Technometrics

ISSN: 0040-1706 (Print) 1537-2723 (Online) Journal homepage: https://www.tandfonline.com/loi/utch20

Anomaly Detection in Images With Smooth Background via Smooth-Sparse Decomposition

Hao Yan, Kamran Paynabar & Jianjun Shi

To cite this article: Hao Yan, Kamran Paynabar & Jianjun Shi (2017) Anomaly Detection in Images With Smooth Background via Smooth-Sparse Decomposition, Technometrics, 59:1, 102-114, DOI: 10.1080/00401706.2015.1102764

To link to this article: https://doi.org/10.1080/00401706.2015.1102764

Supplementary materials for this article are available online. Please go to http://www.tandfonline.com/r/TECH

Anomaly Detection in Images With Smooth Background via Smooth-Sparse Decomposition

Hao YAN, Kamran PAYNABAR, and Jianjun SHI

H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332
(*yanhao@gatech.edu; kamran.paynabar@isye.gatech.edu; jianjun.shi@isye.gatech.edu*)

In various manufacturing applications such as steel, composites, and textile production, anomaly detection in noisy images is of special importance. Although there are several methods for image denoising and anomaly detection, most of these perform denoising and detection sequentially, which affects detection accuracy and efficiency. Additionally, the low computational speed of some of these methods is a limitation for real-time inspection. In this article, we develop a novel methodology for anomaly detection in noisy images with smooth backgrounds. The proposed method, named smooth-sparse decomposition, exploits regularized high-dimensional regression to decompose an image and separate anomalous regions by solving a large-scale optimization problem. To enable the proposed method for real-time implementation, a fast algorithm for solving the optimization model is proposed. Using simulations and a case study, we evaluate the performance of the proposed method and compare it with existing methods. Numerical results demonstrate the superiority of the proposed method in terms of the detection accuracy as well as computation time. This article has supplementary materials that includes all the technical details, proofs, MATLAB codes, and simulated images used in the article.

KEY WORDS: Anomaly detection; Convex optimization; High-dimensional; Image; Regression; Smooth background

1. INTRODUCTION

Image sensing systems have been widely deployed in a variety of manufacturing processes for online process monitoring and fault diagnosis. The reasons for this range from their low implementation cost and high acquisition rate of image sensors to the rich process information they provide. One of the main applications of these systems is real-time product inspection in which a snapshot of a product or part is analyzed to detect defects or anomalies. One example is in continuous casting manufacturing where molten metal is solidified into a semi-finished billet used in the subsequent rolling process. To inspect the quality of billets and detect anomalies on their surfaces, a vision sensing system is set up to take snapshots of billets at short time intervals. A sample of the surface image with a vertical curved line defect is shown in Figure 1. Considering the high speed of production, an automatic, quick, and accurate image analysis technique is crucial to an effective quality inspection and anomaly detection system.

Another example of image-based quality inspection, as shown in Figure 2, is in the photoelasticity test (Prasath, Skenes, and Danyluk 2013). The photoelasticity test is a nondestructive evaluation method used for stress and strain analysis of translucent parts or material. The output is often presented by a colormap in which regions with high-tensile stress, often associated with anomalies, are shown by warm colors. The stress maps of a silicon beam and a silicon surface laminate with surface indentation are shown in Figure 2(a) and 2(b), respectively. Although anomalous regions are clear in both samples, developing an automatic algorithm that can accurately detect and separate these regions from the image background in real time is imperative for effective process monitoring. Other applications where image-based inspection and anomaly detection have been used include the rolling process (Jin, Zhou, and Chang 2004), composite material fabrication (Sohn et al. 2004), liquid crystal display manufacturing (Jiang, Wang, and Liu 2005), fabric and textile manufacturing (Kumar 2008), and structural health monitoring (Balageas, Fritzen, and Güemes 2010), to name a few.

The high dimensionality and complex spatial structure of images coupled with measurement noises that reduce contrast between anomalies and the background pose significant challenges in developing real-time anomaly detection methods. Owing to its importance, anomaly detection in images has been extensively studied in the literature, and considerable research has been conducted to address these challenges. Most of existing image-based anomaly detection methods follow a common two-step procedure in which first, a smoothing or denoising method such as spline (De Boor 1978) or wavelet analysis (Wink and Roerdink 2004) is applied to reduce image noises, then a detection algorithm is exploited to identify anomalous regions. The major problem of the two-step approach is that the smoothing step often blurs the sharp boundaries of anomalous regions, which in turn makes the detection step challenging (Nagao and Matsuyama 1979). Qiu (2007) and Qiu and Mukherjee (2012) developed edge-preserved smoothing techniques to preserve the boundary structure applying the smoothing algorithm. Although effective methods for denoising of images with a general

TECHNOMETRICS, FEBRUARY 2017, VOL. 59, NO. 1
DOI: 10.1080/00401706.2015.1102764
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/tech.

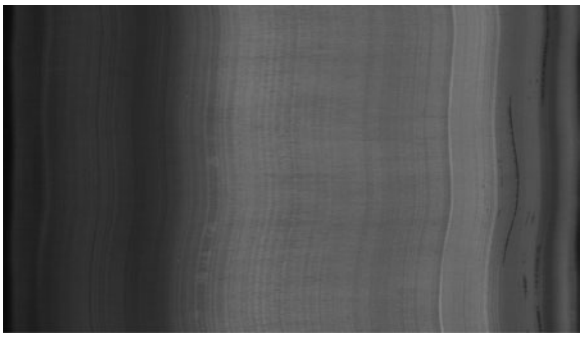


Figure 1. A sample image for online surface inspection in rolling system.

background structure, they should be followed by an anomaly detection or image segmentation algorithm to identify anomalous regions. This may increase the analysis time.

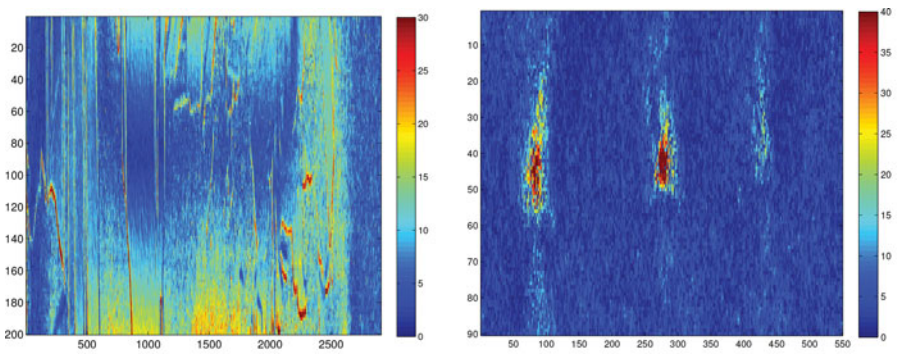
In this article, we propose an accurate and fast method for image-based anomaly detection that overcomes the drawbacks of existing two-step methods. Our method integrates smoothing and anomaly detection tasks into one step through a novel smooth-sparse decomposition (SSD) approach. SSD decomposes an image into three components: namely, the smooth image background, the sparse anomalous regions, and the random noises, as illustrated in Figure 3. In addition to anomaly detection, SSD helps retrieve an image by removing anomalies and random noises. SSD is developed based on the premise that the image background is smooth and anomalous regions are sparse or can be expressed by sparse bases. Decomposition is achieved by constructing a penalized nonparametric regression model that enforces background smoothness and anomaly sparsity through penalty terms added to the loss function. To estimate the parameters of the regression model and perform the decomposition in real-time, we propose efficient optimization algorithms.

The remainder of the article is organized as follows. Section 2 reviews existing methods for image-based anomaly in the literature. Section 3 elaborates the one-step SSD approach for anomaly detection and presents efficient optimization algorithms for the real-time implementation of SSD. In Section 4, we use simulated image data with different anomaly structures to evaluate the performance of the proposed SSD method and compare it with some existing two-step methods in terms of detection accuracy as well as computation time. In Section 5, we illustrate a case study in which we apply the proposed SSD for anomaly detection in silicon samples. We conclude the article with a short discussion and an outline of our future work in Section 6.

2. LITERATURE REVIEW

As mentioned earlier, considerable research has been conducted on anomaly detection in single image observation, most of which is based on a two-step approach comprised of denoising and detection. For the denoising step, smoothing methods such as spline regression (De Boor 1978), B-spline (Unser 1999), penalized spline (Eilers and Marx 1996), kernel smoothing (Wand and Jones 1994), and various filtering techniques (Milanfar 2013) have been widely used. To preserve the boundaries of anomalous regions while applying smoothing, Qiu (2007) and Qiu and Mukherjee (2012) developed edge-preserved smoothing techniques by assigning small weights to the observations located on either side of the boundary. In the area of detection, a considerable body of research focuses on anomaly detection in images with patterned background—for example, images taken from textured materials. Most methods in this category rely upon identifying the areas that differ from the background pattern. Examples include analysis of variance (Jiang, Wang, and Liu 2005), the filter technique (Kumar and Pang 2002), SVD decomposition (Lu and Tsai 2004), wavelet transformation (Sohn et al. 2004), Fourier transformation (Chan and Pang 2000), etc. These methods, however, are not effective when the background of an image is smooth.

Funck et al. (2003) categorized current anomaly detection methods, applicable to images with smooth backgrounds, into three categories: edged-based, thresholding-based, and region-based methods. Edge-based methods focus on locating sharp discontinuities in the image. Most traditional edge-detection methods are based on gradient estimation, that is, the estimation of first- or second-order derivatives of the image intensity function. For example, Sobel (Sobel and Feldman 1968) and Prewitt (Prewitt 1970) are the two widely used first-order derivative operators for edge detection. Laplacian operator or Laplacian of Gaussianedge detector (Marr and Hildreth 1980) is a second-order derivative operator for edge detection. However, gradient-based edge-detection methods are not robust to noise because the edge-detection mask is usually small. To overcome this problem, Qiu and Yandell (1997) proposed jump regression that uses local surface information. Jump regression enables precise



(a) Stress map of multi-crystalline bending beam

(b) Stress map of silicon surface with indentation

Figure 2. Examples of stress maps in photoelasticity experiment.

