On-Line Seam Detection in Rolling Processes Using Snake Projection and Discrete Wavelet Transform

Jing Li

Jianjun Shi

Department of Industrial and Operations Engineering,
The University of Michigan,
1205 Beal Avenue,
Ann Arbor, MI 48109-2117

Tzyy-Shuh Chang
OG Technologies, Inc.,
4300 Varsity Drive, Suite C,
Ann Arbor, MI 48108

This paper describes the development of an on-line quality inspection algorithm for detecting the surface defect "seam" generated in rolling processes. A feature-preserving "snake-projection" method is proposed for dimension reduction by converting the suspicious seam-containing images to one-dimensional sequences. Discrete wavelet transform is then performed on the sequences for feature extraction. Finally, a T^2 control chart is established based on the extracted features to distinguish real seams from false positives. The snake-projection method has two parameters that impact the effectiveness of the algorithm. Thus, selection of the parameters is discussed. Implementation of the proposed algorithm shows that it satisfies the speed and accuracy requirements for on-line seam detection. [DOI: 10.1115/1.2752519]

Keywords: seam detection, discrete wavelet transform (DWT), feature extraction, T^2 control chart

1 Introduction

Rolling is a high-speed bulk deformation process that reduces the thickness or changes the cross section of a long workpiece by compressive forces applied through a set of rolls [1]. Surface defects are critical quality concerns in the rolling industry. They are usually generated due to material overfills, nonmetallic inclusion, or porosities. Among surface defects, a seam is one of the most serious types. Because seams result in stress concentration on the bulk material that could cause catastrophic failures when the rolled product is in use, products with severe seams have to be scraped. Therefore, detection of seams is important for quality assurance.

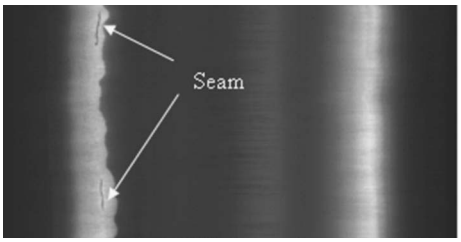
Seam detection has traditionally been restricted to off-line manual inspection because of a lack of effective sensors that can inspect the product surface on-line, under harsh environmental conditions, such as heat, dust, and lubricants. In recent years, with the development of advanced imaging technologies, vision sensors have been successfully adopted in the rolling process, collecting high-quality sensing images of the product surface. As a result, automatic on-line detection of seams becomes possible. A portion of a sensing image from a bar-rolling process (i.e., the products are steel bars) is given in Fig. 1.

Because a seam has a high contrast against its background in terms of image gray levels, as shown in Fig. 1, edge detection techniques can be adopted to detect seams. A qualified edge detector should possess two properties for on-line seam detection. First, considering the rolling velocity, 100% inspection of the product surface requires a data processing speed of at least 80 Mb/s. Thus, the edge detector should be computationally fast. Second, the sensing images are noisy, as seams usually intermix with surface marks and material impurities. Therefore, for effectively identifying seams and reducing the false alarm rate, the edge detector should be noise insensitive. In the edge detection literature, although some edge detectors, such as Prewitt detectors and Sobel detectors [2], are simple and fast, their performance deteriorates unacceptably when the image is noisy. More sophisticated edge detectors use smoothing operations to reduce noise [2], but some useful information for detecting edges is inevitably lost. Optimal detectors [3] were proposed to ensure an acceptable compromise between noise reduction and edge conservation. A typical optimal detector is called "Canny," which has been extensively used as a standard gauge in the edge detection research [4]. However, optimality is achieved with a sacrifice in the detection speed. Other edge detection approaches include multiscale methods, statistical tools, and neural networks [5], which are computationally more intensive.

To detect seams fast and effectively under noisy rolling surface conditions, a two-stage seam detection methodology has been developed in this research. In the first stage, a Sobel edge detector is adopted to rapidly identify the suspicious seam-containing local regions (referred to as sub-images from here on) in a sensing image. As the Sobel edge detector is noise sensitive, a large number of the identified sub-images are indeed false positives. Thus, a re-inspection on the sub-images is needed in the second stage to distinguish real seams from false positives. Because the sub-images have a substantially smaller size than the original sensing images, a more sophisticated algorithm can be adopted. The developed re-inspection algorithm consists of three steps: (i) converting the sub-images into 1-D sequences by a proposed snake-projection method that reduces the data dimension while preserving seam characteristics; (ii) extracting features from the sequences by discrete wavelet transform (DWT); and (iii) identifying seams by a T^2 control chart based on the extracted features.

Because the Sobel edge detector in the first-stage seam detection methodology is well documented in the literature, this paper focuses on the development of the second-stage algorithm (referred to as snake-projection-wavelet algorithm). The remainder of this paper is organized as follows. Section 2 gives a brief description of the sub-images and data structure. Section 3 provides the detailed procedure for developing the snake-projection-wavelet algorithm. Section 4 presents a case study to demonstrate the effectiveness of the developed algorithm. Finally, Sec. 5 summarizes this research.

Contributed by the Manufacturing Science Division of ASME for publication in the JOURNAL OF MANUFACTURING SCIENCE AND ENGINEERING. Manuscript received December 1, 2004; final manuscript received May 3, 2007. Review conducted by C. James Li.



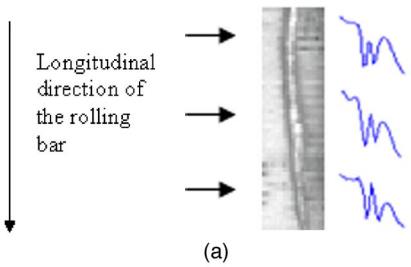


Fig. 1 A portion of a sensing image from a bar-rolling process

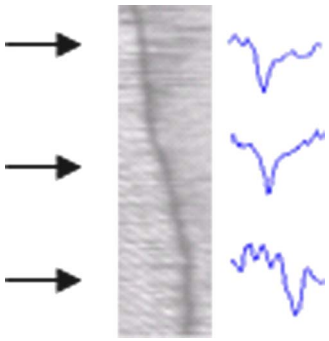


Fig. 2 A seam sub-image

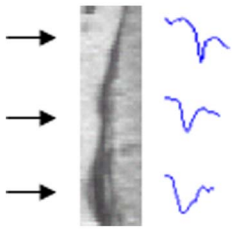


Fig. 3 (*a*) Two ridge-based false positive sub-images (left: containing two dark strips; right: containing one dark strip), (*b*) a mark-based false positive sub-image

2 Description of Sub-Images and Data Structure

This section introduces the physical forms of seams and false positives, their appearances in sub-images, and the structure of the sub-image data.

