localization capabilities of the algorithm has been shown here. There is no ground truth for a correctly localized component since the algorithm may detect a part of an object that is not a complete object [e.g., the side of the tie plate in Fig. 12(a)]. For this reason, only T_r was used in the experiments conducted on 71 panoramas, which is detailed in Section VI-A.

IV. Turnout Inspection

Turnout detection is done according to Fig. 13, which is a modified version of Fig. 5. The input signal is the rail web or the middle part of the rail when viewed from a lateral viewpoint. $y(m)$ is the blockwise Gabor transform of the rail web (shown in Fig. 13 as a 1-D signal). The output is a threshold detector for $J(\omega)$.

To inspect a turnout, one must first verify that certain frames of an inspection video contain a turnout. A signal processing-based method is introduced for detecting periodic components indicative of turnouts, such as frog bolts or joint bar bolts [see Fig. 14(a)], and estimating that period T. The lateral viewpoint inspection video is converted along the rail web into a panoramic mosaic [see Fig. 14(a)]. The periodicity of the components in the panoramic mosaic is then estimated and the components subsequently localized. A blockwise Gabor transform is again utilized, where the web area that is analyzed in Fig. 14(b) is $N_b = 64$ pixels in height.

The image is transformed in a blockwise manner into the Gabor frequency domain [see Fig. 14(c)]. Each block's height is identical to the height of the rail web area shown in Fig. 14(a), and each block's response is computed using an overlapping window with respect to each block's right neighbor. This window overlaps by half of the block's width. This block-wise Gabor response is then processed as a 1-D signal [see Fig. 15(a)]. The MUSIC algorithm is subsequently applied to find periodic components [see Fig. 15(b)].

The MUSIC algorithm outputs a frequency analysis, in which the input signal's frequency response $J(\omega)$ is computed, as shown in Fig. 15(b). The output of Fig. 15(b) shows the power at each radial frequency ω. Each radial frequency relates to period T by the formula $\omega = 2\pi/T$. Hence, the peak at $\omega = 0.14\pi$ represents a component that repeats every $T = 14.3$ blocks.

Further experiments are run on a set of 43 panoramas, and details are given in Section VI-B.

A. Turnout Component Inspection

Once the turnout area has been isolated in the inspection video, the components from Section I-B1 are identified. The heel of the switch has a strong gradient that is perpendicular to the rail and is easily detected with spatial templates. The switch point is found using spatial templates on frames that are obtained from both the over-the-rail and lateral viewpoints. The frames are aligned by synchronizing the videos and maintaining an accurate tie count, which is possible using the algorithm in Section V.

V. Tie Detection

Tie detection is done according to Fig. 16, which is a modified version of Fig. 5. The input signal is the video (either lateral or over-the-rail). $y(m)$ is a 1-D signal formed after filtering each Gabor-transformed video frame with a binary tie detection mask. The ties are detected as the peaks of the 1-D signal, and for this algorithm, the detected T is just used to constrain the search window for peak amplitudes.

The 1-D signal $y(m)$ is obtained by performing texture classification on individual video frames, followed by spatial filtering with a user-created binary filter. For example, Fig. 17(a) contains a tie. A binary mask is created by dividing the image into blocks and classifying each block as ballast and nonballast, as shown in Fig. 17(b). To do this, each image block is Gabor

