

Streak Detection in Color Printed Customer Content*

Runzhe Zhang^a, Eric Maggard^b, Yousun Bang^c, Minki Cho^c, Mark Shaw^b, Jan Allebach^a

^a School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47906, U.S.A.

^b HP Inc., Boise, ID 83714, U.S.A.

^c HP Inc., Suwon City, KOREA.

Abstract

Streaks are one of the most common print defects in electrophotographic printers that influence print quality. Streak defects are dark or light lines with the major axis along the printing process direction and are usually caused by a defective Intermediate Transfer Belt (ITB) or Organic Photo Conductor (OPC) component in the printer. Previously, we designed an algorithm to detect the streak defects only on standard printed test pages, which have uniform color [1]. In this paper, we design an algorithm to detect the streak defects on the customer content area, which we call a raster ROI. It is more complicated than the uniform color printed page because the customer content influences our streak detection result. Sometimes, the customer content has some dark or light straight lines along the printing process direction, and they are similar to streak defects. To detect the streak defects on the customer pages, we must separate the straight lines of customer content and the streak defects. To detect the streak in a raster ROI, we apply the Sobel edge detection algorithm and morphological operations to the master image, which includes the customer content without defects, to remove the straight customer content lines on the scanned test pages. The remaining dark or light straight line along the printing process direction may be streak defects. For the detected streak defects, we use the DAG-SVM multi-classification method to classify the rank of streak defects in the raster ROI.

1. Introduction

In previous work on streak defect detection, we designed an algorithm to detect the streak defects only on standard printed test pages, which are uniform color test pages, as shown in Figure 1 (a) [1][2]. In this paper, we design an algorithm to detect the streak defects on the customer pages, including customer content, as shown in Figure 1 (b).

This raster ROI streak detection and classification project is part of a larger effort to analyze a variety of types of print quality defects in pages containing customer content [3]. In our previous work, we designed a method to extract four different ROIs from the master image [4], as shown in Figure 2. In this project, we need to use the raster ROIs from the master image and the scanned test image. Because of their unstructured content, customer content in the raster ROIs will be the most challenging ROIs in which to detect the streak defect compared with the symbol, vector, and background ROIs. The symbol, vector, and background ROIs contain uniform or slowly varying color. We can detect the streak



Figure 1: The streak defects sample images.

defects using the method we designed for the standard test images. In this paper, we design a method to detect the streak defects in a raster ROI. After detecting the streak defects in a raster ROI, we classify the raster ROI streak defect into four ranks: A, B, C, and D.

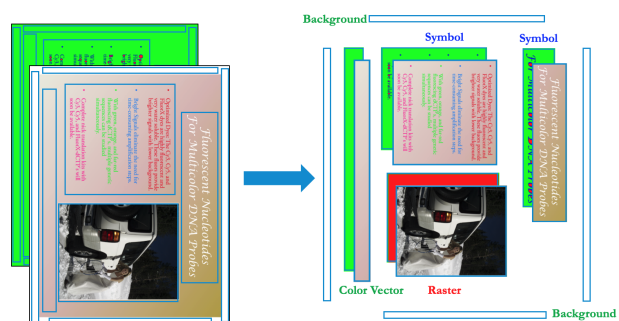


Figure 2: The ROI extraction input and output.

*Research supported by HP Inc., Boise, ID 83714

2. Raster ROI streak detection and classification procedure

In this section, we introduce the details of the raster ROI streak defect detection and classification procedure. Figure 3 shows the overall pipeline of the proposed method.

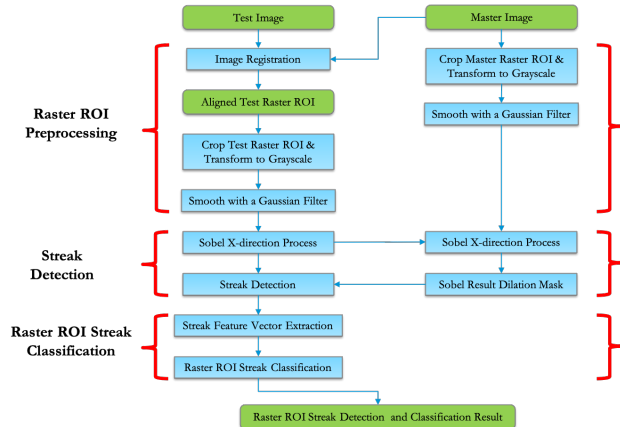


Figure 3: The overall pipeline of the raster ROI streak defect detection and classification.

