Automatic Surface Defect Inspection System Using Convolutional Neural Networks

X Zhu\textsuperscript{1,2*}, S K Kundu\textsuperscript{1,3*}, and N K Bangalore Ramaiah\textsuperscript{1,4}

\textsuperscript{1}Hitachi America, Ltd. Research & Development Division, Farmington Hills, MI, 48335 USA
\textsuperscript{2}xiaoliang.zhu@hal.hitachi.com, \textsuperscript{3}subrata.kundu@hal.hitachi.com,
\textsuperscript{4}naveenkumar.bangaloreramaiah@hal.hitachi.com
*Corresponding authors: xiaoliang.zhu@hal.hitachi.com, subrata.kundu@hal.hitachi.com

Abstract. Surface quality of the piston rod is very important for the durability of the shock absorbers. However, currently in production, it is very difficult to use the available systems to inspect the whole rod surface online, in real time (~4s), with a high sensitivity and accuracy. To overcome this, in this paper, an online automatic rod inspection system has been developed, which allows to inspect the whole rod surface in 4s, with a sizing accuracy around 75\%, and detection accuracy >90\%. An integrated software is used for rod rotation control, image capturing, image processing, and decision making. Convolutional Neural Network is used to processing the 360° surface image with a high accuracy, eliminating errors caused by environmental lighting. A new method based on aspect ratio and size information is used for defects classification. Experimental results show that the system is capable of detecting defects as small as 25µm and differentiating nodule, dent, and scratch with a processing time around 4s per rod.

Keywords: Surface quality; machine vision; CNN; dent; nodule; scratch

1. Introduction

Automotive suspension system, which can be classified as passive, semi-active, and active suspension, plays a more and more important role in driving stability and comfort by reducing or eliminating the vibrations transmitted to vehicle body [1]. As one of the key moving component of the shock absorbers in the suspension system, the piston rod is forced moving reciprocally by the vibrations or bumps from the road, moving hydraulic oil between different chambers in the absorber and converting the vibration energy into heat which is then dissipated. During the piston rod’s reciprocal movement, if there are any defects on the rods surface, namely, nodule, dent, and scratches, the hydraulic oil seal will be quickly damaged, causing oil leakage, noise, and eventually a failure of the suspension system. Thus, it is critical to inspect the rod surface quality before assembling.

The traditional way to inspect the rod surface is performed manually by either visual inspection (eyeballing the rod surfaces) or scratching rod surfaces with a knife. For the latter, by feeling the resistances caused by defects, defective rods can thus be screened. Although simple, those methods not only highly rely on workers experience which is subjective and labor-intensive, but also time consuming and raise risk of surface damage due to the additional manual handling. In an attempt to solve this problem, some non-destructive online sensors have been developed. Ismagilov et al. have developed a method for detecting surface defects in micron-thickness metallic films by scanning the object surface with a focused laser beam. A piezoelectric sensor is then used to detect the generated
acoustic waves (Raleigh wave) on the object surface [2]. Schmitt Europe Ltd. also developed a product, LaserCheck™ Inline surface roughness sensor, for measurement of surface qualities based on laser light scatter detection technology. While high sensitivities have been achieved by these methods, only one line or multiple lines on the rods surface are inspected. Thus, it is highly biased and cannot provide enough information for the whole rod surface quality. On the other hand, eddy current sensors have also been widely developed and implemented for surface quality inspections. Despite its non-destructive, simple structure, and fast inspection time, its accuracy can be easily affected by the distance between the sensor, i.e. sensing coils, and the rod surface, and the subsurface material properties, e.g. porosity, slag, or even different grain size or hardness [3-5]. Thus, complicated sensor adjustments are needed or large errors could occur. Optical methods, including optical microscopy, light scattering, and machine vision system, have more potential and possibilities for surface quality characterization due to their non-destructive and online inspection capabilities. Although many research works have been demonstrated based on machine vision in areas of touch panel inspection [6], aluminum surface roughness measurement [7], and tile/fabric defect detection [8, 9], these technologies cannot be applied for rod inspection due to 1), the small size of the defects, e.g. nodule diameters are in the range of 1 µm to 90 µm; 2), the high reflective, cylindric, and the wide variety of defects on the rods surface, which are caused by various chemical and mechanical processes in the production line; 3) uneven illumination due to environmental light condition and constructive imperfection; and 4) the short time needed for inspecting the whole rod surface online.