2.1 Seam and False Positive: Physics and Image Appearance. A seam is a thin deep crack along the longitudinal direction of a rolling bar. Thus, it shows as a nearly vertical thin dark strip with a constant width of 2–3 pixels in the sensing image. Figure 2 is a seam-containing sub-image.

There are two types of false positives: ridge based and mark based, as shown in Fig. 3. A ridge-based false positive is a longitudinal ridge on the surface caused by material overfills. It is usually pictured as a thin bright strip with one or two dark strips on its sides, depending on the angle of the lighting source, where the bright and the dark strips are the ridge and its shadow(s), respectively. A mark-based false positive is a longitudinal mark on the surface, usually characterized by a dark strip of varying width. Note that, although the false positives have different physical forms and image patterns, they each have a dark strip in the sub-images, which is misidentified as a seam. However, in the ridge-based false positive, the dark strip is paralleled by a bright strip (i.e., the ridge); and in the mark-based false positive, the width of the dark strip (i.e., the mark) varies. These two characteristics form the fundamental difference between false positives and seams.

2.2 Data Structure of Sub-Images. The data of a sub-image is a matrix with each element corresponding to one image pixel. The elements take integer values ranging from 0 to 255 gray levels, where 0 and 255 represent black and white, respectively. A typical matrix for one sub-image containing $m \times n$ pixels is given in (1), where a_{ij} is the gray level of the pixel at the ith row and the jth column ($0 \leq a_{ij} \leq 255$):

$$\begin{bmatrix}
a_{11} & \cdots & \cdots & a_{1(t_1-k)} & \cdots & a_{1(t_1-1)} & \underline{\underline{a_{1t_1}}} & a_{1(t_1+1)} & \cdots & a_{1(t_1+k)} & \cdots & \cdots & a_{1n} \\
a_{21} & \cdots & \cdots & a_{2(t_2-k)} & \cdots & a_{2(t_2-1)} & \underline{\underline{a_{2t_2}}} & a_{2(t_2+1)} & \cdots & a_{2(t_2+k)} & \cdots & \cdots & a_{2n} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
a_{i1} & \cdots & a_{i(t_i-k)} & \cdots & a_{i(t_i-1)} & \underline{\underline{a_{it_i}}} & a_{i(t_i+1)} & \cdots & a_{i(t_i+k)} & \cdots & \cdots & \cdots & a_{in} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
a_{(m-1)1} & \cdots & \cdots & a_{(m-1)(t_{m-1}-k)} & \cdots & a_{(m-1)(t_{m-1}-1)} & \underline{\underline{a_{(m-1)t_{m-1}}}} & a_{(m-1)(t_{m-1}+1)} & \cdots & a_{(m-1)(t_{m-1}+k)} & \cdots & \cdots & a_{(m-1)n} \\
a_{m1} & \cdots & \cdots & \cdots & a_{m(t_m-k)} & \cdots & a_{m(t_m-1)} & \underline{\underline{a_{mt_m}}} & a_{m(t_m+1)} & \cdots & a_{m(t_m+k)} & \cdots & a_{mn}
\end{bmatrix} \quad (1)$$

In addition to the matrix, the location of the center line of the dark strip in the sub-image is also taken as input to the snake-projection-wavelet algorithm. This location is identified by the Sobel edge detector and recorded in a vector $[t_1,\ldots,t_i,\ldots,t_m]^T$, where t_i is the column index of element a_{it_i} $(i=1,2,\ldots,m)$ on the center line of the dark strip. Because a dark strip is not strictly vertical, t_i is not necessarily equal to t_j $(1\le i,j\le m, i\ne j)$. Thus, elements $[a_{1t_1},\ldots,a_{it_i},\ldots,a_{mt_m}]^T$ form a curved line along the vertical direction. This line, shown in (1) by the elements with double underlines, is referred to as the "characteristic line" of a sub-image in this paper. Physically, the characteristic lines of a seam sub-image, a ridge-based false positive sub-image, and a mark-based false positive sub-image correspond to the center lines of the seam, one shadow strip of the ridge, and the mark, respectively. Detailed procedures for obtaining a sub-image matrix as well as the characteristic line can be found in [6].

3 Snake-Projection-Wavelet Algorithm

The procedure of the snake-projection-wavelet algorithm is shown in Fig. 4. The sub-images are first converted to 1-D sequences by a feature-preserving snake-projection method. A discrete wavelet transform is then performed on the sequences for feature extraction. Finally, a T^2 control chart is constructed, based on the features to distinguish seams from false positives.

3.1 Dimension Reduction by a Feature-Preserving Snake-Projection Method. Dimension reduction samples the pixels in the sub-image and converts the sub-image into a 1-D sequence. With fewer pixels and reduced dimension, the computational speed of the algorithm can be increased. In dimension reduction, it is important that the features that distinguish seams from false positives are preserved. These features are embedded in the row pixels of the sub-image. To further illustrate this point, three rows are extracted from each sub-image in Figs. 2 and 3, with the gray levels of the pixels on each row plotted on the right-hand side of the sub-image. (The arrows on the left-hand side indicate the locations of the three rows.) It can be seen that (i) a ridge-based false positive differs from a seam in the shape of the curves, and (ii) the width of the valleys in the curves of a mark-based false positive varies while this width in the curves of a seam is almost constant. Therefore, to capture the shape difference and the width variation, multiple rows need to be extracted from a sub-image, which motivates the following "snake-projection" method.