This procedure includes three parts: the raster ROI preprocessing part, the streak defects detection part, and the raster ROI streak defects classification part. In the first part, we preprocess the master and test images before extracting the streak defects from the raster ROIs. Because there is misalignment between the corresponding master and test images, we need to do the alignment first [5]. Then, we extract the corresponding master and test raster ROIs and transform them to grayscale. Besides, the scanned printed test raster ROI includes a halftone pattern. So we use a Gaussian filter [6] to smooth both the master and test raster ROIs and to remove the halftone pattern. This process can improve the streak detection result.

In the second part, we use the Sobel X-direction filter [7], which is the same direction as the printing process direction, to process the master raster ROI. The Sobel X-direction filter process result can extract the straight lines of customer content along the printing process direction in the master raster ROI. We then use the dilation [8] of the Sobel X-direction filter result as a mask image to remove the customer content straight lines from the test raster ROI. The remaining dark or light straight lines along the printing process direction in the test raster ROI are the streak defects.

In the third part, we classify the raster ROI streak defects. In this part, we extract the feature vector of the raster ROI streak defects. The multi-classification method classifies [9] the raster ROI streak defects based on the streak defect feature vector. We use the Directed Acyclic Graph-Support Vector Machine (DAG-SVM) [10] for the multi-classification method.

After completing Parts 2 and 3 of the raster ROI streak detection and classification procedure, we will get the streak defects detection result and the raster ROI streak defects classification result. We will introduce the details of the raster ROI streak defects detection and classification procedure in the following sections.

3. Raster ROI Preprocessing

In this section, we introduce the raster ROI preprocessing, which includes four steps: image registration, raster ROI extraction, color space transformation, and Gaussian smoothing.

3.1 Raster ROI image registration

The input is the master image and scanned test image, as shown in Figures 4 (a) and (b). Because of the limited printing and scanner precision, there is always misalignment between the master image and the scanned test image, as shown in Figure 4 (c). So, before we use the master image raster ROI result to crop the corresponding raster ROI from master and scanned test image, we must do the image registration between the master and scanned test images to remove this misalignment.

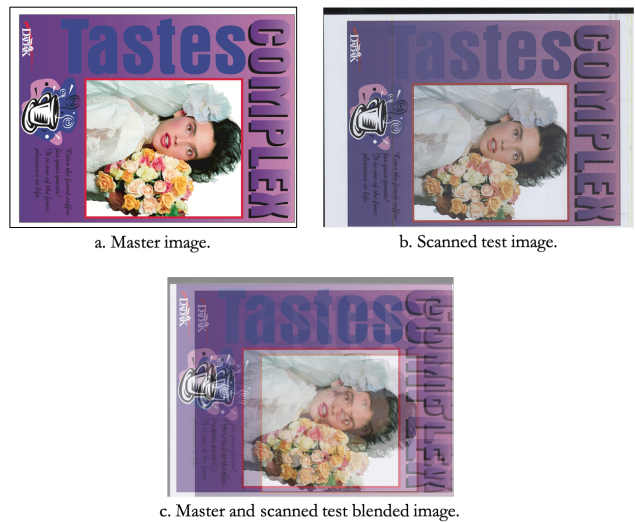


Figure 4: The master image, scanned test image, and the master/scanned test blended image.



Figure 5: The image registration result: master and aligned test blended images.

We use the SIFT method [11] [12] first to extract the interest points from the master and scanned test images. Then, the Normalized Cross-Correlation (NCC) [13] can establish the correspondences of the interest points between the master and scanned test images. Finally, the random sample consensus (RANSAC) algorithm [14] can calculate the optimal transformation matrix based on the corresponding interest

points [15]; and we can use the optimal transformation matrix to process the scanned test image to remove the misalignment. Figure 5 shows the master and aligned test blended image. The master and aligned test images match much better than in the original blended image Figure 4 (c). But we still can find that there is a small misalignment between the master and test images, as shown in the Figure 5 detail image.