In this paper, a unique automatic vision-based rod surface inspection system with defect detection and differentiation algorithm is presented. The system consists of a LED light, a rod manipulation stage, a high-resolution line scan camera, and a laptop with an integrated software for system control and image processing. Convolutional Neural Network (CNN) is used for image processing to eliminate uneven background intensity. After image processing, data of defect size, count, types, and location are extracted and saved on the PC, allowing users for accessing the surface quality (defects types, sizes, and counts). Experimental results show that, the developed system is capable of detecting defects as small as 25µm and has the ability to differentiate nodule, dent, and scratch with a processing time around 4s per rod, which has not been achieved before.

2. System design
2.1. Overall system

The structure of the automatic vision-based rod surface inspection system is shown in figure 1. A high intensity LED light is used for illuminating the rod surface. The robotic manipulation station is motorized with two DC motors to fully automate the inspection; one motor for rod rotation (Motor 1) and the other one for moving the rod along its axial direction (Motor 2). A line scan camera with 4096 pixels and a Field of View (FOV) of 6 cm (around 14.6 µm/pixel), which is oriented parallel to the rod axis, is used to capture the surface image of a rod with 22 mm in diameter and 30 cm in length. It is worth mentioning here that higher resolution camera (e.g. 16K resolution) can be used to capture the whole surface in one revolution. A laptop with a self-developed software is used to coordinate the robotic manipulation stage and the camera for image capturing, and control the light indicator based on the rod condition: red light means bad rods and green light means good rods. Image processing and defect extraction are also performed by the integrated software developed in C++.

The working mechanism of the inspection system is illustrated in figure 2. When there is no defect on the rods surface, a mostly dark image will be recorded by the camera since most of the light is reflected, as shown in figure 2 (a). However, if there are defects on the surface (i.e. nodules, dents, or scratches), some light will be scattered toward the camera direction. Thus, a bright area will be recorded in a dark background, as shown in figure 2(b). The size and shape of the bright area is related to the defect size and shape. To be noted that the rod rotation speed was determined based on rod diameter, camera line rate, and FOV. The camera position is adjusted by a 3-axis linear stage and a 2-axis goniometer stage to get the best image quality.
2.2. Image processing framework
The main functions of the integrated software are shown in figure 3. First, the system will initialize for the camera (e.g. line rate, exposure time, frame size, capture time, etc.) and the two motors (e.g. rotation speed, acceleration/deceleration speed, etc.). After that, if the light sensor (in the rod shadow, not shown in figure 1) detected rod presence, Motor 1, which is controlled by the developed software, will start to rotate at the initialized speed. Second, camera will start to record the 360º surface images with the initialized parameters. It is worth to mention here that the line scan camera is calibrated with methods described in [10] for distortion removal. Third, the recorded images (18 frames) are transferred to the allocated PC RAM for processing to get the defect type, size, and count information. To be noted that the FOV is 6 cm in the current design, thus, a total of 5 sections are needed to get the 360º surface images for a 30cm long rod. Motor 2 is used to move the rod from section to section. Finally, the total defect number on the rods surface will be compared with a pre-set threshold. If any defect size overpasses the threshold, the red light will be turned on. If none, the green light will be turned on.

2.3. Convolutional neural networks
Classical image processing is effective when captured image frames have a constant background intensity. When the captured image has varied background intensity, as shown in figure 4(b), and figure 9, binary thresholding becomes unreliable. To overcome this, CNN which has been wide used for speech analysis, object detection and categorization in recent years [11,12], is used to pre-process the captured images. To use CNN, basically, training data sets including input image and label image are needed to train a network/architecture, if the error rate of the output image is less than a threshold, the training is done, and the final architecture is then used in figure 4 (a). Figure 5 gives an overview of the underlying parts of the neural network development. The starting point is the generation of a dataset. After capturing 342 images with a size of 4096 × 262 from different rods, for each image a corresponding label image need to be created. The label images represent the information that the neural network needs to learn, i.e. the locations and sizes of defects. For the neural network, an architecture including types and number of layers as well as the structure of their interconnectivities
needs to be defined. Then, the dataset – consisting of both captured images and label images – are put into the network during training. In every training step, the output of the network is compared with the label image and the resulting error is projected backwards through the network.