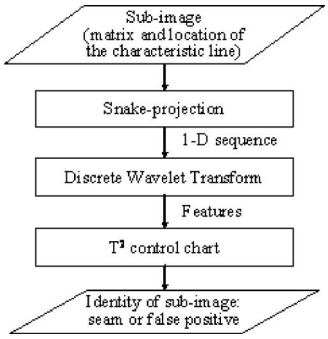


Fig. 4 Procedure of the snake-projection-wavelet algorithm

Step 1. Extract the elements $\{a_{i(t_i-k)},a_{i(t_i-k+1)},\ldots,a_{i(t_i-1)},a_{it_i},a_{i(t_i+1)},a_{i(t_i+2)},\ldots,a_{i(t_i+k)}: i=1,2,\ldots,m\}$ from (1), i.e., extract the pixels on the characteristic line, together with k adjacent pixels on each side of the line for each row of a sub-image. The extracted elements constitute a $m\times(2k+1)$ matrix, as shown in (2). In this matrix, the characteristic line corresponds to the center column.

Step 2. In (2), extract the elements on every s row and the elements on the endmost columns connecting those rows. The inclusion of the endmost columns helps maintain the continuity of gray levels between two extracted rows. The extracted elements, highlighted by gray strips in (2), constitute a sequence in which the ordering of the elements is specified by the arrows. This sequence is denoted by a vector $\mathbf{V}^{k,s}$.

$$\begin{bmatrix}
a_{1(t_1-k)} & \cdots & a_{1(t_1-1)} & a_{1t_1} & a_{1(t_1+1)} & \cdots & a_{1(t_1+k)} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
a_{(s+1)(t_{s+1}-k)} & \cdots & a_{(s+1)(t_{s+1}-1)} & a_{(s+1)t_{s+1}} & a_{(s+1)(t_{s+1}+1)} & \cdots & a_{(s+1)(t_{s+1}+k)} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
a_{(2s+1)(t_{2s+1}-k)} & \cdots & a_{(2s+1)(t_{2s+1}-1)} & a_{(2s+1)t_{2s+1}} & a_{(2s+1)(t_{2s+1}+1)} & \cdots & a_{(2s+1)(t_{2s+1}+k)} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
a_{(3s+1)(t_{3s+1}-k)} & \cdots & a_{(3s+1)(t_{3s+1}-1)} & a_{(3s+1)t_{3s+1}} & a_{(3s+1)(t_{3s+1}+1)} & \cdots & a_{(3s+1)(t_{3s+1}+k)} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
a_{m(t_m-k)} & \cdots & a_{m(t_m-1)} & a_{mt_m} & a_{m(t_m+1)} & \cdots & a_{m(t_m+k)}
\end{bmatrix} \tag{2}$$

Due to the shape of the connective gray strips in (2), the proposed projection method is called "snake projection." By applying the snake projection to a sub-image, a 1-D sequence can be obtained. Figures 5 and 6 plot sequences $\mathbf{V}^{8,5}$ (i.e., $k=8, s=5$) for the seam sub-image in Fig. 2 and the false positive sub-images in Fig. 3, respectively.

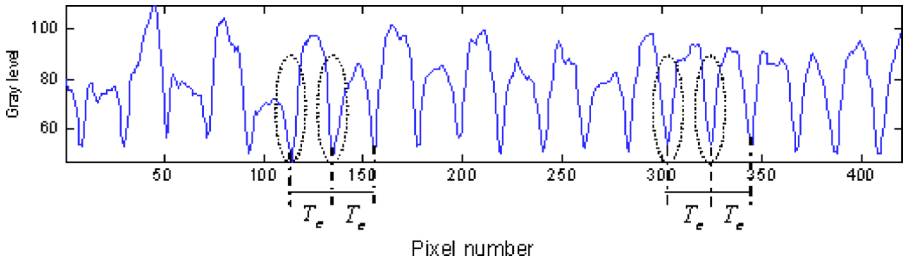


Fig. 5 Sequence of the seam sub-image in Fig. 2 by snake projection (k=8,s=5)

Three important issues of the snake-projection method need to be addressed.

(i) *Step 1* re-aligns a sub-image by its characteristic line, such that in the resulting new image matrix, (2), the characteristic line is strictly vertical and becomes the center column. As a result, *Step 2* generates a cycle-based sequence. The starting and ending points of a cycle are the pixels on the characteristic line. Furthermore, it can easily be derived that the length of the cycles, denoted by T_c, is $T_c=2k+s$. For illustration purposes, several cycles are labeled for each sequence in Figs. 5 and 6. The frequency of the cycles f_c is:

$$f_c = 1/(2k+s) \tag{3}$$

(ii) In *Step 2*, the snake projection starts from the first row and ends with the last possible row in (2) for effectively detecting the mark-based false positives. Since there is no fixed pattern in the width variation of the mark-based false positives (i.e., the mark could be thin in the beginning and wide in the end, or the reverse, or wide in the middle and thin in both ends, etc.), the most effective strategy to capture the different patterns of width variations and distinguish any mark-based false positive from seams is to make the snake to wind over the whole length of the mark (i.e., from the first to the last possible row in (2)).