3.2 Raster ROI extraction

The next step is to crop out the corresponding raster ROI from the master and aligned test images. The raster ROI result is based on our previous work [4], as shown in Figure 6 (a) (only the white area is the ROI). The raster ROI result is four numbers for each ROI: the left top pixel position and the right bottom pixel position. This “Complex” sample image only includes one raster ROI and the raster ROI cropped result shown in Figures 6 (b) and (c).

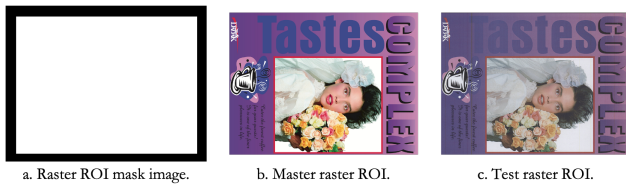


Figure 6: The master raster ROI and the aligned test raster ROI.

3.3 Color space transformation and Gaussian filtering to smooth the raster ROI

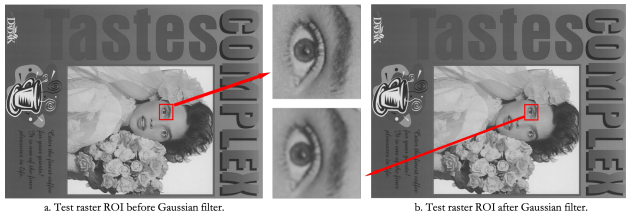


Figure 7: Processing the test raster ROI with a Gaussian filter to remove the halftone pattern. The reader is advised to zoom in to see the halftone pattern in the upper center enlargement.

After we crop out the corresponding raster ROI from the master and test images, we transform the master and aligned test raster ROI color space to grayscale. Because there is a halftone pattern in the printed test image, we use a Gaussian filter to process the test image and remove the halftone pattern. For our 300 dpi scanned test images, we set the Gaussian filter size to be 15 pixels × 15 pixels with $\sigma = 2$. Equation 1 shows the Gaussian smoothing. Here, $f(x, y)$ is the input grayscale image; $g(i, j)$ is the 15 × 15 Gaussian filter; and $f'(x, y)$ is the Gaussian smoothed result. Figure 7 shows the grayscale raster ROI Gaussian smoothed result. Figure 7 (a) shows the pre-smoothed raster ROI, including the halftone pattern, and Figure 7 (b) shows the Gaussian smoothed raster ROI. To maintain consistency, we also use

the same filter to smooth the master grayscale ROI.

$$f'(x, y) = \sum_{i=-7}^7 \sum_{j=-7}^7 f(x-i, y-j)g(i, j) \quad (1)$$

After completion of this step, we get the smoothed test raster grayscale ROI and the smoothed master grayscale ROI. In the next section, we will introduce how to detect streak defects using these two images.

4. Raster ROI streak detection

Streak defects are dark or light straight lines along the printing process direction. To detect the streak defects, we first need to remove the customer content straight lines along the printing process direction directly from the test raster ROI. The remaining straight lines along the printing process direction will be the streak defects.

To extract the customer content straight lines, we use the Sobel operator. The Sobel operator is an edge detection algorithm, and it includes two 3 × 3 edge detection kernels: one for horizontal edge detection, and one for vertical edge detection. Equation 2 shows the Sobel operator computation process. A is the source image, which is the grayscale raster ROI in our project. G_x and G_y are two images that, at each point, contain the vertical and horizontal derivative approximations, respectively. The “*” denotes the 2-dimensional signal processing convolution operation.

$$G_x = \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix} * A \text{ and } G_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * A \quad (2)$$

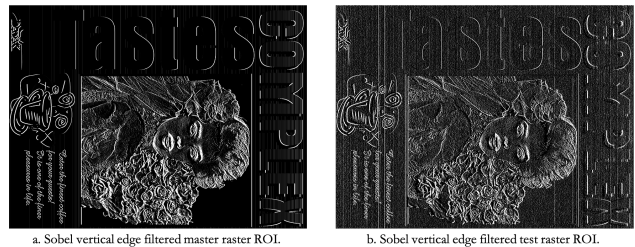


Figure 8: The Sobel vertical edge filter processed master raster ROI and test raster ROI.