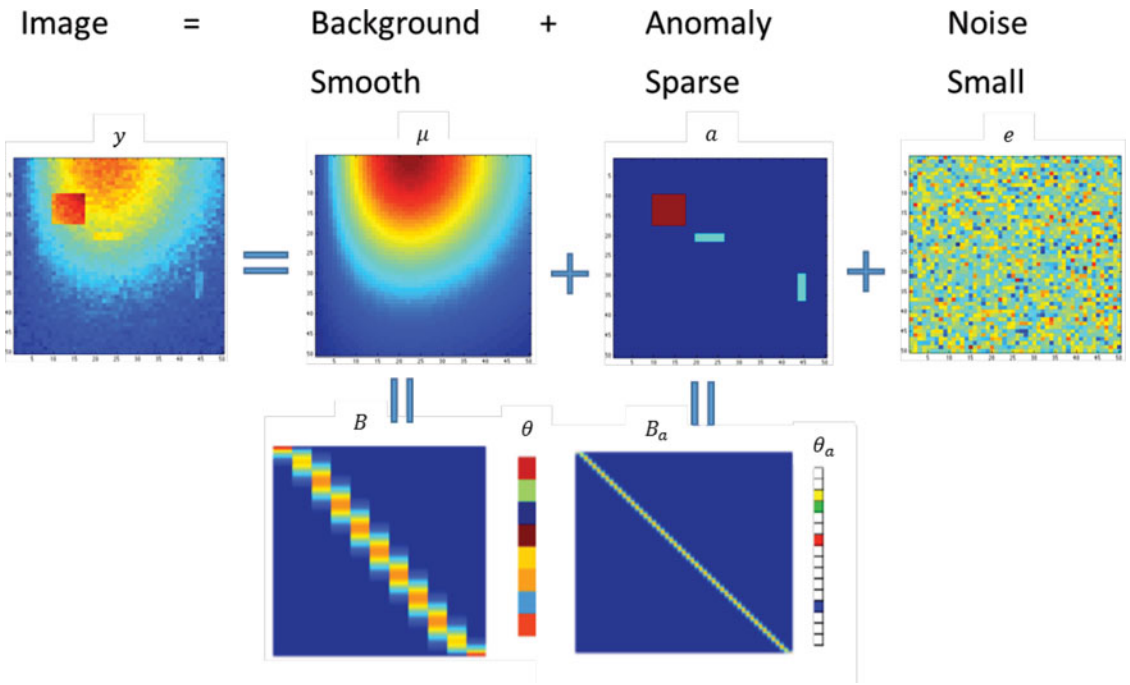


Figure 3. Decomposition of image to background, anomaly, and noise.

estimation of jump locations or edges on a smooth background in a noisy setting (Qiu and Yandell 1997; Qiu 2007). The output of edge-detection algorithms are often discontinuous pixels, which may not form a closed-boundary region or continuous curves. Consequently, most edge-based methods require a post-processing step such as edge linking or filling algorithm to link them for creating closed-boundary regions and fill the area inside these regions. The main problem of edge-based methods is that even after applying edge-linking algorithms, the detected edges still do not form a closed region (Funck et al. 2003). To address this, Qiu and Sun (2009) developed a curve estimation algorithm to effectively link scattered pixels with a closed curve. Qiu and Sun (2007) also developed a local nonparametric segmentation methods for spotted microarray images. Park et al. (2012) developed a multistage semi-automated procedure to extract the shape information of nanoparticles based on the boundary points from edge detectors. However, these methods are case-specific (i.e., for spotted microarray image detection and nanoparticles detection) and may not work well for other types of anomalies such as scattered or line defects. Moreover, they require information of the centroid location, which, in some cases, may be difficult to find or estimate. Computational speed is another issue because these methods often entail multiple processing steps.

Unlike edge-based methods, thresholding-based methods use the intensity difference between anomalies and the background to find closed-boundary regions directly. The thresholding algorithm can be further classified into two categories: global thresholding and local thresholding. Otsu's method (Otsu 1975) is one of the global thresholding methods that uses a single value obtained by minimizing in-class variation to globally threshold the entire image and produce a binary map for anomalous regions. However, as pointed by Jiang and Mojon (2003), global thresholding algorithms do not perform well when large intensity variation exists in the image background. On the contrary, local thresholding methods (e.g., Bradley and Roth 2007; Sauvola and Pietikäinen 2000) divide an image into small regions and then perform the thresholding locally in each region. There are two main problems with local thresholding methods: First, in nondefective regions, local thresholding algorithms still pick a threshold value, which leads to false detection. Second, anomalous regions are often larger than the thresholding window, which leads to inconsistent thresholding of the anomalous region.

Similarly, like local thresholding methods, region-based methods give closed-boundary regions. For example, the "seeded region growing" method (Adams and Bischof 1994) starts from a small initial region (pixels) and grows it by exploring neighbor pixels and adding pixels similar to the initial ones. However, region-based algorithms often require a small set of initial anomalous pixels from which the region can be grown. Consequently, when lacking the location information of anomalies or in the case of multiple anomalous regions, this method is not practical. Extended-maxima transformation (Soille 2013) is one of the morphological filtering methods that can be used for anomaly detection. This method searches for connected pixels with a constant intensity that are distinguished from the background by an external boundary with a different intensity. However, this method does not take advantage of the smooth structure of the background and anomalies, thus may fail when the intensity contrast between anomalies and the background is small.

There is also substantial research focusing on analysis and inference on multiple images including image registration techniques (Qiu and Xing 2013), pattern learning (Wu et al. 2007), object tracking (Comaniciu, Ramesh, and Meer 2003), image monitoring (Yan, Paynabar, and Shi 2014), and classification (Epifanio 2008) to name a few. Although important, these topics are beyond the scope of this article that focuses only on anomaly detection on single images.

3. SMOOTH-SPARSE DECOMPOSITION

In this section, we present the penalized nonparametric regression model used for SSD and propose efficient algorithms for its implementation. For simplicity, we first discuss the methodology for one-dimensional (1-D) signals and then generalize it to n-D images ($n > 1$). When the underlying manufacturing process is unstable, faulty parts may be produced. In this case, the measured signal is comprised of not only the functional mean and noises but also anomalies as shown in Figure 2(a) and 2(b). In this article, anomalies are defined as faults whose functional structure differ from the functional mean of the background and their magnitude is larger than that of the noise. Therefore, SSD aims to decompose the signal into a smooth functional mean, sparse anomalous regions, and random noises. SSD aims to decompose the signal into a smooth functional mean, sparse anomalous regions, and random noises. Specifically, the signal is decomposed as $y = \mu + a + e$, where μ is the smooth mean of the signal, a is the vector of anomalies assumed to be sparse in a certain functional space, and e is the vector of random noises. We further expand the mean and anomalies using a smooth basis (e.g., spline basis) denoted by B and B_a, respectively. Consequently, the signal decomposition model can be rewritten as $y = B\theta + B_a\theta_a + e$, where θ and θ_a are, respectively, the basis coefficients corresponding to μ and a. Least square regression is used to estimate the model parameters (i.e., θ and θ_a). To ensure the smoothness of the estimated mean and the sparsity of the detected anomalies, the least square loss function is augmented by L_1 and L_2 penalties, which results in the following penalized regression criterion:

$$\underset{\theta,\theta_a}{\operatorname{argmin}} \|e\|^2 + \lambda\theta^T R\theta + \gamma\|\theta_a\|_1, \quad \text{subject to } y = B\theta + B_a\theta_a + e, \tag{1}$$

where $\|\cdot\|$ and $\|\cdot\|_1$ are L_2 and L_1 norm operators, and λ and γ are tuning parameters to be determined by the user. R is the roughness matrix, and can be defined as $R = D^T D$ in the 1-D case, which is related to the difference between the nearby spline coefficients, that is, $\|\Delta^d\theta\|^2$, where Δ^d is the dth order difference operator. It is not hard to show this penalization term can also be written as $\|\Delta^d\theta\|^2 = \theta^T R\theta$, in which $R = D^T D$, D is the dth order difference matrix.

For example, if $k = 1$,

$$D = \begin{bmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{bmatrix}.$$

Therefore, the L_2 penalty term, $\lambda\theta' R\theta$, regularizes the level of smoothness of the mean function, while the L_1 penalty term, $\gamma\|\theta_a\|_1$, encourages the sparsity of the anomalous regions. The constraint guarantees that the signal can be reconstructed using the linear combination of the estimated components. Note that if $\lambda = 0$, the SSD model boils down to LASSO (Tibshirani 1996).

3.1 Optimization Algorithms for SSD

The loss function in (1) is convex and can be solved via general convex optimization solvers like the interior point method (Nesterov, Nemirovskii, and Ye 1994). However, the interior point method is often slow for large-scale problems and hence cannot be used for real-time inspection and monitoring. In this section, we propose a set of efficient algorithms for the real-time implementation of SSD for two cases: orthogonal- and general-anomaly bases (i.e., B_a).

3.1.1 Orthogonal Basis B_a. If the basis B_a is orthogonal, the block coordinate descent (BCD) method is used to break down the SSD model into two simpler optimization problems. The BCD is a class of algorithms that groups domain variables into different blocks and finds a local minimum for a function by iteratively minimizing this function with respect to one block given all other blocks. By defining θ and θ_a as two variable blocks, a two-step iterative algorithm based on the BCD can be used to find the minimizer of (1). In each iteration k, given $\theta_a^{(k-1)}$, the SSD loss function in (1) reduces to a weighted ridge regression, which has a closed-form solution in the form of $\theta^{(k)} = (B^T B + \lambda R)^{-1} B^T (y - B_a\theta_a^{(k-1)})$. Equivalently, it can be shown that $y^{(k)} = H(y - B_a\theta_a^{(k-1)})$, where $H = B(B^T B + \lambda R)^{-1} B^T$. In the second step of the optimization algorithm, according to Proposition 1, given $\theta^{(k)}$ or $\mu^{(k)}$, $\theta_a^{(k)}$ is updated by a simple soft-thresholding operation.