(iii) k and s are two parameters associated with the snake projection and jointly determine the data reduction rate. A higher reduction rate results in less data to be processed, which helps to increase the data processing speed. However, it may also lead to more information loss such that seams cannot be effectively distinguished from false positives. To balance the seam detection

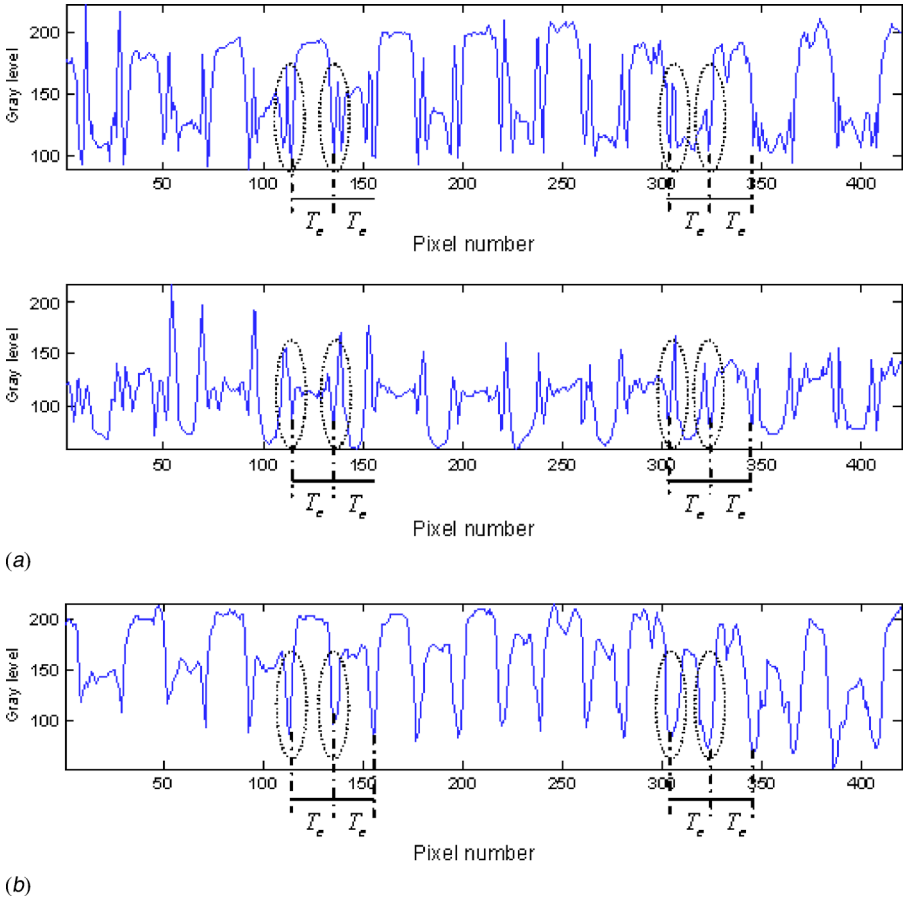


Fig. 6 (*a*) Sequences of the ridge-based false positive sub-images in Fig. 3(*a*) by snake projection (k=8,s=5) (top: corresponding to the sub-image with two dark strips; bottom: corresponding to the sub-image with one dark strip), (*b*) sequence of the mark-based false positive sub-image in Fig. 3(*b*) by snake projection (k=8,s=5)

speed and accuracy, the values of k and s must be cautiously selected. A detailed discussion of this problem can be found in Sec. 4.4.

3.2 Feature Extraction by Discrete Wavelet Transform. Based on the specific time[1] -frequency characteristics of the 1-D sequences, DWT is utilized to decompose the sequences, and then partial wavelet coefficients are selected as the critical features that distinguish seams from false positives.

3.2.1 Time-Frequency Characteristics of the Sequences. The time-frequency characteristic that is important for distinguishing seams from false positives is the local time-domain behavior of a sequence at a specific frequency f_c (defined in (3)). Specifically, the local shape of a sequence at the boundaries of each cycle T_c can be used to distinguish a seam sequence from a ridge-based false positive sequence; and the width variation of the valleys at the boundaries of the cycles can be used to distinguish a seam sequence from a mark-based false positive sequence. For illustration purposes, several cycle boundaries are highlighted by dash-lined circles for each sequence in Figs. 5 and 6. The shape difference and width variation of the valleys at the boundaries can be clearly seen.

Typically, Fourier transform [7] is used to study the time-frequency behavior of a signal. However, it has two limitations. First, in transforming a signal to the frequency domain, the time information is lost. This is because Fourier transform produces only one coefficient corresponding to a frequency such that it discards the information on how the signal behavior changes over time. This deficiency of Fourier transform makes it ineffective in capturing the width variation of valleys in a mark-based false positive sequence. As a result, the mark-based false positive may not be distinguished from a seam. Second, because the basis function of Fourier transform is sinusoidal and the shape of a sinusoid does not resemble the local shape of a seam or a false positive sequence at the boundaries of each cycle T_c, the magnitude of the Fourier coefficient at f_c will not be significantly different between the seam and false positive sequences. Thus, this coefficient can hardly be used for distinguishing the sequences.

In contrast, wavelet transform is more favorable in this application, as it overcomes the above limitations of Fourier transform. First, it produces a set of coefficients corresponding to a frequency and each coefficient describes the signal behavior at a particular location in the time domain. Thus, by applying wavelet transform to a mark-based false positive sequence, the width variation of the valleys can be preserved. Second, there are various wavelet bases with different shapes. If a wavelet basis is chosen that has a shape similar to the local shape of a seam sequence at the boundaries of each cycle T_c, the wavelet coefficients at f_c of seam sequences should be larger than those of false positive sequences. Thus, those coefficients can be used for distinguishing the sequences. Due to this consideration, SYM4, a wavelet basis from the Symlets family, is adopted in this research.

3.2.2 Introduction of DWT and Physical Interpretation of Wavelet Coefficients. Let $\mathbf{R}$ denote the real set and $L^2(\mathbf{R})$ be the space of square integrable real functions defined on $\mathbf{R}$. If $g(x) \in L^2(\mathbf{R})$, it can be expressed as:

$$g(x) = \sum_{h \in \mathbf{Z}} c_{l_0,h} \phi_{l_0,h}(x) + \sum_{l=l_0}^{\infty} \sum_{h \in \mathbf{Z}} d_{l,h} \psi_{l,h}(x)$$

where $\mathbf{Z}$ is the integer set, i.e., $l_0, l, h \in \mathbf{Z}$, and $\phi(x)$ and $\psi(x)$ are known as the scaling function and the wavelet function, respectively. These two functions can be used to create a set of time-frequency atoms through dilation and translation, thus composing an orthonormal basis for $L^2(\mathbf{R})$. Consequently, the wavelet coefficients can be computed by

$$c_{l_0,h} = \int_{\mathbf{R}} g(x) \phi_{l_0,h}(x) dx \quad \text{and} \quad d_{l,h} = \int_{\mathbf{R}} g(x) \psi_{l,h}(x) dx$$

where $c_{l_0,h}$ are approximation coefficients and $d_{l,h}$ are detail coefficients.