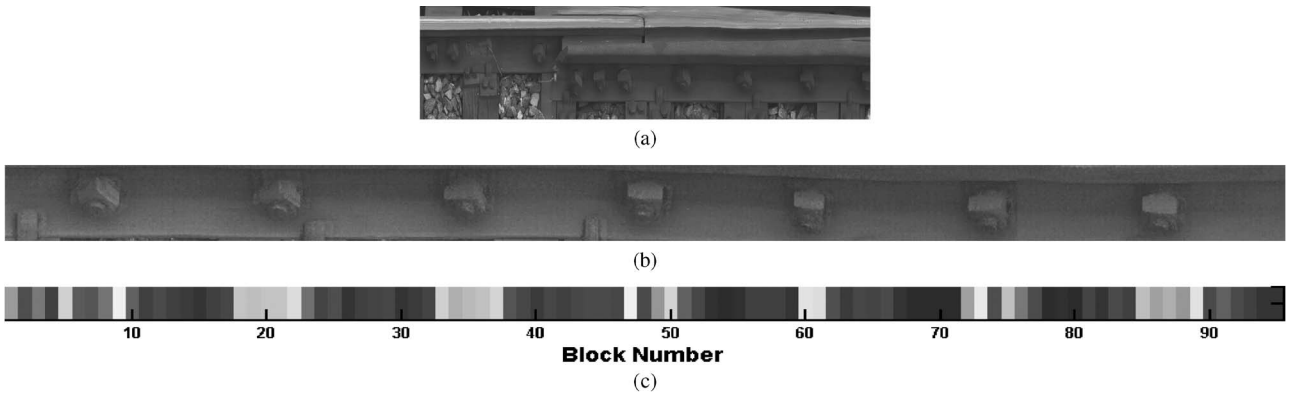


Fig. 14. Turnout bolts used in recognition. (a) Original panoramic image. (b) Rail web area. (c) Blockwise Gabor image.

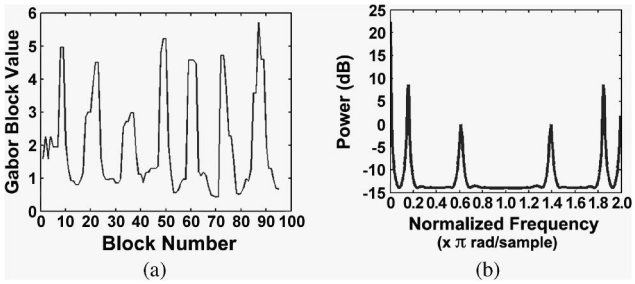


Fig. 15. Turnout detection. (a) $\mathcal{G}(I_r(x))$. (b) $J(\omega)$ of $\mathcal{G}(I_r(x))$.

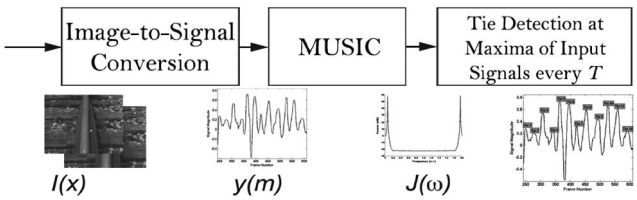


Fig. 16. Tie detection as it relates to the main algorithm in Fig. 5.

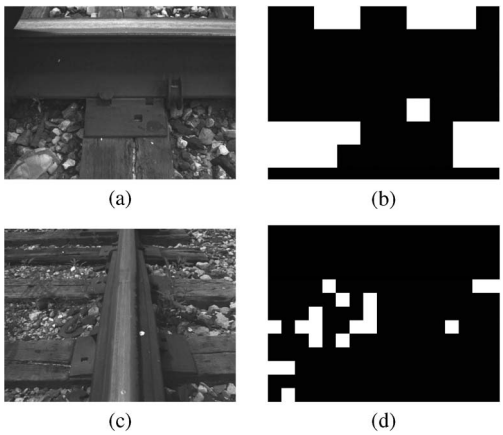


Fig. 17. (a) Original lateral image. (b) Binary texture classification mask for lateral image, where black is "nonballast" and white is "ballast." (c) Original over-the-rail image. (d) Binary texture classification mask for over-the-rail image.

transformed using the same Gabor filters, as in Section II-B, and the resulting d-dimensional signal for each block is used to classify the block using a d-dimensional support vector machine (SVM) [23]. Two manually labeled images from a different railroad track were used to train the SVM. The over-the-rail viewpoint in Fig. 17(c) is similarly decomposed into blocks and classified, and the resulting mask is shown in Fig. 17(d).

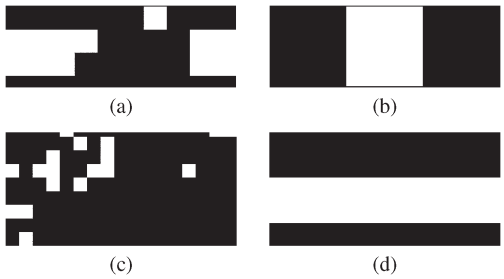


Fig. 18. (a) Lower half of the binary texture classification mask for the lateral viewpoint. (b) Template that the lateral mask is compared against. (c) Lower half of the binary texture classification mask for the over-the-rail viewpoint. (d) Template that the over-the-rail mask is compared against.