Proposition 1. If B_a is orthogonal, in iteration k, the subproblem $\theta_a^{(k)} = \operatorname{argmin}_{\theta_a} \|y - B\theta^{(k)} - B_a\theta_a\|^2 + \gamma\|\theta_a\|_1$ has a closed-form solution in the form of $\theta_a^{(k)} = S_{\frac{\gamma}{2}}(B_a^T(y - B\theta^{(k)}))$, in which $S_\gamma(x) = \operatorname{sgn}(x)(|x| - \gamma)_+$ is the soft-thresholding operator, and $\operatorname{sgn}(x)$ is the sign function and $x_+ = \max(x, 0)$.

The proof is given in Appendix A in the online supplementary materials.

The fact that both subproblems have closed-form solutions significantly speeds up the optimization algorithm. A summary of the BCD algorithm for the orthogonal case is given in Algorithm 1. Although the BCD algorithm, in general, may not converge to an optimum, even if the function is convex (Friedman et al. 2007), the following proposition guarantees that the BCD attains the global optimum of problem (1).

Proposition 2. The BCD algorithm attains the global optimum of the SSD loss function in (1).

The proof is given in Appendix B in the online supplementary materials.

Algorithm 1: Optimization algorithm for SSD based on BCD method for orthogonal B_a

initialize
  Choose a basis for the background as B and for the anomalous regionsas $B_a\theta_a^{(0)} = 0$, $k = 1$
end
while $\left\|\theta_a^{(k)} - \theta_a^{(k-1)}\right\| > \epsilon$ **do**
  Update $\mu^{(k)} = B\theta^{(k)}$ via $\mu^{(k)} = H(y - B_a\theta_a^{(k-1)})$, $H = B(B^T B + \lambda R)^{-1} B^T$
  Update $\theta_a^{(k)}$ by $\theta_a^{(k)} = S_{\frac{\gamma}{2}}(B_a^T(y - \mu^{(k)}))$
  Update $k = k + 1$
end

3.1.2 General Basis B_a. For the general basis B_a, we first show that the SSD problem can be reduced to a weighted LASSO problem via Proposition 3.

Proposition 3. The SSD problem in (1) is equivalent to a weighted LASSO problem in the form of

$$\arg\min_{\theta_a} F(\theta_a) = (y - B_a\theta_a)^T(I - H)(y - B_a\theta_a) + \gamma\|\theta_a\|_1 \tag{2}$$

with $H = B(B^TB + \lambda R)^{-1}B^T$.

The proof is given in Appendix C in the online supplementary materials.

Common LASSO solvers such as least angle regression (LARS) (Efron et al. 2004) and quadratic programming (Tibshirani 1996) cannot solve the above weighted LASSO problem for high-dimensional data. For example, for an image of the size 350 by 350, that is, $p \approx 10^5$, the LARS algorithm (Efron et al. 2004) will find the entire solution path in about 60 hr, which is impractical for real-time purposes. Alternatively, we develop an efficient algorithm based on the accelerated proximal gradient method (Tseng 2008) for solving the large-scale optimization problem in (2).

The proximal gradient (PG) method is a class of optimization algorithms focusing on minimization of the summation of a group of convex functions, some of which are nondifferentiable. The function $F(\theta_a)$ in (2) is comprised of $f(\theta_a) = (y - B_a\theta_a)^T(I - H)(y - B_a\theta_a)$, which is convex differentiable if R is a positive semidefinite matrix (see Appendix D in the online supplementary materials for the proof of convexity) and $g(\theta_a) = \gamma\|\theta_a\|_1$, which is a nondifferentiable function. Another assumption of the PG algorithm is that the continuous part of the objective function $f(\theta_a)$ (1) is convex differentiable with the Lipschitz continuous gradient L, that is, there exists a constant L that for every $\alpha, \beta \in \mathbb{R}, \|\nabla f(\alpha) - \nabla f(\beta)\| \le L\|\alpha - \beta\|$, where $\nabla f(\cdot)$ is the gradient function. As shown in Appendix E in the online supplementary materials, $f(\theta_a)$ is Lipschitz continuous gradient with $L = 2\|B_a\|_2^2$. Provided the above assumptions, the PG method optimizes $F(\theta_a)$ through an iterative algorithm given by $\theta_a^{(k)} = \text{argmin}_{\theta_a}\{f(\theta_a^{(k-1)}) + \langle\theta_a - \theta_a^{(k-1)}, \nabla f(\theta_a^{(k-1)})\rangle + \frac{L}{2}\|\theta_a - \theta_a^{(k-1)}\|^2 + \gamma\|\theta_a\|_1\}$, where super-indices (k) and $(k-1)$ denote iteration numbers and $\langle\cdot,\cdot\rangle$ is the inner product operator. For more details on proximal gradient method, readers can refer to Parikh and Boyd (2013). Through the following proposition, we show that in each iteration PG results in a closed-form solution for the SSD problem in the form of a soft-thresholding function.

Proposition 4. The proximal gradient method for the SSD problem in (1), given by $\theta_a^{(k)} = \text{argmin}_{\theta_a}\{f(\theta_a^{(k-1)}) + \langle\theta_a - \theta_a^{(k-1)}, \nabla f(\theta_a^{(k-1)})\rangle + \frac{L}{2}\|\theta_a - \theta_a^{(k-1)}\|^2 + \gamma\|\theta_a\|_1\}$, has a closed-form solution in each iteration k, in the form of a soft-thresholding function as follows:

$$\theta_a^{(k)} = S_{\frac{\gamma}{L}}\left(\theta_a^{(k-1)} + \frac{2}{L}B_a^T\left(y - B_a\theta_a^{(k-1)} - \mu^{(k)}\right)\right) \tag{3}$$

with $L = 2\|B_a\|_2^2$.

The proof is given in Appendix F in the online supplementary materials.

Algorithm 2: Optimization algorithm for solving SSD based on APG

initialize
  $L = 2\|B_a\|_2^2, \quad \theta_a^{(0)} = 0, \quad x^{(0)} = 0, \quad t_0 = 1, \quad k = 1$
end
while $\|\theta_a^{(k)} - \theta_a^{(k-1)}\| > \epsilon$ **do**
  Let $\mu^{(k)} = H(y - B_a\theta_a^{(k-1)})$
  Update $\theta_a^{(k)} = S_{\frac{\gamma}{L}}(x^{(k-1)} + \frac{2}{L}B_a^T(y - B_a x^{(k-1)} - \mu^{(k)}))$
  Update $t_k = \frac{1+\sqrt{1+4t_{k-1}^2}}{2}$
  Update $x^{(k)} = \theta_a^{(k-1)} + \frac{t_{k-1}-1}{t_k}(\theta_a^{(k-1)} - \theta_a^{(k-2)})$
  Update $k = k + 1$
end

The soft-thresholding solution provided by PG can significantly expedite the SSD implementation and anomaly detection. Suppose B_a is of size $n \times k_a$, and B is of size $n \times k_\mu$. The most computationally expensive operator in the soft-thresholding solution is $\theta_a^{(k-1)} + \frac{2}{L}B_a^T(y - B_a\theta_a^{(k-1)} - \mu^{(k)})$. So, the total computational complexity in each iteration is around $k_\mu^3 + 6n^2k_\mu$ flops, which is quadratic in n.

To increase the convergence speed of the proximal gradient method, Nesterov (1983) showed that with the adjustment of the step size, it is possible to achieve the quadratic convergence rate $O(\frac{1}{k^2})$. With this adjustment, the proposed optimization algorithm for solving SSD based on the accelerated proximal gradient (APG) algorithm is summarized in Algorithm 2.

It is worth noting that if B_a is orthogonal ($B_a^TB_a = I$), the PG algorithm is reduced to the BCD algorithm discussed in Section 3.1.1.

3.2 Generalization to 2-D Image Case

In the previous section, we discussed SSD for 1-D cases and presented efficient algorithms for performing SSD. In this section, we extend our formulation and algorithms to two-dimensional (2-D) images. Suppose a 2-D image $Y_{n_1\times n_2}$ is available. We define B_i and $B_{a,i}$; $i = 1, 2$ as the basis for the background and anomalous regions, respectively, where i ($i = 1, 2$) denotes the basis in the x and y direction of the image. Therefore, the tensor product of these 1-D bases (i.e., $B = B_1 \otimes B_2$) can give the proper 2-D basis for the background as well as anomalous regions. Consequently, the SSD problem for the 2-D case can be written as

$$\underset{\theta,\theta_S}{\text{argmin}}\ \|\tilde{e}\|^2 + \lambda\theta^TR\theta + \gamma|\theta_a|_1,\ \text{s.t.}\ \tilde{y} = B\theta + B_a\theta_a + \tilde{e}, \tag{4}$$

where $B = B_2 \otimes B_1$, $B_a = B_{a,2} \otimes B_{a,1}$, $\tilde{y} = \text{vec}(Y)$, $\tilde{e} = \text{vec}(e)$, $\otimes$ is the tensor product, and $\text{vec}(\cdot)$ is an operator that unfolds a matrix to a column vector. Note that the size of the resulting basis is defined by the size of their individual bases—for example, if the size of matrix B_i is $n_i \times k_{\mu_i}$, in which k_{μ_i} is the number of basis in the ith direction, then the size of B is $n_1n_2 \times k_{\mu_1}k_{\mu_2}$. Similarly, assume that $B_{a,i}$ is of size $n_i \times k_{a_i}$, then B is an $n_1n_2 \times k_{a_1}k_{a_2}$matrix.