In our project, we orient all the scanned test images in the same direction relative to the printing process direction, so the streak defects are vertical lines on the scanned test images. The G_x image will show all the vertical lines on the master image and the test image. We use the Sobel vertical edge filter to process the grayscale master raster ROI and corresponding test raster ROI; and we normalize the vertical edge detection result to the range 0–255, as shown in Figure 8.

Figure 8 (a) shows the result of applying the Sobel vertical edge filter to the master raster ROI. The white color pixels label all the vertical lines in this raster ROI. Figure 8 (b) shows the result of applying the Sobel vertical edge filter to the test raster ROI. We can find that there are many noise test raster ROI vertical edge detection results, even

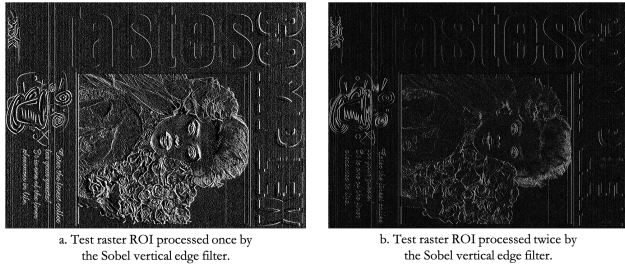


Figure 9: Comparison of the once and twice Sobel vertical edge filtered results for the test raster ROI.

though we have used a Gaussian filter to smooth the test image. Here, we want to remove these noise lines from the test raster ROI vertical edge detection result, and only keep the customer content vertical lines and streak defects. The method is to use the Sobel vertical edge filter to process the test raster ROI twice. Intuitively, we calculate the second derivative of the test raster ROI in the horizontal direction to extract the vertical lines in test raster ROI. We also normalize the twice Sobel vertical edge filtered result to the range 0 – 255, as shown in Figure 9 (b). In Figure 9, comparing (a) and (b), we can find that applying Sobel vertical edge filter twice can remove most of the noise in the test raster ROI and only keep the customer content vertical lines and streak defects.

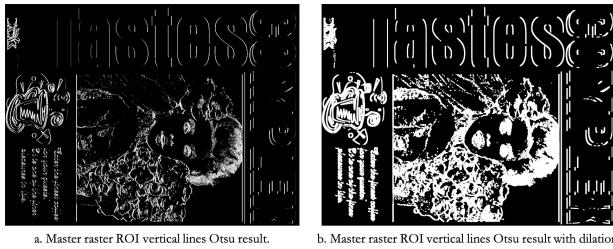


Figure 10: The master raster ROI vertical lines Otsu result and dilation result.

We apply the Otsu method [15] to the master raster ROI Sobel vertical edge detection result and extract the customer content vertical lines, as shown in Figure 10 (a). This Otsu result is a binary image, and the white pixels are the customer content vertical lines. We want to use this binary image as a mask to remove all the customer content vertical lines from the test raster ROI. However, as shown in Figure 5, even though we did the image registration between the master and the test images, there is a remaining 5 – 20 pixel misalignment. To remove the influence of this misalignment between master raster ROI and test raster ROI, we dilate the Otsu result of the master raster ROI, as shown in Figure 10 (b). The dilation kernel is 9 pixels × 9 pixels. We will use this dilation result as a mask to remove the customer content vertical lines from test raster ROI.

For the next step, we use the dilation result of the master raster ROI vertical lines image to process the test raster ROI vertical lines detection result. Because the master raster ROI dilation result is the binary image and the foreground is the dilated vertical lines, we should invert this dilation image and set the vertical lines in the dilation areas to zero.

Then, we use this inverse dilation master raster ROI image to mask the test raster ROI vertical lines result, thereby removing the customer content. Figure 11 (a) is the test raster ROI Sobel vertical lines detection result, and Figure 11 (b) is the processing result using the master raster ROI vertical lines dilation mask image. We can find most of the customer content vertical lines are removed from the test raster ROI vertical lines detection, but the streak defects remain.