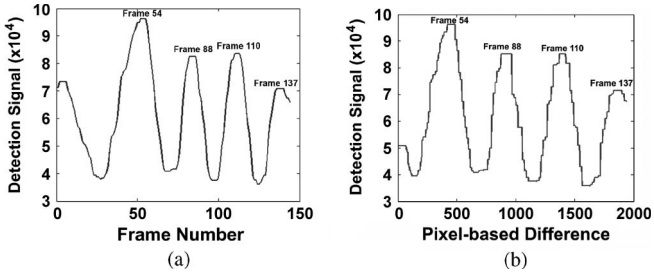


Fig. 19. Response of binary classification mask to template as a function of (a) frame number and (b) pixel-based distance.

Only the lower half of the lateral and over-the-rail masks are examined, as indicated in Fig. 18(a) and (c). These are subtracted with the tie detection masks for lateral and over-the-rail [see Fig. 18(b) and (d), respectively]. From this, the 1-D signal is formed.

This results in a signal that is sinusoidal with respect to time when the inspection video is acquired at a constant speed, as shown in Fig. 19(a). To avoid the requirement of constant speed, the signal can also be recorded as a function of interframe displacement [see Fig. 19(b)].

Video frames that contain a tie are located at the maximum amplitude of the signal in Fig. 19(a). If velocity is nonconstant, the video frames that contain a tie are computed by determining which frame corresponds to the pixel-based distance at the maximum amplitude in Fig. 19(b).

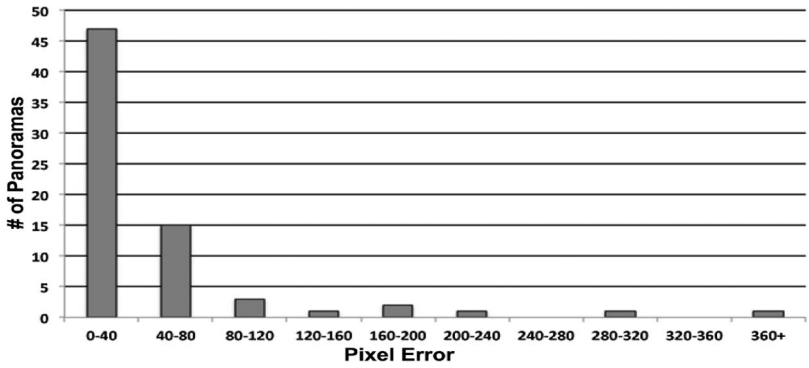


Fig. 20. Pixel error in the periodicity estimation of ties in panoramas.

TABLE I
TURNOUTS

Number of Panoramas	Precision	Recall
43	0.75	0.81

TABLE II
TIE DETECTION

Type	Number of ties	Precision	Recall
Lateral	351	0.97	0.99
Over-the-rail	273	0.97	1.0

The experimental results are shown in Section VI-C.

VI. EXPERIMENTS

A. Component Detection on Panoramas

Further experiments were conducted on 71 panoramas of the track. The goal of this experiment was to detect the dominant periodicity, in pixels, in the lower half of the panorama. The estimate for T was obtained, and then, error was computed from the ground-truth estimate of the tie periodicity within that panorama. The periodicity prediction error (in pixels) for all panoramas is shown in Fig. 20. Note that 87% of the panoramas had their tie periodicity correctly predicted within 80 pixels.

The prediction errors of 80 pixels or more are attributed to one of the following: occlusion, oversaturation, and anomalous lighting. Mild occlusion is acceptable, as was shown in the previous experiments on the panorama in Fig. 6. Occlusion that completely covers two or more of the ties, however, is detrimental. Oversaturation that causes the ballast to appear white and without edges is also detrimental as the ballast will lose its characteristic textures, and the Gabor filtering will be ineffective. Anomalous lighting is also a problem since a bright concentrated light from either a natural or unnatural source can alter the texture properties.