After that, the weights of the network are adjusted in a way, which reduces the overall error. When training is completed, the trained weights can be used together with the defined architecture to deploy the network in an application, as shown in figure 4(a). Furthermore, the accuracy and performance of the network can be evaluated with a set of test images to see whether the results are satisfying.

Since a lot of defects appear really small on the image, having only a few pixels, the network should make for each pixel a decision if it belongs to a defect or not. For that reason, a binary label image is generated, on which every defect pixel is white and every background pixel is black. How this is done, is shown in the red rectangle in figure 5. As shown in the right part of the red rectangle, the captured image is binary threshold to mark the bright defect clearly. In the left part of the red rectangle, sharp operation is used to increase its contrast before binary threshold. With this step, less bright defects can be found on the resulting binary image with the cost of a very noisy detection of the bright defects. However, the two techniques can be combined to get both advantages of clearly found bright defects and the detection of less bright defects by simply adding the two binary images. Then region of interest (ROI) is used to crop the label image for CNN output training. For the CNN input training data, same ROI on the captured image is cropped then varied its brightness multiplying a random factor between 0.25 and 4, while the corresponding label images remain the same. With this action, the problem of having a dataset with label images, which highly depend on the brightness of their underlying source images, is resolved. Thus, the network can learn to not only rely on brightness, but also on shape and the appearance of the defects.

It should be noted that, as long as the image dimensions are significantly bigger than the size of the biggest filter kernel of the network, the quality of the output is nearly the same for different input dimensions. This fact allows us to crop the images into smaller parts and, by doing so, increases the

![Image](image.png)  
**Figure 4.** (a) image processing framework for defect information extraction; (b) illustration of uneven background intensity; (c)-(e) illustration of morphological transform, discard and merge functions.
dataset. Since the cropping of a small image part (128 × 128 pixels) is random, not only the number of resulting images can be increased, but also the amount of information in the images. The final dataset has 102,600 images with 128 × 128 pixels for training and validation, and 180 images with 4,096 × 262 pixels for testing.

To make a pixel-wise prediction using CNN, the output dimensions of the last layer must fit the image input size. Since some of the defects might only have a few pixels, there are no pooling layers in the architecture, because small defects would be lost after down sampling the feature maps. Hence, the size of the feature maps is the same all over the network. In addition to that, there are no fully connected layers neither, to keep the spatial information of the feature maps. All the mentioned constraints lead to an architecture which only consists of convolutional layers followed by ReLU (Rectified Linear Unit) layers as non-linear activation functions.

After generating the dataset and setting up the architecture, training can be started, which is the process of optimizing the weights of the network with respect to a loss function, which calculates the difference between the ground truth represented by the label image and the output image of the network. Since the outputs of our network represent probabilities, it is appropriate to use the Sigmoid-Cross-Entropy-Loss E, which is described in the following formula:

\[
E = -\frac{1}{N} \sum_{n=0}^{N} \left[ p_n \log \hat{p}_n + (1 - p_n) \log(1 - \hat{p}_n) \right].
\]

where \( N \) denotes the number of pixels of the image, \( p_n \) the label value of the \( n^{th} \) pixel \( \{0,1\} \) and \( \hat{p}_n \) the predicted value of the \( n^{th} \) pixel. As optimization algorithm, Stochastic Gradient Descent is used. To determine the best architecture, try and error in terms of both structural parameters (number of filters, kernel size, etc.) as well as hyperparameters (learning rate, batch size, etc.) can be used. For architectures with similar performance (error rate of predication), additional method can be used to quantify the quality of the neural network result, which is average of width and height accuracy (ACC), as shown in figure 6.