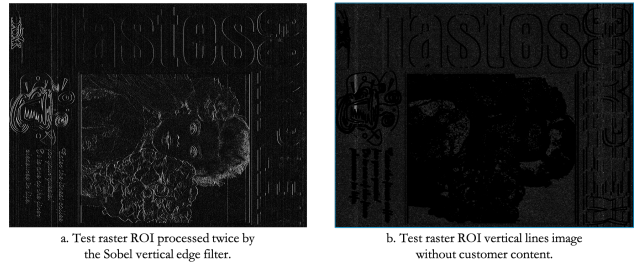


Figure 11: Removing customer content vertical lines from the test raster ROI using the master raster ROI vertical lines dilation mask image. The reader is advised to zoom in to see the noise in (b).

In Figure 11 (b), it is evident that the white lines are the streak defects in the test raster ROI. If we look closely, we can find many small white noise points in Figure 11 (b). These small white areas may not be the streak defects and usually are noise. We prefer to extract the apparent streak defects in the vertical lines from Figure 11 (b).

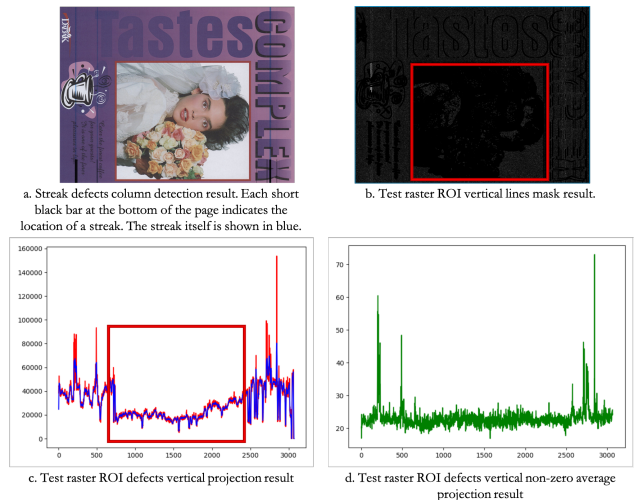


Figure 12: The test raster ROI streak defects column detection result. The reader is advised to zoom in to see the fine streaks in (a).

We calculate the vertical projection of the test raster ROI vertical lines image without customer content Figure 12 (b) and get the red projection result in Figure 12 (c). Because the vertical projection result fluctuates greatly, we use a 9 pixel unweighted average kernel to smooth the vertical projection result [16], as shown in Figure 12 (c) blue line. In the smoothed vertical projection result, we use the average vertical projection smoothed result plus the standard deviation of the vertical projection smoothed result as the threshold to

label the streak defects. The streak defects column detection results are indicated by the short black bars in Figure 12 (a). This streak defects column result includes all the streak defects in the test raster ROI. However, in Figure 12 (c), we find a large low-value vertical projection area indicated by the red bounding box. This area does not include streak defects, but its projection value is much less than that of other good print quality areas. This is because the corresponding columns include a large area of zero-valued mask, which results in very small vertical projection values. This unsmooth projection result will sometimes influence the streak detection result because a lot of small vertical projection values will pull down the streak detection threshold. To solve this problem, we normalize the vertical projection value by the number of non-zeros pixels in the corresponding column and get the vertical non-zero average projection result, as shown in Figure 12 (d). With this new vertical non-zero average projection result, it is much more accurate to extract streak defects in the raster ROI using the average plus the standard deviation of the non-zero averaged projection value as the threshold.