To solve the SSD problem in (4), we can still use algorithms presented in 1-D cases. However, since both APG and BCD

algorithms require matrix inversion operations to compute the projection matrix $H = B^T(B^TB + \lambda R)^{-1}B$, and the computational complexity of the matrix inversion is nonlinearly proportional to the size of $(B^TB + \lambda R)$, that is, $O\left((k_{\mu_1}k_{\mu_2})^3\right)$, the complexity of the APG or BCD algorithm is given by a sixth-order polynomial of $(n_1, n_2, k_{\mu_1}, k_{\mu_2}, k_{a_1}, k_{a_2})$ with the leading term $(k^3_{\mu_1}k^3_{\mu_2} + 6n_1^2n_2^2k_{\mu_1}k_{\mu_2})$. Consequently, this becomes computationally intractable as the size of the image increases. To reduce the computational complexity, we define matrix R in such a way that matrix H can be computed by a tensor product of two low-dimensional matrices. Following Xiao, Li, and Ruppert (2013), we define matrix R as $R = B_2^TB_2 \otimes D_1^TD_1 + D_2^TD_2 \otimes B_1^TB_1 + \lambda D_2^TD_2 \otimes D_1^TD_1$, which results in a decomposable projection matrix, that is, $H = H_2 \otimes H_1$, where $H_i = B_i(B_i^TB_i + \lambda D_i^TD_i)^{-1}B_i^T$, with the dimensions of $k_i; i = 1, 2$. D_i is the first-order difference matrix in ith dimension. This trick makes the algorithm very efficient for 2-D images, as it requires the inversion of matrices with lower dimensions, that is, $B_i^TB_i + \lambda D_i^TD_i$. Hence, the computational complexity of the matrix inversion operation is reduced from $O((k_{\mu_1}k_{\mu_2})^3)$ to $O(k^3_{\mu_1} + k^3_{\mu_2})$.

Since $(P \otimes Q)\theta = \text{vec}(P\Theta Q)$ with Θ as the matrix form of θ, updating steps in Algorithms 1 and 2 can be shown in the matrix form as $\mu^{(k)} = H_1(B_{a,1}\Theta_a^{(k-1)}B_{a,2}^T - Y)H_2$, $\Theta_a^{(k)} = S_{\frac{\gamma}{2}}(B_a^T(Y - \mu^{(k)}))$ for BCD, and $\Theta_a^{(k)} = \text{S}_{\frac{\gamma}{2}}(X^{(k-1)} + \frac{2}{L}B_{a,1}^T(Y - B_{a.1}X^{(k-1)}B_{a,2}^T - \mu^{(k)})B_{a,2})$ for APG. The total computational complexity is computed by a third-order polynomial of $(n_1, n_2, k_{\mu_1}, k_{\mu_2}, k_{a_1}, k_{a_2})$, with the leading term $(k^3_{\mu_1} + k^3_{\mu_2} + 6n_1^2k_{\mu_1} + 6n_2^2k_{\mu_2} + 2(n_1 + n_2)k_{a_1}k_{a_2} + 2n_1n_2(n_1 + n_2) + 2n_1n_2(k_{a_1} + k_{a_2}))$, which is computationally more efficient.

3.3 Choice of Tuning Parameters λ and γ

In the SSD model, two tuning parameters, λ and γ, are used to control the smoothness of the background $\hat{\mu}$ and the sparsity of anomalous regions $\hat{a}$, respectively. A common approach for choosing these tuning parameters is to use the k-fold cross-validation method on a 2-D grid of parameters (λ, γ) and find the pair of parameters that minimizes the mean squared error. However, this approach requires solving the SSD problem for each pair of (λ, γ) on the grid, which may not be feasible. Alternatively, we propose an iterative approach that updates the tuning parameters in each iteration of the APG and BCD algorithms without exploring all pairs of (λ, γ).

We begin with some initial values for the tuning parameters. In the kth iteration, along with $\theta^{(k)}$, λ is also updated based on the general cross-validation (GCV) criterion as follows: $\lambda^{(k)} = \arg\min_\lambda \text{GCV}(\lambda) = \arg\min_\lambda \frac{\|Y - H_1(\lambda)(Y - a^{(k-1)})H_2(\lambda) - a^{(k-1)}\|^2/n}{(1-n^{-1}\text{tr}(H^{(k)}(\lambda))^2}$ (Eilers and Marx 1996), where $a^{(k)} = B_{a,1}X^{(k-1)}B_{a,2}^T$. As $H_i(\lambda)$ that involves matrix inversion should be computed for different values of λ in each iteration, we use a series of transformations and operations inspired by Ruppert (2002) to increase the computational speed. We, first, calculate the Cholesky decomposition of $B_i^TB_i$ that gives the square matrix Z_i. Then, using the eigenvalue decomposition, we calculate the eigenvalues and eigenvectors of matrix $Z_i^{-1}D_i^TD_i(Z_i^{-1})^T$. That is, $U_i\text{diag(s}_i)U_i^T = Z_i^{-1}D_i^TD_i(Z_i^{-1})^T$. Next, the calculated eigenvectors are used to define matrix $V_i = B_i(Z_i^{-1})^TU_i$, which is calculated prior to optimization. Therefore, in each iteration, $H_i(\lambda)$ can be computed by $H_i(\lambda) = V_i^T\text{diag}(\frac{1}{1+\lambda s_1}, \ldots, \frac{1}{1+\lambda s_n})V_i$. As can be seen, the calculation of $H_i(\lambda)$ does not involve matrix inversion, which makes its computation much more efficient. The detailed derivation is shown in Appendix G in the online supplementary materials.

To select the tuning parameter $\gamma^{(k)}$ in the kth iteration, the GCV criterion can still be used. However, GCV usually tends to select more pixels, leading to a larger false positive rate. This is because GCV is a function of the residual sum of square (RSS) in the following way: $\text{GCV} = \frac{\text{RSS}/n}{(1-n^{-1}\text{Tr}(H))^2}$. However, in the case of anomaly detection, the goal is to precisely identify the anomalous regions rather than to achieve a smaller RSS. Therefore, we use the Otsu' s method (Otsu 1975) for finding $\gamma^{(k)}$. Otsu shows that minimizing the intra-class variance $\sigma_w^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t)$ is the same as maximizing interclass variance $\sigma_b^2(t) = \sigma^2 - \sigma_w^2(t) = \omega_1(t)\omega_2(t)\left[\mu_1(t) - \mu_2(t)\right]^2$, where weights $\omega_i(t)$ are the probabilities of the two classes separated by a threshold t and σ_i^2 are the variances of these classes. In our case, classes are defined as the background and the anomalous regions. Both $\omega_i(t)$ and $\mu_i(t)$ can easily be computed from the histogram $p(j)$ by $\omega_1(t) = \sum_{j=0}^t p(j), \omega_2(t) = 1 - \omega_1(t), \mu_1(t) = [\sum_{j=0}^t p(j)x(j)]/\omega_1$, and $\mu_2(t) = [\sum_{j=t+1}^n p(j)x(j)]/\omega_2$. For simplicity, we use otsu(y) to represent the function that returns $\gamma^{(k)}$ by applying Otsu's method on y. For detailed information on Otsu's method, see Otsu (1975).

The detailed optimization algorithm with parameter selection is shown in Algorithm 3.

Algorithm 3: Optimization algorithm for solving SSD based on APG and BCD with tuning parameter selection

initialize

Choose the basis for background as B_1, B_2 and anomalies as $B_{a,1}$, $B_{a,2}$, respectively

$Z_iZ_i^T = B_i^TB_i$, $U_i\text{diag(s}_i)U_i^T = Z_i^{-1}D_i^TD_i(Z_i^{-1})^T$,

$V_i = B_i(Z_i^{-1})^TU_i, i = 1, 2$

$H_i(\lambda) = V_i^T(I + \lambda\text{diag}(s_i))^{-1}V_i, i = 1, 2$ is a function of λ

$L = 2\|B_{a,1}\|_2^2\|B_{a,2}\|_2^2$,

$\Theta_a^{(0)} = 0$, $X^{(0)} = 0$, $t_1 = 1$, $k = 1$

end

while $\left\|\Theta_a^{(k-1)} - \Theta_a^{(k)}\right\| > \epsilon$ **do**

Select $\lambda^{(k)} = \text{argmin}_\lambda\, GCV(\lambda) = \arg\min_\lambda \frac{\|Y - H_1(\lambda)(Y - A^{(k-1)})H_2(\lambda) - A^{(k-1)}\|^2/n}{(1-n^{-1}\text{tr}(\hat{H}(\lambda))^2}$

$H_i^{(k)} = H_i(\lambda^{(k)}), i = 1, 2$

$A^{(k)} = B_{a,1}X^{(k-1)}B_{a,2}^T$, $M^{(k)} = H_1^{(k)}(Y - A^{(k-1)})H_2^{(k)}$

$\Theta_e^{(k)} := X^{(k-1)} + \frac{2}{L}B_{a,1}^T(Y - M^{(k)} - A^{(k)})B_{a,2}$

(In BCD algorithm, for orthogonal $B_{a,i}, i = 1, 2$, $\Theta_e^{(k)} := B_{a,1}^T(Y - M^{(k)})B_{a,2}$)Select $\gamma^{(k)}$ by Otsu's method $\gamma^{(k)} = \text{otsu}(\Theta_e^{(k)}) \times L$

$\Theta_a^{(k)} = S_{\frac{\gamma}{L}}(\Theta_e^{(k)})$

Update $t_k = \frac{1+\sqrt{1+4t_{k-1}^2}}{2}$

Update $X^{(k)} = \Theta_a^{(k-1)} + \frac{t_{k-1}-1}{t_k}(\Theta_a^{(k-1)} - \Theta_a^{(k-2)})$

Update $k = k + 1$

end

3.4 Choice of Basis for Background and Anomalous Regions

Another important factor in the implementation of SSD is the type of basis chosen for the background and anomalous regions. In this section, we provide some general guidelines for basis selection. As it is assumed that the background is smooth, any smooth basis, such as splines or kernels, can be used for the background. If a spline basis is used, the number of knots is another parameter that should be chosen. As pointed out by Ruppert (2002), as long as the number of knots is sufficiently large to capture the variation of the background, a further increase in knots will have little effect due to the regulation of smoothness via λ. Ruppert (2002) proposed using the GCV criterion to select the right number of knots. A similar approach can be used to select the number of knots for the background, assuming that anomalies are only a small part of the entire image.