DWT is a special case of wavelet transform that is applied to a signal $\mathbf{X} = (x_1, \ldots, x_N)$ of size N. An efficient way to implement DWT was developed by Mallat [8]. In Mallat's algorithm, the decomposition starts with the signal $\mathbf{X}$, which can be considered as the 0th level approximation coefficients $\mathbf{C}_0$. Then $\mathbf{X}$ is convolved with a low-pass filter followed by down-sampling to get the first-level approximation coefficients $\mathbf{C}_1$. Concurrently, $\mathbf{X}$ is convolved with a high-pass filter followed by down-sampling to get the first-level detail coefficients, $\mathbf{D}_1$. The next step splits $\mathbf{C}_1$ into two parts using the same scheme, producing $\mathbf{C}_2$ and $\mathbf{D}_2$, and so on. Thus, the vector of DWT coefficients, denoted by $\mathbf{\Lambda}$, at decomposition level d, is composed of $[\mathbf{C}_d, \mathbf{D}_d, \mathbf{D}_{d-1}, \ldots, \mathbf{D}_1]$.

Because approximation coefficients $\mathbf{C}_d$ are obtained from a low-pass filter, they capture the low-frequency content of a signal. In a seam or false positive sequence, this content is the baseline drift of gray levels, reflecting the lighting condition when the sensing image is taken.

The detail coefficients at different levels (i.e., $\mathbf{D}_d, \mathbf{D}_{d-1}, \ldots, \mathbf{D}_1$) capture the transient features of the sequence at different frequencies. For example, the densely occurring squiggles in the sequence, which are introduced by image noise, are mostly captured by $\mathbf{D}_1$, because these squiggles have a very high frequency. More importantly, the local time-domain behavior of the sequence at frequency f_c, which is the key for distinguishing seams from false positives, is captured by one of the $\mathbf{D}_i$'s $(i=2, \ldots, d)$, denoted by $\mathbf{D}_{j_c}$. Note that j_c must be larger than 1 because f_c is lower than the frequency of the noise-induced squiggles.

3.2.3 Selection of the Decomposition Level. To distinguish seams from false positives, it is important to find $\mathbf{D}_{j_c}$, i.e., the detail coefficients corresponding to frequency f_c. Once $\mathbf{D}_{j_c}$ are identified, the decomposition level can be determined.

It was pointed out in [9] that the detail coefficients $\mathbf{D}_j$ $(j \in \{1, \ldots, d\})$ produced by Mallat's algorithm can be equivalently obtained by passing a signal through a high-pass filter with pass-band $1/2^{j+1} \le f \le 1/2^j$; namely, $\mathbf{D}_j$ capture the contents of the signal at frequencies $1/2^{j+1} \le f \le 1/2^j$. Therefore, if $1/2^{j_c+1} \le f_c \le 1/2^{j_c}$, the detail coefficients corresponding to f_c are $\mathbf{D}_{j_c}$. Take base 2 logarithm of $1/2^{j_c+1} \le f_c \le 1/2^{j_c}$ and express j_c as a function of f_c,

$$j_c = \lfloor -\log_2 f_c \rfloor \tag{4}$$

where $\lfloor \bullet \rfloor$ rounds a number to the nearest integer that is smaller than or equal to this number. Replace the f_c in (4) with the right-hand side of (3),

$$j_c = \lfloor \log_2(2k+s) \rfloor \tag{5}$$

Thus, for a given sequence (i.e., known k and s), the detail coefficients that are used to distinguish seams from false positives, $\mathbf{D}_{j_c}$, can be obtained from (5). j_c is also the decomposition level d of DWT because there is no need for further decomposition to generate higher-level detail coefficients with the discriminating coefficients (i.e., $\mathbf{D}_{j_c}$) already identified, i.e.,

$$d = \lfloor \log_2(2k+s) \rfloor \tag{6}$$

3.2.4 Handling Border Conditions. To apply Mallat's algorithm for decomposing a finite-length sequence, it is required to specify a border extension method for any wavelet basis with a filter length larger than two [10]. Commonly used border exten-

[1]Here, the "time" is not clock time, but is defined in a broad sense, as the order of the pixels in the sequence.

sion methods such as zero padding, wraparound, and symmetric extension [11] add artificial points to the sequence borders. The artificial points may create abrupt changes at the sequence borders in the time or frequency domain. Because these abrupt changes may be confounded with the critical time-frequency characteristics of the sequences used to distinguish seams from false positives, seam detection effectiveness may be reduced. Moreover, because both seams and false positives have a finite length on the rolling surface, the sequences are finite in nature and cannot reasonably be considered as a portion of any infinite sequences. Therefore, the sequence is "distorted" at its borders regardless of the border extension method used.

Based on the above considerations, the following strategy is proposed to handle border conditions. First, artificial points are added to the borders to extend the sequence. Second, Mallat's algorithm is adopted to decompose the sequence and $\mathbf{D}_{j_c}$ are obtained. Finally, the boundary coefficients are removed from $\mathbf{D}_{j_c}$, where a "boundary coefficient" is a wavelet coefficient whose computation involves any of the artificial points added at the sequence borders. The details to find the boundary coefficients will be discussed in Sec. 3.2.5.

Note that for a sequence of length N, the number of boundary coefficients at each decomposition level for a certain wavelet basis is invariant regardless of the border extension method selected to add the artificial points. Thus, the border extension method can be selected for convenience.

3.2.5 Feature Extraction From Wavelet Coefficients. The features for distinguishing seams from false positives are given:

$$\bar{\mathbf{D}}_{j_c}^B = \mathbf{D}_{j_c} \setminus \mathbf{D}_{j_c}^B \tag{7}$$

i.e., $\bar{\mathbf{D}}_{j_c}^B$ is the complement of $\mathbf{D}_{j_c}^B$ with respect to $\mathbf{D}_{j_c}$, where $\mathbf{D}_{j_c}^B$ are the detail boundary coefficients at the j_cth level of wavelet decomposition.