B. Turnout Detection

Our goal is to identify turnout images from a collection of panoramic images of the lateral viewpoint. We selected 43 panoramas, 26 of which contained turnout components and 17 of which did not contain turnout components (see Table I). A true-positive (tp) is a turnout that is indeed a turnout, a true-negative (tn) is a nonturnout that is indeed a nonturnout, a false-positive (fp) is a detected turnout in a nonturnout panorama, and a false-negative (fn) is a turnout that was not detected. Precision is $tp/tp + fp$, and recall is $tp/tp + fn$ [22].

Precision is lower than recall, thus indicating a slight bias toward overidentifying panoramas as turnout panoramas. The periodic detection used for turnout detection occurs along the rail, and the assumption is that nonturnout panoramas have a smooth rail area, whereas turnout panoramas have some component repeating. Since there are a variety of bolt configurations on the turnout area, the detection of a turnout is not determined by the exact spacing of the repeating object but, rather, by the PSNR of a detected object using the MUSIC algorithm. Hence, several approximately periodic rail irregularities, particularly in the form of dirt, shadows, and reflections, can cause the algorithm to falsely detect a turnout.

C. Tie Detection

Table II evaluates the method presented here for the lateral and over-the-rail viewpoints. We defined a true-positive (tp) as detected ties that are indeed ties, true-negative (tn) as ballast between ties that was correctly identified as being not a tie (for our purposes, we count only the area to the right of the tie), false-positive (fp) as a detected tie where there was none, and a false-negative (fn) as a tie that was not detected.

Precision is slightly lower than recall for both lateral and over-the-rail viewpoints; hence, when an error occurs, it is more likely due to false prediction of a tie rather than not identifying a tie. Because the algorithm incorporates global information by treating the detection as a signal [see Fig. 21(a) and (b)], it is robust to sources of noise on individual ties, including noise from occlusion, worn parts, and lighting inconsistencies.

VII. CONCLUSION

Computer vision can be used to create railroad track inspection algorithms that are objective and provide a reliable assessment of track conditions. Video containing track inspection data will inherently contain periodically recurring components. MUSIC is a valuable signal processing technique that extracts periodic signals from a 1-D signal, and its robustness to noise allows it to effectively estimate periodicity in real-world inspection video and images.

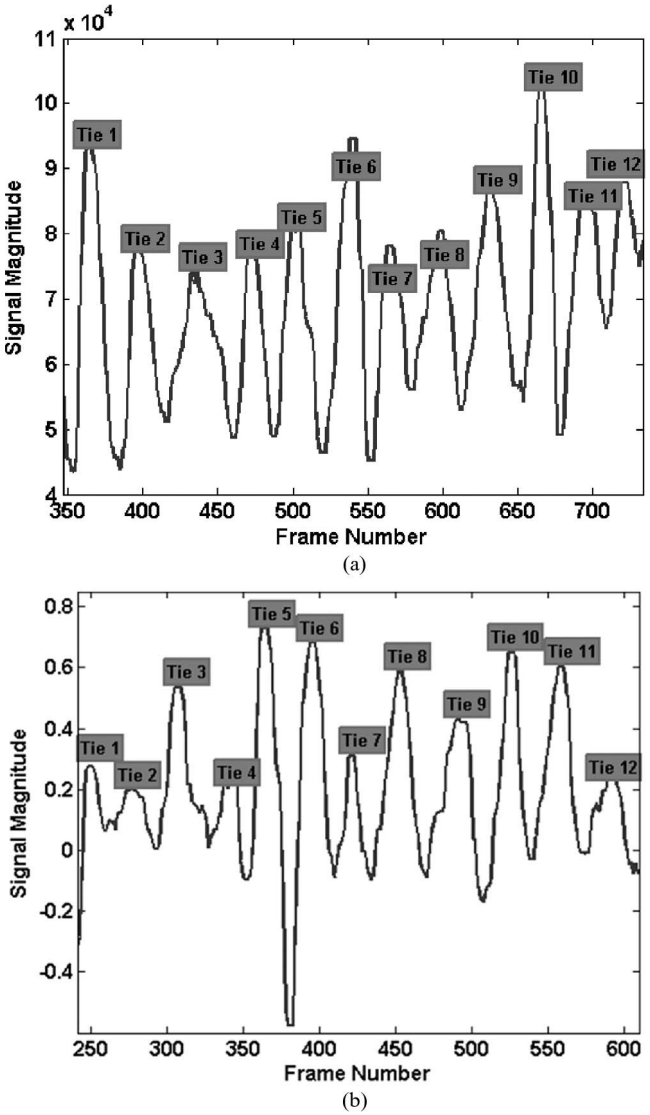


Fig. 21. Tie detection within turnout area for (a) lateral viewpoint and (b) over-the-rail viewpoint.