Selecting the basis to better represent anomalous regions is a more challenging task, and prior information about the size and shape of anomalies would be useful for choosing a suitable basis. For example, if anomalies are small regions scattered over the background or are in the form of thin lines, then it is recommended to use an identity basis, that is, $B_a = I$. However, if the anomalies form clustered regions, a spline basis can be a better choice. For example, if anomalous regions have sharp corners (e.g., a rectangular shape), a 0-order or linear B-splines will suffice. For regions with curved boundaries, quadratic or cubic B-spline bases are recommended. Prior information about the size of anomalies, if available, could help in the selection of the right number of knots for a spline basis. In general, a smaller number of knots may result in the loss of detection accuracy, and a larger number of knots will lead to the selection of normal regions with large noises.

4. SIMULATION STUDY

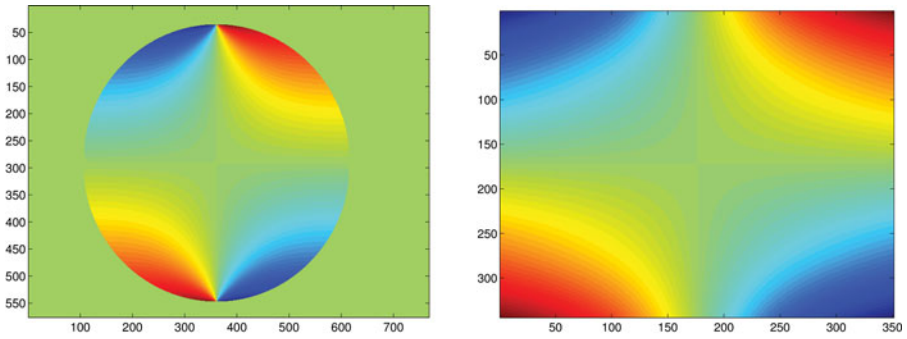
In this section, the performance of the proposed SSD method is evaluated through simulations under different conditions. Specifically, we consider three different types of anomalies: the scattered type, in which anomalies are some random pixels in the image; the line type, in which anomalies are represented by several thin lines; and the clustered type, in which the clusters of anomalies are spread over the image. A sample of these generated images with different types of anomalies is shown in Figure 5. We simulate a 350×350 image Y according to the following model $Y = M + A + E$, in which M is the true background, A represents anomalous regions, and E is the random noise. The smooth background M (see Figure 4(b)) is obtained from the photoelasticity experiments from the center part of the theoretical stress direction of a loaded circle where the load is applied to its top and bottom points (see Figure 4(a)). The anomalies are generated by $A = \delta \cdot I(a \in A_s)$, in which A_s is the set of anomalous pixels, and δ characterizes the intensity difference between anomalies and the background, which is set to be 0.3. For the scattered case, A_s is defined based on randomly generated 5×5 squares. For the line case, a set of line-shaped systematic anomalies caused by numerical errors in photo-elasticity experiments is used. For the clustered case, anomalies are randomly generated clusters comprising about 341 pixels. We generate the random noise E by $E_i \sim \text{NID}(0, \sigma^2)$ with $\sigma = 0.05$.

We compare our proposed SSD method with four existing methods in the literature that follow the two-step approaches described in the Introduction. The benchmark algorithms that we used for comparison include Sobel edge detection (Sobel and Feldman 1968), jump regression with local polynomial kernel regression (Sun and Qiu 2007), the Otsu global thresholding method (Otsu 1975), and the Nick local thresholding method (Khurshid et al. 2009). As per reviewer's suggestion, we also compare our method with the extended-maxima transformation method. The comparison results are reported in the online supplementary materials. For Sobel edge detection, Otsu global thresholding method, and Nick local thresholding method, a smoothing Garcia (2010) is first applied as preprocessing step to remove the noises. For the edge-based methods, namely, the Sobel edge detection and jump regression model, edge-thinning algorithm, edge-linking algorithms, and filling algorithms (Soille 2013) are applied to close the boundary and fill the area inside the boundary.

For the proposed SSD method, we use the identity matrix, cubic B-spline basis with 175×175 knots, and cubic B-spline basis with 85×85 knots in the case of line anomalies, scattered anomalies, and clustered anomalies for anomaly basis B_a, respectively. The cubic B-spline basis B with 7×7 knots is chosen for the background. The proposed methodology is applied for different anomaly cases and the estimated background and anomaly are shown in Figure 6. The tuning parameters λ and γ are selected automatically based on the GCV criterion and Otsu's method as discussed in Section 3.3.

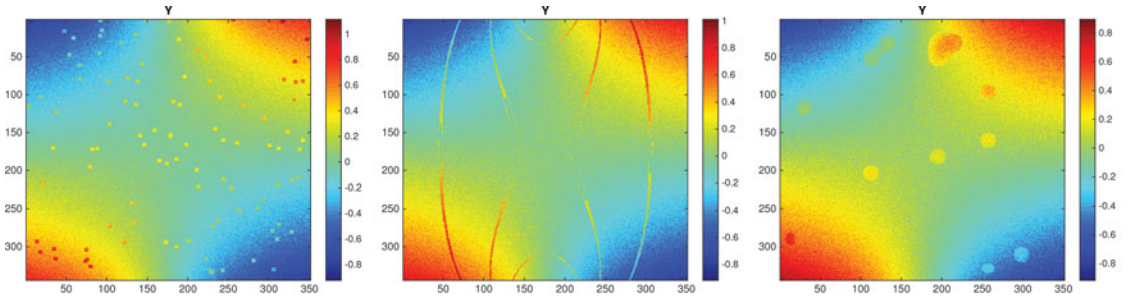
To evaluate the performance of the proposed methodology and benchmark methods, we repeat the simulation procedure 100 times and the following criteria are calculated and compared: false positive rate (FPR), defined as the proportion of normal pixels predicted as anomaly; false negative rate (FNR), defined as the proportion of anomalous pixels predicted as normal; background recovery square root mean square error (e_μ), defined as the square root of mean square error of the background estimator $\hat{\mu}$: $e_\mu = \sqrt{\|\mu - \hat{\mu}\|^2}$; anomalies recovery square root mean square error (e_a), defined as the square root of mean square error of the anomalies estimator $\hat{a}$: $e_a = \sqrt{\|a - \hat{a}\|^2}$; and the computation time. The FPR, FNR, and computation time of all methods for all scenarios are reported in Table 1. Since other benchmark method cannot give the estimation of the background and anomalies, only e_μ and e_a of the proposed SSD is reported in Table 2. Detected regions along with true anomalies for one of the simulation replications with $\delta = 0.3$ is shown in the binary plots in Figure 7.

From Figure 7 and Table 1, it can be seen that in terms of the FPR and FNR, our SSD method overall exhibits a better performance than other benchmark methods. For example, in the line case, the FPR and FNR of SSD are 0.001 and 0.003, respectively, which are the least among all methods. The edge-detection method has a lower FPR than SSD in the clustered case and scattered case. However, it shows an FNR of 0.754 in the case of clustered anomalies, which is much larger than that of SSD, 0.001. The reason for such a high FNR is that the edge-based method only detects the boundaries and often fails to give closed-boundary regions even with edge linking and filling



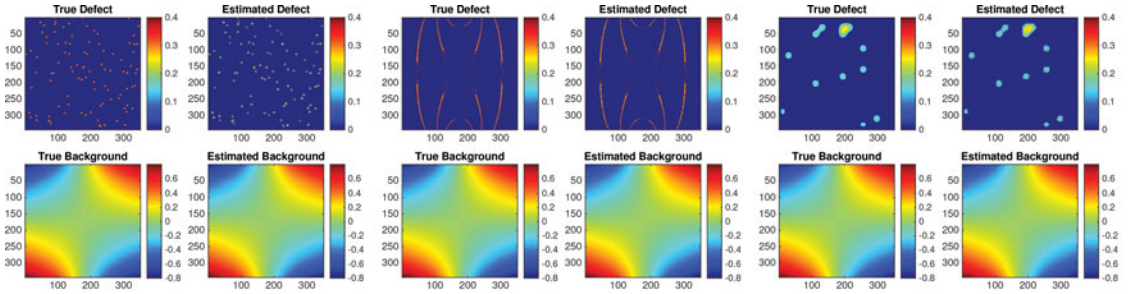
(a) Theoretical principal stress direction of a loaded circle (b) Background

Figure 4. Simulated background from principal stress direction of a loaded circle.



(a) The simulated image of scattered anomaly (b) The simulated image of line anomaly (c) The simulated image of clustered anomaly

Figure 5. The simulated image of scattered and clustered anomalies.



(a) Result of the proposed method in scattered anomaly (b) Result of the proposed method in line anomaly (c) Result of the proposed method in clustered anomaly

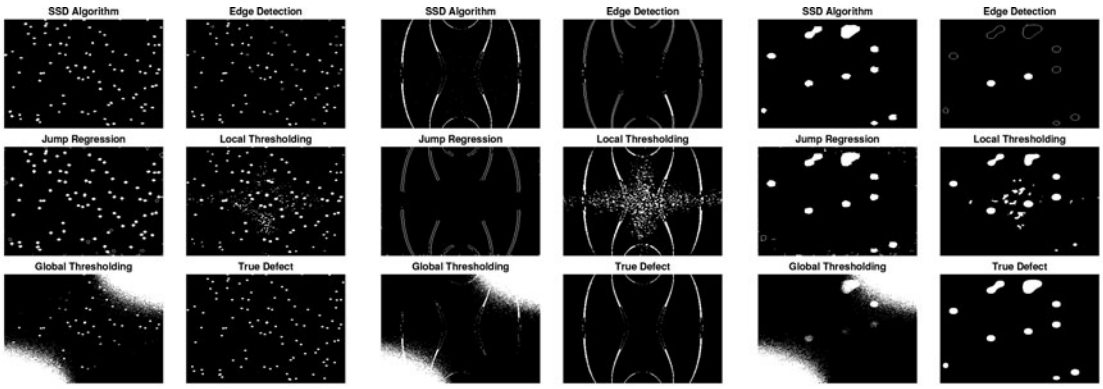
Figure 6. SSD decomposition results for scattered, line, and clustered anomalies.