$\mathbf{D}_{j_c}^B$ consist of several consecutive coefficients at the beginning and end of $\mathbf{D}_{j_c}$, called left-boundary and right-boundary coefficients, respectively. The number of left-boundary coefficients does not depend on sequence length N [10], whereas that of right-boundary coefficients slightly varies with N. In addition to N, the number of boundary coefficients also depends on the length of the wavelet filter, denoted by L, and the decomposition level j_c. For example, the number of left-boundary coefficients at $j_c=4$ for SYM4 wavelet filter ($L=8$) is 6; and the number of right-boundary coefficients is 6 if $0 \le \mathrm{mod}(N,16) \le 6$, and 7 otherwise, where $\mathrm{mod}(N,16)$ returns the remainder of $N/16$. Thus, the total number of boundary coefficients for SYM4, denoted by $\dim(\mathbf{D}_{j_c}^B)$, is

$$\dim(\mathbf{D}_4^B) = \begin{cases} 12 & \text{if } 0 \le \mathrm{mod}(N,16) \le 6 \\ 13 & \text{otherwise} \end{cases} \tag{8}$$

The dimension of the feature space is the difference between the number of coefficients in $\mathbf{D}_{j_c}$ and that of boundary coefficients $\mathbf{D}_{j_c}^B$. The number of wavelet coefficients at a certain decomposition level j_c for a sequence with length N can be found in [12]. As for the above example, the number of wavelet coefficients at $j_c=4$ is

$$\dim(\mathbf{D}_4) = \begin{cases} (N - \mathrm{mod}(N,16))/16 + 6 & \text{if } 0 \le \mathrm{mod}(N,16) \le 6 \\ (N - \mathrm{mod}(N,16))/16 + 7 & \text{otherwise} \end{cases} \tag{9}$$

Therefore, the dimension of the feature space can be computed as

$$\dim(\mathbf{D}_4) - \dim(\mathbf{D}_4^B) = (N - \mathrm{mod}(N,16))/16 - 6 \tag{10}$$

3.3 Seam Identification by T^2 Control Chart. In multivariate process monitoring, the T^2 control chart [13] has been widely adopted to distinguish two process conditions: in control and out of control. Due to its computational ease, on-line rapid detection of process change can be achieved. In addition, constructing a T^2 control chart requires only the in-control data to follow a multivariate normal distribution, which is a favorable property if the out-of-control data come from heterogeneous sources having mixed patterns such as the features of different types of false positives. Therefore, a T^2 control chart [13] is used for on-line monitoring of the product surface and the detection of seams, in which the features of seam sub-images are treated as in-control data and their normality is verified by a Mardia's skewness-kurtosis test and a chi-square quantile-quantile plot [14].

In using the T^2 control chart for seam identification, the T^2 statistic of the features of a sub-image is first computed by

$$T^2 = (\bar{\mathbf{D}}_{j_c}^B - \hat{\boldsymbol{\mu}}_{\bar{\mathbf{D}}_{j_c}^B})^T \mathbf{S}_{\bar{\mathbf{D}}_{j_c}^B}^{-1} (\bar{\mathbf{D}}_{j_c}^B - \hat{\boldsymbol{\mu}}_{\bar{\mathbf{D}}_{j_c}^B})$$

where $\bar{\mathbf{D}}_{j_c}^B$ is the column vector of features, defined in (7), and $\hat{\boldsymbol{\mu}}_{\bar{\mathbf{D}}_{j_c}^B}$ and $\mathbf{S}_{\bar{\mathbf{D}}_{j_c}^B}$ are sample mean and sample covariance matrices of $\bar{\mathbf{D}}_{j_c}^B$, respectively, estimated using a training data set of seams. The T^2 statistic is further compared with the control limit given by

$$\mathrm{UCL} = \frac{(n-1)\dim(\bar{\mathbf{D}}_{j_c}^B)}{n - \dim(\bar{\mathbf{D}}_{j_c}^B)} F_{\dim(\bar{\mathbf{D}}_{j_c}^B), n-\dim(\bar{\mathbf{D}}_{j_c}^B)}(\alpha)$$

where n is the sample size of the training data set, $\dim(\bar{\mathbf{D}}_{j_c}^B)$ is the number of features, and $F_{\dim(\bar{\mathbf{D}}_{j_c}^B), n-\dim(\bar{\mathbf{D}}_{j_c}^B)}(\alpha)$ is the upper (100α)th percentile of a F-distribution with degrees of freedom $\dim(\bar{\mathbf{D}}_{j_c}^B)$ and n-$\dim(\bar{\mathbf{D}}_{j_c}^B)$. If the T^2-statistic is smaller than the control limit UCL, the sub-image is considered to be in control, i.e., containing a seam; otherwise, it is considered to be out of control, i.e., containing a false positive.

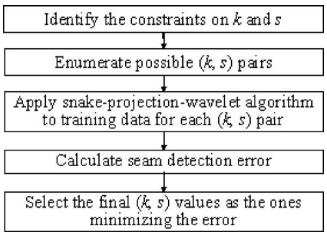


Fig. 7 Flow chart of the procedure for selecting snake-projection parameters

3.4 Selection of Snake-Projection Parameters. The snake-projection method requires specifying the values of two parameters, i.e., k and s. Different values of k and s result in different sequences in terms of the number of cycles a sequence contains and the length of each cycle. Because the subsequent algorithms are based on the sequences, this variability will propagate, impacting the feature extraction and further impacting the effectiveness of using the T^2 control chart for seam identification. To select k and s, the snake-projection-wavelet algorithm is applied to training data for each possible (k,s) pair subject to a set of constraints. The final (k,s) is the one that minimizes the seam detection error. This procedure is illustrated in Fig. 7. As constraint identification is the key part of the procedure, it is discussed in detail in this section.