Future railroad track inspection technology should incorporate automatic detection and segmentation of periodically occurring objects to achieve a more robust system. This could lead to an autonomous system that could inspect thousands of miles of track without human supervision by automatically adjusting spatial detection filters as the track cart travels on previously uninspected track.

REFERENCES

[1] L. Molina, E. Resendiz, J. Edwards, J. Hart, C. Barkan, and N. Ahuja, "Condition monitoring of railway turnouts and other track components using machine vision," in *Proc. Transp. Res. Board 90th Annu. Meeting*, Washington, DC, 2011, pp. 1–17.

[2] E. Resendiz, L. Molina, J. Hart, J. Edwards, S. Sawadisavi, N. Ahuja, and C. Barkan, "Development of a machine vision system for inspection of railway track components," in *Proc. 12th WCTR*, Lisbon, Portugal, 2010, Paper 3355.

[3] S. Sawadisavi, J. Edwards, E. Resendiz, J. Hart, C. Barkan, and N. Ahuja, "Development of a machine vision system for inspection of railroad track," in *Proc. Amer. Railway Eng. Maintenance Way Assoc. Annu. Conf.*, 2009.

[4] S. Sawadisavi, J. Edwards, E. Resendiz, J. Hart, C. Barkan, and N. Ahuja, "Machine-vision inspection of railroad track," in *Proc. Transp. Res. Board 88th Annu. Meeting*, Washington, DC, 2009, Paper 09-1369.

[5] T. Hoang, N. Haas, Y. Li, C. Otto, and S. Pankanti, "Enhanced rail component detection and consolidation for rail track inspection," in *Proc. IEEE Workshop Appl. Comput. Vis.*, 2012, pp. 289–295.

[6] F. Kaleli and Y. Akgul, "Vision-based railroad track extraction using dynamic programming," in *Proc. 12th Int. IEEE Conf. Intell. Transp. Syst.*, 2009, pp. 1–6.

[7] M. Singh, S. Singh, J. Jaiswal, and J. Hempshall, "Autonomous rail track inspection using vision based system," in *Proc. IEEE Int. Conf. Comput. Intell. Homeland Secur. Pers. Safety*, 2006, pp. 56–59.

[8] J. Velten, A. Kummert, and D. Maiwald, "Image processing algorithms for video-based real-time railroad track inspection," in *Proc. 42nd Midwest Symp. Circuits Syst.*, 2000, pp. 530–533.

[9] E. Stella, N. Ancona, and A. Distante, "Ballast 3D reconstruction by a matching pursuit based stereo matcher," in *Proc. IEEE Intell. Veh. Symp.*, 2004, pp. 653–657.

[10] C. Alippi, E. Casagrande, F. Scotti, and V. Piuri, "Composite real-time image processing for railways track profile measurement," *IEEE Trans. Instrum. Meas.*, vol. 49, no. 3, pp. 559–564, Jun. 2000.

[11] J. Edwards, J. Hart, S. Todorovic, C. Barkan, N. Ahuja, Z. Chua, N. Kocher, and J. Zeman, "Development of machine vision technology for railcar safety appliance inspection," in *Proc. Int. Heavy Haul Conf., Spec. Tech. Session*, 2007, pp. 745–752.

[12] Y.-C. Lai, C. Barkan, J. Drapa, N. Ahuja, J. Hart, P. Narayanan, C. Jawahar, A. Kumar, L. Milhon, and M. Stehly, "Machine-vision analysis of the energy efficiency of intermodal freight trains," *J. Rail Rapid Transit*, vol. 221, pp. 353–364, 2007.

[13] A. Kumar, N. Ahuja, J. Hart, U. Visesh, P. Narayanan, and C. Jawahar, "A vision system for monitoring intermodal freight trains," in *Proc. IEEE Workshop Appl. Comput. Vis.*, 2007, p. 24.