Table 1. FPR, FNR, and computation time for line anomalies, clustered anomalies, and scattered anomalies with $\delta = 0.3$

	Line anomalies			Clustered anomalies			Scattered anomalies		
	FPR	FNR	Time (sec)	FPR	FNR	Time (sec)	FPR	FNR	Time (sec)
SS decomposition	0.001	0.003	0.129	0.018	0.001	0.19	0.012	0.007	0.267
Edge detection	0.015	0.783	0.945	0.001	0.754	0.409	0.003	0.257	0.667
Jump regression	0.035	0.111	38.43	0.081	0.054	37.736	0.11	0.063	37.796
Local thresholding	0.054	0.063	0.043	0.046	0.289	0.045	0.02	0.087	0.045
Global thresholding	0.195	0.456	0.046	0.211	0.572	0.048	0.203	0.407	0.048

Table 2. The square root of mean square error of background and anomalies estimator with $\delta = 0.3$

	Line anomalies		Clustered anomalies		Scattered anomalies	
	e_μ	e_a	e_μ	e_a	e_μ	e_a
SS decomposition	$3.7e-4$	$6e-4$	$3.6e-4$	$1.5e-3$	$4.2e-4$	$7e-4$

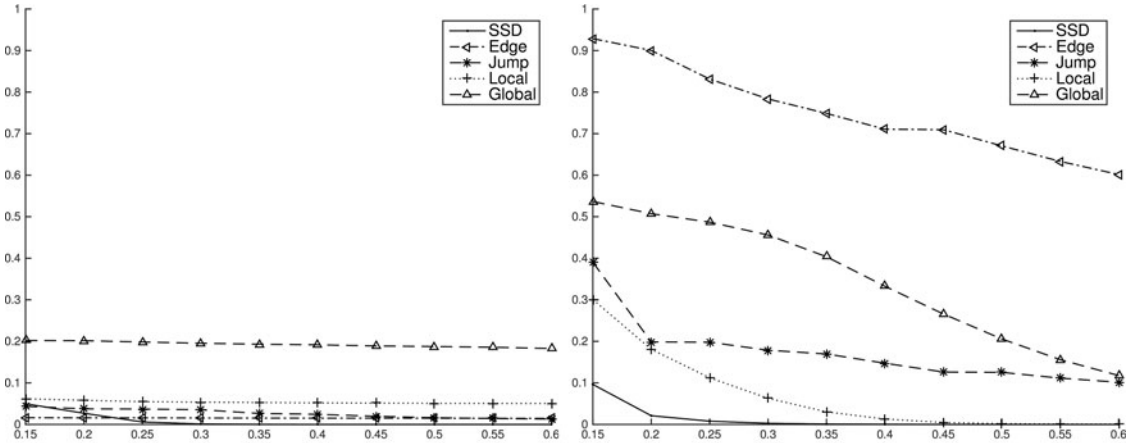


(a) Result of benchmark method in scattered anomaly (b) Result of benchmark method in line anomaly (c) Result of benchmark method in clusterd anomaly

Figure 7. Detected anomalies for scattered, line, and clustered cases.

algorithm applied. Therefore, the inner sections of anomalous regions are not detected as seen in Figure 7(b) and 7(c). Jump regression, on the other hand, can detect the edge more precisely by the local kernel polynomial regression, and thus reduce the FNR dramatically. Global thresholding has the worst performance with FPR and FNR values around 0.20 and 0.50 in all cases. The poor performance of this method is because it can only identify the anomalies globally with one single thresholding value. In contrast, local thresholding has better FPR and FNR values than global thresholding techinques, since it identifies the anomalous regions directly by thresholding the image locally. However, this method is sensitive to noises, and hence falsely detect more anomalies in normal regions than the SSD method. For example, in the clustered case, the FPR of the local thresholding method is 0.289, which is almost 300 times larger than that of SSD, 0.001. In terms of computation time, local and global thresholding methods as well as SSD have comparable computational times. The edge detection is slow because it entails post-processing steps. Jump regression has the highest computation time as it requires fitting local polynomial kernel regression models for each pixel.

To study the sensitivity of these algorithms, we also run a similar simulation setting for different anomaly magnitudes δ. FPR and FNR values of all methods for all cases are reported in Figures 8–10. In terms of FNR, one can see that as δ increases, the FNR of the SSD methods converges to 0 much faster than the rest. In addition, local thresholding and jump regression have much smaller FNR value than other benchmark methods. This indicates that if the intensity contrast between anomalous regions and the background is large, both of these methods will accurately detect all the regions. Jump regression has smaller performance (smaller FNR) especially in the clustered case indicating that it identifies the jump location precisely between the clustered defect and the background. However, in the case of line defect, due to the difficulty to close the boundary, it has larger FNR rate than local thresholding. Both global thresholding and Sobel edge detection have very large FNR rate even when δ is large. In terms of FPR, both local thresholding and jump regression have larger FPR than SSD methods implying that it selects some noisy pixels as anomaly. Global thresholding has much larger FPR. Sobel edge detection sometimes has a very small FPR at the cost of a very large FNR. Furthermore, the comparison of the FNR and FPR plots for different methods indicates that our comparative observations in the previous simulation with $\delta = 0.3$ is valid for other δ values. In short, considering the reported error rates and computation time, the



(a) Result of FPR for different δ (b) Result of FNR for different δ

Figure 8. Sensitivity study for line anomaly.

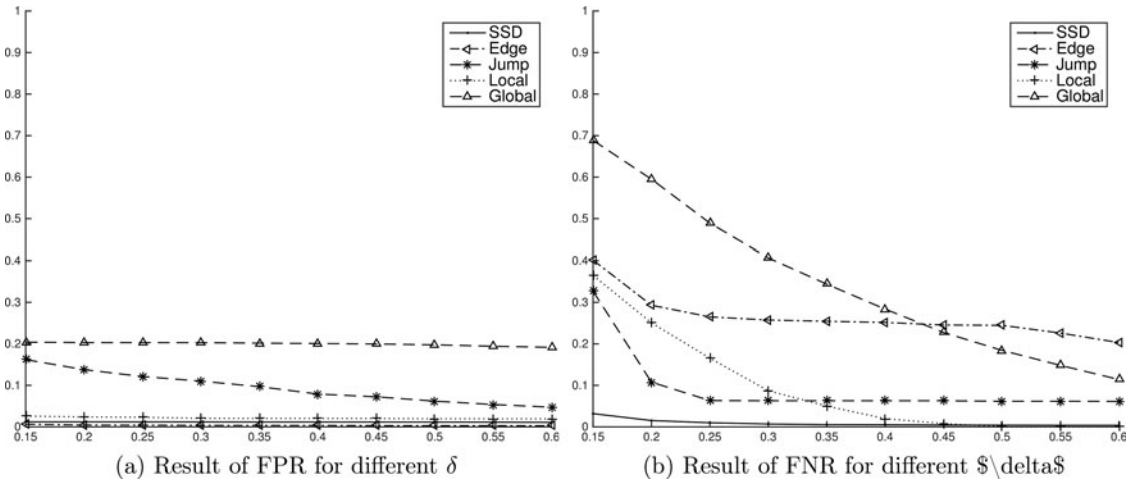


Figure 9. Sensitivity study for scattered anomaly.

proposed SSD method overall outperforms other benchmark methods.

5. CASE STUDY

In this section, the proposed SSD method is applied to a case study of anomaly detection and feature extraction in the area of photo-elasticity, a nondestructive evaluation method. Photo-elasticity is a noncontact optical technique developed based on the birefringence property exhibited by translucent materials. This method can acquire maximum shear stress (isochromatics) and principal stress direction (isoclinics) by recording the transmitted light through translucent materials. Unlike point-by-point techniques such as X-ray diffraction, ultrasonic test, and micro-Raman spectroscopy, photo-elasticity is able to obtain the full-field stress quantification directly from the digital camera, making it popular in the stress analysis of various manufacturing products, such as silicon wafers and translucent composite laminates.

The setup of photo-elasticity experiments used in this article is shown in Figure 11. In this setup, a near infra-red light source of wavelength 1150 nm and a digital camera equipped with a low-pass filter was used to record necessary images. Two

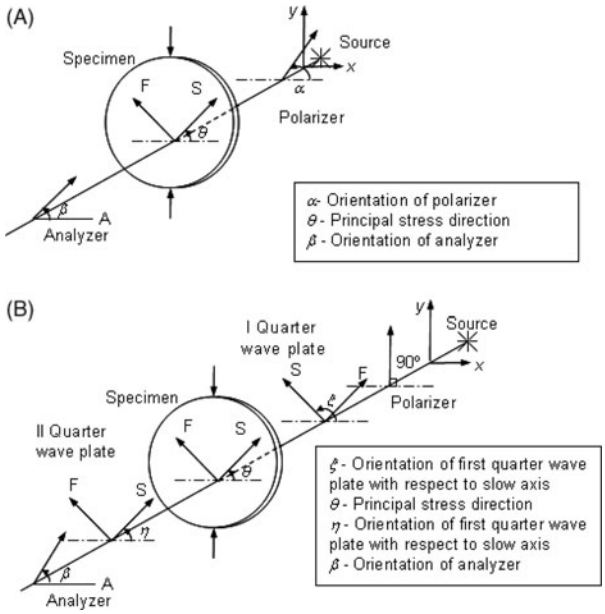


Figure 11. Photo-elasticity experiment setup (Ramji and Prasath 2011).