3.4.1 Constraint Set by Sub-Image Boundaries. The sequences are confined by sub-image boundaries that result in constraints on k and s. In an $m \times n$ sub-image matrix, because k determines the

Table 1 Dimension of feature space

	s										
k	2	3	4	5	6	7	8	9	10	11	12
7	19	11	7	5	4	3	2	—	—	—	—
8	22	13	9	7	5	4	3	—	2	—	—
9	25	15	10	8	6	4	3	2	2	—	—
10	—	17	12	9	7	5	4	3	3	—	2

number of pixels to be included in the sequence on each side of the characteristic line, the following constraint needs to be satisfied:

$$\begin{cases} t_i + k \le n \\ t_i - k \ge 1 \end{cases} \quad \text{for any } 1 \le i \le m$$

i.e., $k \le \min\{n - t_i, t_i - 1\}$ for any $1 \le i \le m$. Thus,

$$k \le \min_{1 \le i \le m} \{\min\{n - t_i, t_i - 1\}\} \tag{11}$$

There are two constraints on s. First, $s \ge 1$ by definition. Second, $s = \lfloor m/q \rfloor$, where q is the number of rows that are extracted from the sub-image and included in the sequence. Because a mark-based false positive differs from a seam in the width variation of the dark strip in the sub-image, at least two rows have to be extracted to capture this variation; i.e., $q \ge 2$. Thus, the constraints on s can be given by

$$1 \le s \le \lfloor m/2 \rfloor \tag{12}$$

3.4.2 Constraint Set by Seam and False Positive Characteristics. The fundamental difference between seams and false positives was discussed in Sec. 2. To capture this difference by snake projection, it is important that k is sufficiently large such that the pixels on the bright strips in ridge-based false positive sub-images and those on the widest part of the dark strips in mark-based false positive sub-images are included in the sequences. Thus,

$$k \ge k_E \tag{13}$$

where k_E can be obtained from the engineering knowledge regarding the width of the ridges and marks on the rolling surface.

3.4.3 Constraint Set by Processing Speed Requirement. Because DWT takes over 90% of the processing time of the entire algorithm and the speed of DWT is linearly related to the sequence length N [9], N needs to be sufficiently small such that the algorithm is fast enough for on-line implementation.

Let t_w denote the processing time of DWT for a given wavelet basis. The relationship between t_w and N can be denoted by $t_w = a_w + b_w N$, where a_w and b_w can be obtained from simulations. Let t_E denote the engineering-specified processing time of the algorithm for on-line seam detection. Then $t_w < 0.9 t_E$. Thus,

$$N < (0.9 t_E - a_w)/b_w \tag{14}$$

It can be easily derived that N is a function of k and s for a given sub-image with m rows, i.e.,

$$N = (2k + s)\lfloor m/s \rfloor \tag{15}$$

Thus, by inserting (15) into (14), a constraint on k and s can be obtained:

$$(2k + s)\lfloor m/s \rfloor \le (0.9 t_E - a_w)/b_w \tag{16}$$

3.4.4 Constraint Set by Required Feature Space Dimension. Because a mark-based false positive sequence is different from a seam sequence in that the width of the valleys at the cycle boundaries varies, at least two features are needed to capture the width variation, i.e., $\dim(D_{j_c}) - \dim(D^B_{j_c}) \ge 2$. Using (10), $[N - \mathrm{mod}(N, 16)]/16 - 6 \ge 2$. Replacing N with the right-hand side of (15) and reorganizing the terms,

$$(2k + s)\lfloor m/s \rfloor - \mathrm{mod}((2k + s)\lfloor m/s \rfloor, 16) \ge 128 \tag{17}$$

4 Case Study

In this case study, the rolling bars from which the sensing images were taken have an average diameter of 16.8 mm and an average moving velocity of 1.1 km/min. The training data consists of 200 sub-images of seams and 200 sub-images of false positives.

The values of k and s that satisfy the constraints in (11)–(13), (16), and (17) are listed in Tables 1 and 2. Table 1 gives the dimension of the feature space. Table 2 gives the seam detection error (the proportion of false positives plus false negatives).

Several conclusions can be drawn from the results:

(1) The optimal setting for k and s based on Table 2 is $k=8$ and $s=5$, which results in a minimum error of 0.02. Because the engineering specification for seam detection accuracy is 0.05, several other settings for k and s are also acceptable, such as $\{k=7, s=4\}$, $\{k=8, s=4\}$, and $\{k=9, s=5\}$.
(2) $k=9$ and 10 have relatively larger errors than $k=7$ and 8.
(3) An adequate number of features is seven to nine. With too few features, the width variation of the valleys at the boundaries of the cycles in a mark-based false positive cannot be captured so it can hardly be distinguished from a seam. In fact, most of the misclassified samples are mark-based false positives when $\dim(D_4) - \dim(D^B_4) < 5$.
(4) Having too many features also results in large errors because the features of seams cannot be well approximated by a multivariate normal distribution when $\dim(D_4) - \dim(D^B_4) > 15$. As an example, the chi-square quantile-quantile plots for the features of 100 seams are shown in

Table 2 Seam detection errors based on training data

	s										
k	2	3	4	5	6	7	8	9	10	11	12
7	0.18	0.09	0.05	0.07	0.1	0.13	0.15	—	—	—	—
8	0.18	0.09	0.04	0.02	0.09	0.12	0.17	—	0.15	—	—
9	0.2	0.14	0.07	0.05	0.11	0.13	0.18	0.18	0.17	—	—
10	—	0.16	0.1	0.08	0.11	0.13	0.21	0.2	0.2	—	0.23

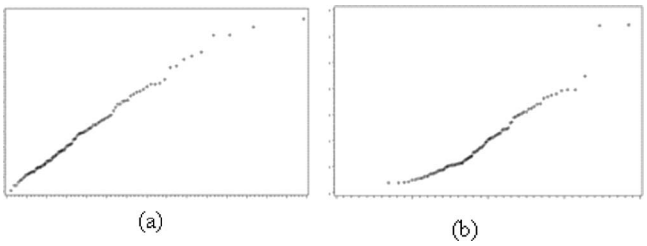


Fig. 8 Chi-square quantile-quantile plot of features (*x*-axis: chi-square quantiles; *y*-axis: squared Mahalanobis distances [14] of features; the labels for *x*- and *y*-axes are omitted)

Figs. 8(*a*) and 8(*b*), where (*a*) corresponds to $k=8$, $s=5$, $\dim(D_4)-\dim(D_4^B)=7$; and (*b*) corresponds to $k=8$, $s=2$, $\dim(D_4)-\dim(D_4^B)=22$. The plots in Fig. 8(*a*) have a straight-line pattern, indicating that the assumption of multivariate normality is valid. However, an apparent curved pattern can be seen in Fig. 8(*b*) suggesting lack of normality.