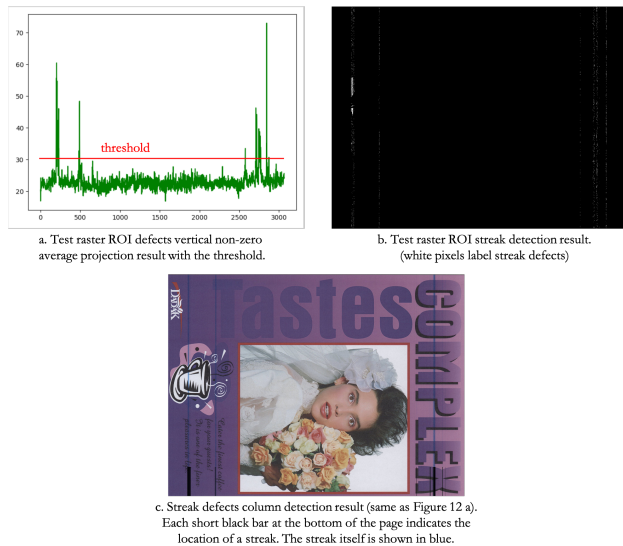


Figure 13: The test raster ROI streak defects result. The reader is advised to zoom in to see the fine details of the streak defects.

The last step of streak defects detection in the raster ROI is to extract the pixels of streak defects from the corresponding columns. We apply Otsu’s method to the non-zero values of the test image raster ROI vertical lines mask result, as shown in Figure 12(b), because the masked zero values will pull down the Otsu threshold and influence the accuracy of streak defects detection result. The pixels’ Sobel X-direction twice processed value, which are above the Otsu threshold and in a streak defects labeled column, are the streak defects pixels, as shown in Figure 13 (b). We use the white pixels to label the streak defects.

5. Raster ROI streak defects classification

In this section, we extract the feature vector of streak defects in the raster ROI and classify the raster ROI streak defects into one of four different ranks. The requirement of

four different ranks is same as the text fading classification requirement: Rank A means no streak defects in the raster ROI; Rank B means there are streak defects in the raster ROI, and the defect does not affect the regular use of the printed page; Rank C means there is an observed text fading defect and people can find it easily; Rank D means there are a lot of text fading defects and the defects influence the regular use of the printed page. Figure 14 shows four different rank samples of raster ROI streak defects.

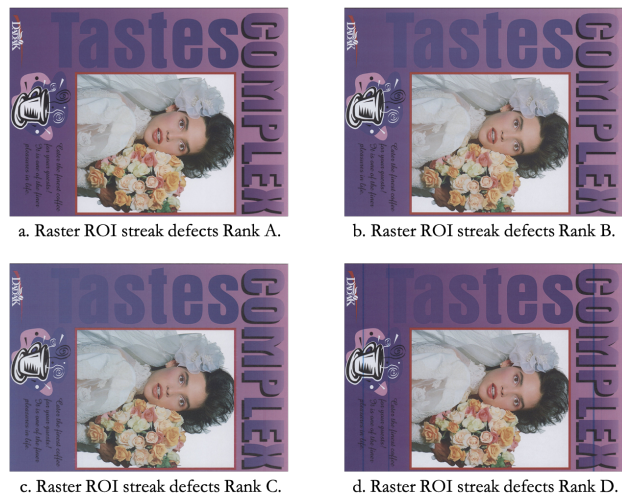


Figure 14: The raster ROI streak defects rank samples. The reader is advised to zoom in to see all the streak defects.

Table 1: Raster ROI Streak Defect Feature Vector for a Single Test Page

1	The area of the streak defects;
2	The total width of the streak defects;
3	The average length of the streak defects;
4	The average sharpness of the streak defects;
5	The average ΔE of the streak defects;
6	The severity ($\Delta E \times streak_{area}$) of the streak defects.

Table 1 shows the six features for raster ROI streak defects on a single page. We will use these features to classify the raster ROI streak defects. The data set includes 37 raster ROI streak defects images. There are 13 images of Rank A, 8 images of Rank B, 7 images of Rank C, and 9 images of Rank D. The multi-classification model is Directed Acyclic Graph-Support Vector Machine (DAG-SVM) [17]. To evaluate this multi-classification result, we apply cross-validation [18]. We separate the 37 raster ROIs into 5 folds, i.e. (7/7/7/8/8) raster ROI streak defects ground truth for each fold. Each fold is used as the test set one time, and the other four folds are used as training sets. The final confusion matrix for this five-fold cross-validation is shown in Table 2. The accuracy of this multi-classification model is 81.1%, and the standard deviation of the accuracy of these five folds is 11.8% with a worst accuracy of 57.1% and a best accuracy of 87.5%.