![Figure 5. Data preparing framework for CNN training.](image)

![Figure 6. Calculation of defect size accuracy.](image)

Since the final defect size accuracy determined by the bounding boxes after applying the finding contours function, the size of each resulting rectangle of the output image is compared to the corresponding rectangle of the label image. Because nearly no false positives occur ( < 0.1%), the lingering ones do not affect the result. On the other hand, missed detections are considered with a size accuracy of zero for the underlying defect. How the size accuracy is calculated in detail, can be found in figure 6. Since most of the defects have a circular shape, we are interested in the accuracy of the diameter prediction. So not the numbers of pixels within the bounding boxes are compared, but their widths and heights (which are the same for the most cases). Both width and height of the bounding box on the output image is compared with the corresponding dimensions of the bounding box of the label image and their accuracies are averaged. This mean value represents the accuracy of the specific
detected defect. Then, this value of every single defect is again averaged to calculate the size accuracy for the whole image, and based on this, the best architecture is selected. The finally selected architecture has 5 layers, with 8, 16, 32, 64, and 1 filters per layer. The kernel size for each layer is 15, 9, 7, 5, 3, and the ACC is 85.3% for the whole testing images. To be noted that, ACC as high as 85.3% was obtained with extra layers and kernels, with an expense of much longer processing time.

2.4. Defect classification and calibration

After determined, the CNN model is implemented into the image processing, as shown in figure 4. It is worth mentioning that in this step, labels that have two bright pixels or lie within other labels are discarded or merged together, as shown in figure 4 (d) and (e). Finally, the total defect sizes and numbers are obtained. To classify the defects, aspect ratio and size information are used. If the defect aspect ratio is larger than 3, it will be classified as scratch. However, this cannot be applied to classify nodule and dent because of their similar circular shapes (aspect ratio between 1 and 2.5). To overcome this problem, the size of a total of 774 nodule defects and 90 dent defects were measured under a microscope (Keyence, VHX-600). Their size distribution is shown in figure 7 (a), where nodule sizes range from 2 - 90 µm with a Gaussian distribution, and dent size ranges from 70 µm to 300 µm. Thus, a size threshold of 88 µm was used to classify nodules from dents with an error rate around 1.5%.

Furthermore, the relationship of nodule diameter and its height is measured with a laser confocal microscope (OLYMPUS OLS5000) from Lurie Nanofabrication Facility (LNF) at the University of Michigan. As shown in figure 7 (b), this information is used to estimate nodule height from the detected nodule diameter, as the height information is also a very critical information to determine the surface quality. Figure 7(c) shows the relationship between detected defect pixel number and the defect size measured under the laser confocal microscope, which is used as the calibration curve to get the real defect size from detected defect pixel numbers.

3. Experiments and results

![Figure 7](image-url)
Figure 8 shows the setup of the automatic rod inspection system. Two limit switches and one reset switch are implemented for safety purposes. Suspension rods with a diameter of 22 mm were loaded onto the robotic manipulation stage manually. For each section, the image processing and defect classification will be performed after the 360° surface image is captured. The rod condition will then be indicated by the indication light after all the sections are inspected and the defects type, size, and counts are collected. The image processing time for each rod is around 4s. In figure 9, the performance of CNN to overcome the uneven background intensity is compared with conventional binary thresholding method. As can be seen, the CNN effectively removed the uneven background intensity, and correctly detected the defects in the image. Figure 10 shows a sample inspection result. Figure 10 (a) is the image captured from the camera, which is cropped from a 4096 × 262 frame to show the details clearly. Figure 10 (b) shows the processed image with labelling, where green rectangles are the detected nodules, blue rectangles are the detected dents, and red rectangles are the detected scratches. The corresponding microscope images, the captured image, and the labelled image of defects (c), (d), and (e) are also shown in figure 10. The inspection results demonstrated that the current inspection system can effectively detect nodules as small as 25 µm and can differentiate nodules, dents, and scratches with a detection accuracy >90% and sizing accuracy >75%.

Figure 8. Setup of the automatic rod inspection system.  
Figure 9. Comparison of image processing results with binary thresholding and CNN.  
Figure 10. (a) captured raw image of a rod.  
(b) image after processing and labelling.  
(c) microscope image of a scratch with 30 µm width and 780 µm length.  
(d) microscope image of a 30 µm nodule (in red dashed rectangle) and a dent.  
(e) microscope image of a 25 µm nodule (in red dashed rectangle) and two dents.