The optimal setting $k=8$ and $s=5$ and the corresponding seam detection algorithm are applied to a separate testing dataset with 100 seam sub-images and 100 false positive sub-images. The result is shown in Fig. 9. The control limit (UCL) is computed at significant level 0.05. Any point falling bellow UCL is considered to be a seam. Otherwise, it is considered to be a false positive.

The 100 seam samples are plotted in the first half of Fig. 9 (left of the vertical dash line), among which two samples are misidentified as false positives. The 100 false positive samples are plotted in the second half of the figure (right of the vertical dash line), among which three samples are misidentified as seams. Therefore, the seam detection error is $5/200=0.025$. In addition, the detection speed of the snake-projection-wavelet algorithm is 40% faster than the Canny edge detector. Implementing the algorithm in an industrial test site showed that the speed satisfies the engineering specification for on-line seam detection.

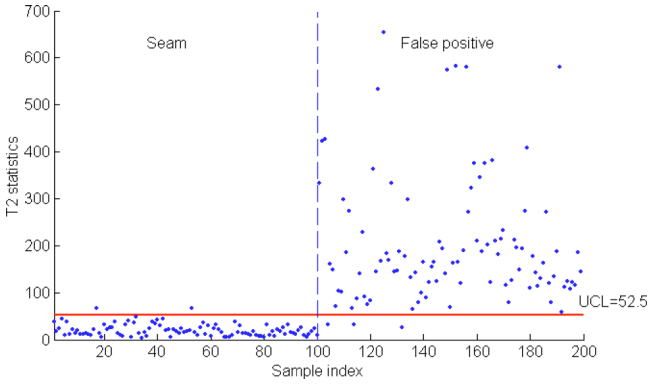


Fig. 9 T^2 control chart at $k=8$, $s=5$ for the testing dataset

5 Conclusion

This paper proposed a snake-projection-wavelet algorithm to detect seams in rolling processes. For data and dimension reductions, a snake-projection method was developed to convert the suspicious seam-containing sub-images into 1-D sequences. A discrete wavelet transform was then performed on the sequence with features extracted from wavelet coefficients. Finally, a T^2 control chart was constructed to distinguish seams from false positives based on the features. Because the snake-projection method has two parameters that jointly determine the data reduction rate and further impact the detection speed and accuracy of the algorithm, how to select the values for these two parameters was discussed. A case study was presented to demonstrate the effectiveness of the algorithm, yielding a seam detection error of 0.025, which is below the engineering specification of 0.05. In addition, on-line testing of the algorithm showed that the detection speed is adequate for vision sensor based automatic seam detection. To assess the robustness of the algorithm with respect to different noise levels and characteristics, this algorithm has been implemented in daily production to detect seams associated with different materials and operational conditions (e.g., rolling speed and lubrication), which are the two key factors impacting the noise levels/characteristics. It is found that this algorithm can successfully detect seams with respect to these different materials and operational conditions.

Acknowledgment

The authors would like to thank OG Technologies, Inc. in Ann Arbor for its invaluable support and contribution to this research.

References

[1] Kalpakjian, S., and Schmid, S. R., 2003, *Manufacturing Processes for Engineering Materials*, Prentice Hall, Upper Saddle River, NJ.
[2] Ziou, D., and Tabbone, S., 1998, "Edge Detection Techniques—An Overview," Int. J. Pattern Recognit. Artif. Intell., **8**(4), pp. 537–559.
[3] Canny, J., 1986, "A Computational Approach to Edge Detection," IEEE Trans. Pattern Anal. Mach. Intell., **8**(6), pp. 679–698.
[4] Basu, M., 2002, "Gaussian-Based Edge-Detection Methods—A Survey," IEEE Trans. Syst. Sci. Cybern., **32**(3), pp. 252–260.
[5] Nixon, M. S., and Aguado, A. S., 2002, *Feature Extraction and Image Processing*, Newnes, Boston, MA.
[6] Li, J., Gutchess, D., Shi, J., and Chang, S., 2003, "Real-Time Surface Defect Detection in Hot Rolling Process," *Proceedings of the Iron and Steel Exposition and 2003 AISE Annual Convention*, Pittsburgh, PA.
[7] Bracewell, R. N., 2000, *The Fourier Transform and Its Applications*, McGraw Hill, Boston.
[8] Mallat, S. G., 1989, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," IEEE Trans. Pattern Anal. Mach. Intell., **11**, pp. 674–693.
[9] Sidney, C. S., Gopinath, R. A., and Gao, H., 1998, *Introduction to Wavelet Transformations: A Primer*, Prentice Hall, Upper Saddle River, NJ.
[10] Percival, D. B., and Walden, A. T., 2000, *Wavelet Methods for Time Series Analysis*, Cambridge University Press, Cambridge, New York.
[11] Strang, G., and Nguyen, T., 1996, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley, MA.
[12] MATLAB User's Manual, v6.1, MathWorks Inc., Natick, MA.
[13] Montgomery, D. C., 2001, *Introduction to Statistical Quality Control*, John Wiley & Sons, Inc., New York.
[14] Mardia, K. V., 1980, "Tests of Univariate and Multivariate Normality," *Handbook of Statistics*, North-Holland, Amsterdam, Vol. 1, pp. 279–320.