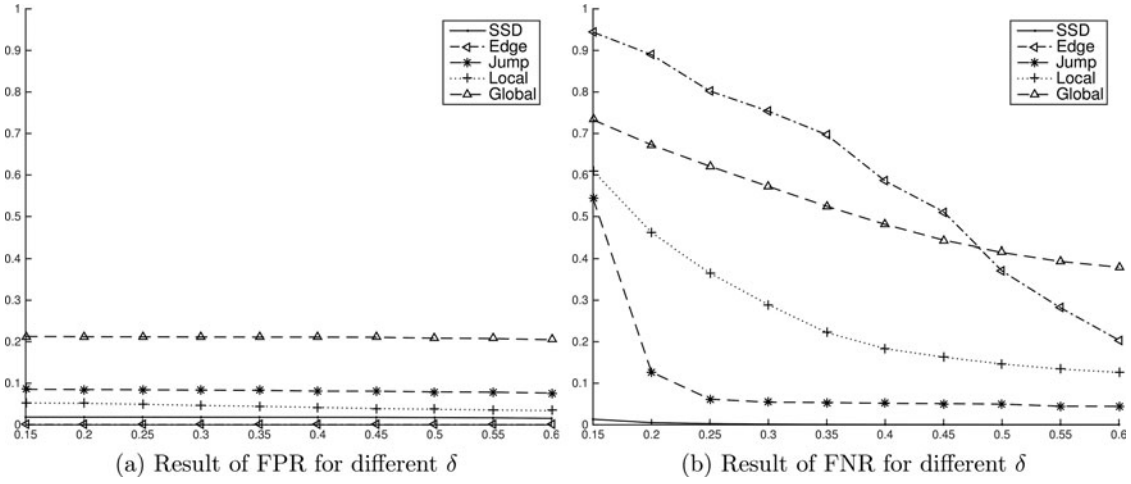


Figure 10. Sensitivity study for clustered anomaly.

Table 3. Case study sample description

Sample	Description	Image size	Defect	Defect type
Sample 1	Multicrystalline silicon	200 by 2910	Grain boundaries	Line
Sample 2	Silicon surface	90 by 550	Surface indentation	Clustered

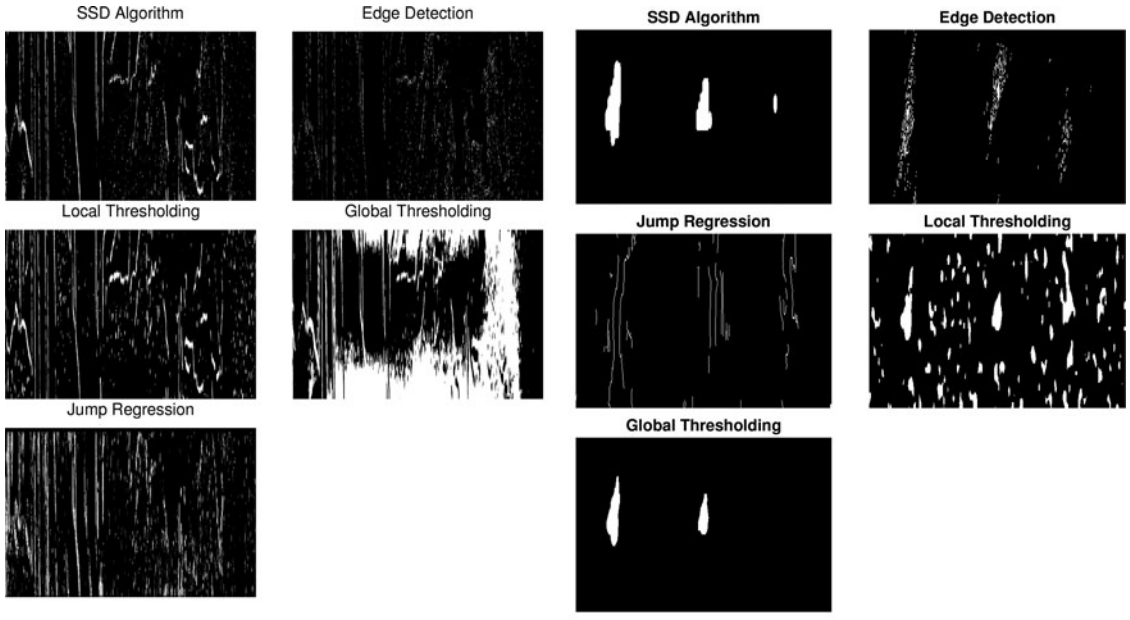
Table 4. Computational time for all methods

Sample	SSD	Edge detection	Local thresholding	Global thresholding	Jump regression
Sample 1	0.350 sec	6.44 sec	0.320 sec	0.270 sec	534.05 sec
Sample 2	0.034 sec	0.543 sec	0.030 sec	0.020 sec	50.069 sec

polarizers are placed on both sides of the specimen, one between the source and the specimen and the other between the specimen and the analyzer. Four images are taken based on the different angles of the two polarizers. Two quarter plates are then added to both sides of the specimen between the specimen and the polarizers. After that, six images are taken by changing the angles of two polarizers and quarter-wave plates. Finally, using Maxwell's stress optic law, the maximum shear stress is obtained from these 10 images to quantify the stress distribution in the specimen (Prasath, Skenes, and Danyluk 2013).

A multicrystalline bending silicon beam and a silicon surface laminate with surface indentation (Skenes, Prasath, and Danyluk 2013) were inspected using the photo-elasticity experiments as described above. The stress maps of these samples are shown in Figure 2(a) and 2(b). In the silicon beam, we are interested in extracting the "grain boundaries." Grain boundaries used in crystallography represent the interface between two crystallines, which is often in the form of swerving lines. In the silicon surface sample, high-stress areas indicate the surface indentation important to be detected in quality inspection. These indentations often form clusters of high-stress areas. A summary of the sample specifications is given in Table 3.

We applied the proposed SSD method and other benchmarks on these stress maps to separate the anomalous regions (i.e., grain boundaries and indentations) from the background of stress maps. In SSD, we used an identity basis for detecting grain boundaries and a cubic B-spline basis with 24×139 knots for detecting the indentations. We also used a cubic B-spline basis with knots 21×21 and 11×51 for the backgrounds of these two samples, respectively. Detected regions from each sample are shown in Figure 12(a) and 12(b). As can be seen in Figure 12(a), the SSD algorithm gives a clearer representation of the grain boundaries than other benchmarks. The edge-detection method is sensitive to noise and hence misses a significant portion of the grain boundaries. Global thresholding performs poorly for Sample 1, since the range of the background intensity in this sample is large. Detected boundaries by local thresholding and jump regression resemble those by SSD. From Figure 12(b), it can be seen that the SSD algorithm can identify all three indentation clusters with no false detection. As the background has a smaller intensity range, the detection performance of global thresholding is better than its performance in Sample 1. However, it still fails to detect the small indentation region because it thresholds the entire image with a single thresholding value without considering the local smoothness. Local-thresholding and edge-detection methods are sensitive to noise, which leads to larger detection errors. Although jump regression can detect all three indentations, its FPR is high. The computation times for each of these methods are also reported in Table 4. Threshold-based methods and SSD have comparable



(a) Detected anomalies in Sample 1 (b) Detected anomalies in Sample 2

Figure 12. Detection results of Samples 1 and 2 for all methods.

computation times, whereas edge detection and jump regression are significantly slower. For example, the computation time of SSD in Sample 2 is around 0.034 sec, 16 times faster than the edge-detection method. The difference in the computation times between Samples 1 and 2 is because of the image size.

6. CONCLUSION

Image data are increasingly used for online inspection and anomaly detection in various manufacturing and service applications. In this article, we proposed SSD for image denoising and anomaly detection. Unlike existing methods, which perform denoising and detection separately, SSD is a one-step approach that is able to model and separate the background, anomalies, and defect. This method improves both the detection accuracy and computation time under the smooth background with various defect types. We formulated the SSD problem in the form of high-dimensional regression augmented with penalty terms to encourage both smoothness of background and sparsity of anomalies in a certain basis. To efficiently solve the large-scale optimization problem for SSD, we used BCD and APG methods and proposed efficient iterative algorithms that have closed-form solutions in each iteration. We also proposed an iterative method for the quick selection of tuning parameters. Using simulations, the performance of proposed SSD was evaluated and compared with some existing methods in terms of detection accuracy and computation time for three types of anomalies. Based on simulation results, we concluded that, overall, the proposed SSD algorithm outperforms other benchmarks. We further showed that the error rate of the proposed SSD algorithm converges to 0 as the anomaly intensity increases. Other benchmarks did not show this property. Additionally, to demonstrate how the proposed method can be applied to real data, we analyzed the stress maps obtained from the photo-elasticity experiments by using SSD. In the case study, we used the stress images of a silicon beam and surface sample that possessed different types of anomalies. We showed that the SSD algorithm can identify anomalous regions precisely in both samples.

The main focus of this article is to provide a framework for one-step, real-time, and automatic algorithms for anomaly detection under smooth background. One extension is to generalize SSD for other types of backgrounds like textured backgrounds. To achieve this, one may use other types of basis such as Fourier basis, wavelet basis, and kernel in the SSD model. Another extension is to propose a data-adaptive method to learn the "best" basis to model the anomaly components. Furthermore, the extension of the proposed SSD algorithm to a dynamic setting where a sequence of time-indexed images is analyzed for process monitoring and anomaly detection is an important, yet challenging problem that requires further research.

SUPPLEMENTARY MATERIALS

Appendix: Proofs of all propositions are given in the online supplementary materials.

Source code: The zipped package that contains MATLAB codes and the simulated images are included in the online supplementary materials. (As it is not authorized to share the dataset used in the case study, it cannot be made publicly available).