[14] Q.-J. Kong, A. Kumar, N. Ahuja, and Y. Liu, "Robust segmentation of freight containers in train monitoring videos," in *Proc. IEEE Workshop Appl. Comput. Vis.*, 2009, pp. 1–6.

[15] D. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*. Upper Saddle River, NJ: Prentice-Hall, 2002.

[16] I. Tang and T. Breckon, "Automatic road environment classification," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 476–484, Jun. 2011.

[17] V. Tyagi, S. Kalyanaraman, and R. Krishnapuram, "Vehicular traffic density state estimation based on cumulative road acoustics," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 3, pp. 1156–1166, Sep. 2012.

[18] Y. Keller and Y. Shkolnisky, "A signal processing approach to symmetry detection," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2198–2207, Aug. 2006.

[19] O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.

[20] S. Marple, *Digital Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[21] B. Manjunathi and W. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 837–842, Aug. 1996.

[22] D. Olson and D. Delen, *Advanced Data Mining Techniques*. New York: Springer-Verlag, 2008.

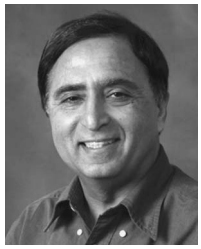
[23] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

Esther Resendiz (S'01–M'11) received the B.S. degree in electrical engineering (with highest honors) from The University of Texas at Austin, in 2001 and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana–Champaign (UIUC), Champaign, in 2006 and 2010, respectively.

She has collaborated with the Rail Transportation and Engineering Center, UIUC, for technology development on several projects, including undercarriage inspection and railroad track inspection for the Transportation Research Board, the Association of American Railroads, and the NEXTRANS Center. Her research interests include applied computer vision for railroad applications, image quality assessment, signal processing, and content-based image retrieval.

John M. Hart received the B.S. degree in electrical engineering technology from DeVry Institute of Technology, Calgary, AB, Canada, in 1984 and the M.S. degree in electrical and computer engineering from the University of Illinois at Urbana–Champaign (UIUC), Champaign, in 1991.

Since 1995, he has over 15 years experience developing advanced camera technologies as a Research Engineer. Since 2001, he has been the Project Leader of six interdisciplinary projects involving the design and development of wayside machine-vision inspection systems for the Association of American Railroads, Transportation Research Board, and Burlington Northern Railway. He is currently a Senior Research Engineer with the Coordinated Science Laboratory and conducts research with the Computer Vision and Robotics Laboratory, Beckman Institute for Advanced Science and Technology, UIUC. His research interests include automated visual inspection systems, advanced visual imaging technologies, and robotic and electromechanical methods for mobile image acquisition.



Narendra Ahuja (S'79–M'79–SM'85–F'92) received the Ph.D. degree in computer science from the University of Maryland, College Park, in 1979.

He was the Founding Director of the International Institute of Information Technology, Hyderabad, India, where he continues to serve as Director International. He is currently a Research Professor with the Department of Electrical and Computer Engineering, with the Beckman Institute for Advanced Science and Technology, and with the Coordinated Science Laboratory, University of Illinois at Urbana–Champaign, Champaign. He currently leads the Information Technology Research Academy, which is an initiative started by the Department of Electronics and Information Technology, Government of India. His current research interests include extraction and representation of spatial structure in images and video, integrated use of multiple image-based sources for scene representation and recognition, versatile sensors for computer vision, and applications including visual communication, image manipulation, and information retrieval.

Dr. Ahuja is a Fellow of American Association for Artificial Intelligence, the International Association for Pattern Recognition, the Association for Computing Machinery, the American Association for the Advancement of Science, and the International Institute of Information Technology. He received the 1998 Technology Achievement Award of the International Society for Optical Engineering, the 1999 Emanuel R. Piore Award of the IEEE, the 2008 TA Stewart-Dyer/Frederick Harvey Trevithick Prize of the Institution of Mechanical Engineers, the Distinguished Alumnus Award from the Department of Computer Science of the University of Maryland in 2008, the 2008–2011 Open Innovation Research Award from Hewlett-Packard, the 2010 Google Research Award, and the Birla Institute for Technology and Science, Pilani, India, in 2012. He was a corecipient of the Best Paper Awards from the IEEE TRANSACTIONS ON MULTIMEDIA in 2006 and the International Conference on Pattern Recognition in 2012.