To be noted that in figure 10 (c), the bottom of the scratch is detected as multiple nodules (green rectangles) as well as some undetected small defects in figure 10 (b). This is because for certain
smaller or shallow defects, their pixel intensity in the captured image are low and inconsistent. After CNN processing, these low intensity pixels are omitted in the output binary image, and thus causing this error in the detection image, which can be overcome with functions of edge drawing algorithm (ED), line segment detection (LSD) or Hough Transform techniques [13].

4. Conclusions
Aiming to overcome the disadvantages of the current inspection method for cylindric objects, in this paper, a unique automatic suspension rod inspection system that can detect and differentiate nodules, dents, and scratches is presented. An integrated software is used for LED light control, rod rotation, image capturing, image processing, and indication lights control. For each rod, a total of 90 frames, each with 4096 × 262 pixels were captured and processed. To be noted that, higher resolution camera e.g. 16K resolution, can be used to replace the current 4K camera to get a higher sensitivity, or capture the whole surface image in one revolution. Experiments results shown that, the system can detect defects as small as 25 µm (75% sizing accuracy, and >90% detection accuracy) with a processing time around 4s per rod, which proved the excellent capability of the developed system.

Acknowledgements
Some measurements were partially performed at the University of Michigan Lurie Nanofabrication Facility.

Reference
[1] Wang W and Song Y 2019 Analytical computation method for steady-state stochastic response of a time-delay nonlinear automotive suspension system. Mech. Syst. Signal Pr., 131, 434-445
[2] Ismagilov I R, Golenishchev-Kutuzov V A, Kalimullin R I, Migachev S A, and Khasanov A A 2014 Detecting surface and volume defects in metals by the laser-acoustic method. Russ. J. Nondestrukt., 50(6), 318-324
[3] Perevertov O, Neslušan M, and Stupakov A 2017 Detection of milled 100Cr6 steel surface by eddy current and incremental permeance methods. NDT & E Int. 87, 15-23
[4] Jiao D, Ni L, Neslušan M, and Stupakov A 2017 Detection of milled 100Cr6 steel surface by eddy current and incremental permeance methods. NDT & E Int. 87, 15-23
[5] Han Y, Zhong C, Zhu X, and Zhe J 2018 Online monitoring of dynamic tip clearance of turbine blades in high temperature environments. Meas. Sci. Technol., 29(4), 045102
[6] Chang M, Chen B C, Gabayno J L, and Chen M F 2016 Development of an optical inspection platform for surface defect detection in touch panel glass. Int. J. Optomechatron., 10(2), 63-72
[7] Ghodrati S, Mohseni M, and Kandi S G 2019 Application of image edge detection methods for precise estimation of the standard surface roughness parameters: polypropylene/ethylene-propylene-diene-monomer blend as a case study. Measurement, 138, 80-90
[8] Xie L, Xiang X, Xu H, Wang L, Lin L, and Yin G 2020 FFCNN: A Deep Neural Network for Surface Defect Detection of Magnetic Tile. IEEE Transactions on Industrial Electronics. DOI: 10.1109/TIE.2020.2982115
[9] Makarem M, Razmjoo N, and Ramezani M 2018 A new method for detecting texture defects based on modified local binary pattern. Signal Image and Video P. 12(7), 1395-1401
[10] Sun B, Zhu J, Yang L, Yang S, and Niu Z 2016 Calibration of line-scan cameras for precision measurement. Appl. optics, 55(25), 6836-6843
[11] Bernal J, Kushibar K, Asfaw D S, Valverde S, Oliver A, Martí R, and Lladó X 2019 Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review. Artifi. Intell. Med. 95, 64-81
[12] Fahad S A, and Yahya A E 2018 July Inflectional review of deep learning on natural language processing. In 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCCEE) (pp. 1-4). IEEE
[13] Liu Y, Xie Z, and Liu H 2019 LB-LSD: A length-based line segment detector for real-time applications. Pattern Recognit. Lett. 128, 247-254.