ACKNOWLEDGMENTS

This work was partially funded by the Strategic University Partnership between Boeing and Georgia Tech. This work was also partially supported by National Science Foundation Grants CMMI-1451088 and CMMI-1362529. We thank Steve Danyluk and Rajaguruprasath Raveendran for providing us the dataset. We also thank the editor, associate editor, and two anonymous referees for their constructive comments and suggestions that have considerably improved the article.

[Received December 2014. Revised September 2015.]

REFERENCES

Adams, R., and Bischof, L. (1994), "Seeded Region Growing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16, 641–647. [104]

Balageas, D., Fritzen, C.-P., and Güemes, A. (eds.) (2010), *Structural Health Monitoring* (Vol. 90), London: ISTE. [102]

Bradley, D., and Roth, G. (2007), "Adaptive Thresholding Using the Integral Image," *Journal of Graphics, GPU, and Game Tools*, 12, 13–21. [104]

Chan, C.-H., and Pang, G. K. (2000), "Fabric Defect Detection by Fourier Analysis," *IEEE Transactions on Industry Applications*, 36, 1267–1276. [103]

Comaniciu, D., Ramesh, V., and Meer, P. (2003), "Kernel-Based Object Tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25, 564–577. [104]

De Boor, C. (1978), *A Practical Guide to Splines* (Vol. 27), New York: Springer-Verlag. [102,103]

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004), "Least Angle Regression," *The Annals of Statistics*, 32, 407–499. [106]

Eilers, P. H., and Marx, B. D. (1996), "Flexible Smoothing With B-Splines and Penalties," *Statistical Science*, 11, 89–102. [103,107]

Epifanio, I. (2008), "Shape Descriptors for Classification of Functional Data," *Technometrics*, 50, 284–294. [104]

Friedman, J., Hastie, T., Höfling, H., and Tibshirani, R. (2007), "Pathwise Coordinate Optimization," *The Annals of Applied Statistics*, 1, 302–332. [105]

Funck, J., Zhong, Y., Butler, D., Brunner, C., and Forrer, J. (2003), "Image Segmentation Algorithms Applied to Wood Defect Detection," *Computers and Electronics in Agriculture*, 41, 157–179. [103]

Garcia, D. (2010), "Robust Smoothing of Gridded Data in One and Higher Dimensions With Missing Values," *Computational Statistics & Data Analysis*, 54, 1167–1178. [108]

Jiang, B., Wang, C.-C., and Liu, H.-C. (2005), "Liquid Crystal Display Surface Uniformity Defect Inspection Using Analysis of Variance and Exponentially Weighted Moving Average Techniques," *International Journal of Production Research*, 43, 67–80. [102,103]

Jiang, X., and Mojon, D. (2003), "Adaptive Local Thresholding by Verification-Based Multithreshold Probing With Application to Vessel Detection in Retinal Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25, 131–137. [104]

Jin, N., Zhou, S., and Chang, T.-S. (2004), "Identification of Impacting Factors of Surface Defects in Hot Rolling Processes Using Multi-Level Regression Analysis," *Transactions of the NAMRI/SME*, 32, 557–564. [102]

Khurshid, K., Siddiqi, I., Faure, C., and Vincent, N. (2009), "Comparison of Niblack Inspired Binarization Methods for Ancient Documents," in *IS&T/SPIE Electronic Imaging Proceedings*, International Society for Optics and Photonics, pp. 72470U–72470U. [108]

Kumar, A. (2008), "Computer-Vision-Based Fabric Defect Detection: A Survey," *IEEE Transactions on Industrial Electronics*, 55, 348–363. [102]

Kumar, A., and Pang, G. K. (2002), "Defect Detection in Textured Materials Using Gabor Filters," *IEEE Transactions on Industry Applications*, 38, 425–440. [103]

Lu, C.-J., and Tsai, D.-M. (2004), "Defect Inspection of Patterned Thin Film Transistor-Liquid Crystal Display Panels Using a Fast Sub-Image-Based Singular Value Decomposition," *International Journal of Production Research*, 42, 4331–4351. [103]

Marr, D., and Hildreth, E. (1980), "Theory of Edge Detection," *Proceedings of the Royal Society of London*, Series B, 207, 187–217. [103]

Milanfar, P. (2013), "A Tour of Modern Image Filtering: New Insights and Methods, Both Practical and Theoretical," *IEEE Signal Processing Magazine*, 30, 106–128. [103]

Nagao, M., and Matsuyama, T. (1979), "Edge Preserving Smoothing," *Computer Graphics and Image Processing*, 9, 394–407. [102]

Nesterov, Y. (1983), "A Method of Solving a Convex Programming Problem With Convergence Rate O (1/k2)," *Soviet Mathematics Doklady*, 27, 372–376. [106]

Nesterov, Y., Nemirovskii, A. S., and Ye, Y. (1994), *Interior-Point Polynomial Algorithms in Convex Programming* (Vol. 13), Philadelphia, PA: SIAM. [105]

Otsu, N. (1975), "A Threshold Selection Method From Gray-Level Histograms," *Automatica*, 11, 23–27. [104,107,108]

Parikh, N., and Boyd, S. (2013), "Proximal Algorithms," *Foundations and Trends in Optimization*, 1, 123–231. [106]

Park, C., Huang, J. Z., Huitink, D., Kundu, S., Mallick, B. K., Liang, H., and Ding, Y. (2012), "A Multistage, Semi-Automated Procedure for Analyzing the Morphology of Nanoparticles," *IIE Transactions*, 44, 507–522. [104]

Prasath, R., Skenes, K., and Danyluk, S. (2013), "Comparison of Phase Shifting Techniques for Measuring In-Plane Residual Stress in Thin, Flat Silicon Wafers," *Journal of Electronic Materials*, 42, 2478–2485. [102,112]

Prewitt, J. M. (1970), "Object Enhancement and Extraction," *Picture Processing and Psychopictorics*, 10, 15–19. [103]

Qiu, P. (2007), "Jump Surface Estimation, Edge Detection, and Image Restoration," *Journal of the American Statistical Association*, 102, 745–756. [102,103]

Qiu, P., and Mukherjee, P. S. (2012), "Edge Structure Preserving 3-D Image Denoising by Local Surface Approximation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34, 1457–1468. [102,103]

Qiu, P., and Sun, J. (2007), "Local Smoothing Image Segmentation for Spotted Microarray Images," *Journal of the American Statistical Association*, 102, 1129–1144. [104]

——— (2009), "Using Conventional Edge Detectors and Postsmoothing for Segmentation of Spotted Microarray Images," *Journal of Computational and Graphical Statistics*, 18, 147–164. [104]

Qiu, P., and Xing, C. (2013), "On Nonparametric Image Registration," *Technometrics*, 55, 174–188. [104]

Qiu, P., and Yandell, B. (1997), "Jump Detection in Regression Surfaces," *Journal of Computational and Graphical Statistics*, 6, 332–354. [103]

Ramji, M., and Prasath, R. (2011), "Sensitivity of Isoclinic Data Using Various Phase Shifting Techniques in Digital Photoelasticity Towards Generalized Error Sources," *Optics and Lasers in Engineering*, 49, 1153–1167. [111]

Ruppert, D. (2002), "Selecting the Number of Knots for Penalized Splines," *Journal of Computational and Graphical Statistics*, 11, 735–757. [107,108]

Sauvola, J., and Pietikäinen, M. (2000), "Adaptive Document Image Binarization," *Pattern Recognition*, 33, 225–236. [104]

Skenes, K., Prasath, R., and Danyluk, S. (2013), "Measurement of Residual Stresses Around Vickers Indentations on Silicon Surfaces via NIR Polariscope," in *Proceedings of the 28th European PV and Solar Energy Conference and Exhibition*, pp. 242–245. [112]

Sobel, I., and Feldman, G. (1968), "A 3x3 Isotropic Gradient Operator for Image Processing," in *Pattern Classification and Scene Analysis*, eds. R. Duda and P. Hart, New York: Wiley, pp. 271–272. [103,108]

Sohn, H., Park, G., Wait, J. R., Limback, N. P., and Farrar, C. R. (2004), "Wavelet-Based Active Sensing for Delamination Detection in Composite Structures," *Smart Materials and Structures*, 13, 153. [102,103]

Soille, P. (2013), *Morphological Image Analysis: Principles and Applications*, Berlin: Springer Science & Business Media. [104,108]

Sun, J., and Qiu, P. (2007), "Jump Detection in Regression Surfaces Using Both First-Order and Second-Order Derivatives," *Journal of Computational and Graphical Statistics*, 16, 289–311. [108]

Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*, Series B, 58, 267–288. [105,106]

Tseng, P. (2008), "On Accelerated Proximal Gradient Methods for Convex-Concave Optimization," submitted to *SIAM Journal on Optimization*. [106]

Unser, M. (1999), "Splines: A Perfect Fit for Signal and Image Processing," *IEEE Signal Processing Magazine*, 16, 22–38. [103]

Wand, M. P., and Jones, M. C. (1994), *Kernel Smoothing* (Vol. 60), London: CRC Press. [103]

Wink, A. M., and Roerdink, J. B. (2004), "Denoising Functional MR Images: A Comparison of Wavelet Denoising and Gaussian Smoothing," *IEEE Transactions on Medical Imaging*, 23, 374–387. [102]

Wu, Y. N., Li, J., Liu, Z., and Zhu, S.-C. (2007), "Statistical Principles in Image Modeling," *Technometrics*, 49, 249–261. [104]

Xiao, L., Li, Y., and Ruppert, D. (2013), "Fast Bivariate P-Splines: The Sandwich Smoother," *Journal of the Royal Statistical Society*, Series B, 75, 577–599. [107]

Yan, H., Paynabar, K., and Shi, J. (2014), "Image-Based Process Monitoring Using Low-Rank Tensor Decomposition," *IEEE Transactions on Automation Science and Engineering*, 12, 216–227. [104]