Table 2: Raster ROI streak defect DAG-SVM classification confusion matrix

	Predict A	Predict B	Predict C	Predict D
Real A	11	1	0	1
Real B	1	6	0	1
Real C	0	1	5	1
Real D	0	1	0	8

Conclusion

In this chapter, we proposed a raster ROI streak defects detection and classification method. This method's input is the master raster image and the corresponding test raster image, and the output is the streak defect detection result with the feature vector and the raster ROI streak defects rank classification result. This method includes raster ROI preprocessing, streak detection, and the raster ROI streak defects classification. In the first part, we use an image registration method and use the Gaussian filter to smooth the test raster ROI and master raster ROI. In the second part, we use the Sobel vertical lines detection filter to extract all the customer content vertical lines and use this result to remove all the customer content vertical lines in the test raster ROI so that only the streak defects vertical lines remain. In the third part, we extract the feature vector from the streak defect detection result and use DAG-SVM to classify the rank of the raster ROI streak defects. All these processes are tested with HP supplied test images.

References

- [1] M. Q. Nguyen, S. Astling, R. J. Jessome, E. Maggard, T. Nelson, M. Q. Shaw, and J. P. Allebach, "Perceptual Metrics and Visualization Tools for Evaluation of Page Uniformity," *Electronic Imaging, Image Quality and System Performance XI*, San Francisco, CA, 3-5 February 2014.
- [2] R. Zhang, E. Maggard, R. Jessome, Y. Bang, M. Cho, J. P. Allebach, "Block Window Method with Logistic Regression Algorithm for Streak Detection," *Electronic Imaging, Image Quality and System Performance XVI*, Burlingame, CA, January 2018.
- [3] R. Zhang, Y. Yang, E. Maggard, Y. Bang, M. Cho, and J. P. Allebach, "A Comprehensive System for Analyzing the Presence of Print Quality Defects," *Electronic Imaging, Image Quality and System Performance XVII*, San Francisco, CA, 26 January 2020.
- [4] R. Zhang, E. Maggard, Y. Bang, M. Cho, J. P. Allebach, "Region of Interest Extraction for Image Quality Assessment," *Electronic Imaging, Image Quality and System Performance XVII*, Burlingame, CA, January 2018.
- [5] R. Zhang, E. Maggard, Y. Bang, M. Cho, M. Q. Shaw, and J. P. Allebach, "Symbol ROI Text Fading Detection," *Electronic Imaging, Color Imaging: Displaying, Processing, Hardcopy, and Applications XVII*, San Francisco, CA, 26 January 2021.
- [6] G. Deng, and L.W. Cahill, "An Adaptive Gaussian Filter for Noise Reduction and Edge Detection," *IEEE Conference Record Nuclear Science Symposium and Medical*

Imaging Conference, San Francisco, CA, 31 Oct.-6 Nov. 1993.

- [7] W. Gao, X. Zhang, L. Yang and H. Liu, "An Improved Sobel Edge Detection," *2010 3rd International Conference on Computer Science and Information Technology*, Chengdu, China, CA, 9-11 July 2010.
- [8] M. L. Comer, and E.J. Delp, "Morphological Operations for Color Image Processing," *Journal of Electronic Imaging*, 8(3), pp. 279-289, 1999.
- [9] D.M.J. Tax, and R.P.W. Duin, "Using two-class classifiers for multiclass classification," *IEEE, Object Recognition Supported by User Interaction for Service Robots*, Quebec City, QC, Canada, 11 August 2002.
- [10] P. Chen, and S. Liu, "An Improved DAG-SVM for Multiclass Classification," *2009 Fifth International Conference on Natural Computation*, Tianjian, China, 14-16 August. 2009
- [11] H. Foroosh, J.B. Zerubia, and M. Berthod, "Extension of Phase Correlation to Subpixel Registration," *IEEE Transactions on Image Processing*, 7 August 2002.
- [12] D.G. Lowe, "Object Recognition from Local Scale-invariant Features," *Proceedings of the Seventh IEEE International Conference on Computer Vision*, Kerkyra, Greece, 6 August 1999.
- [13] F. Zhao, Q. Huang, and G. Wen, "Image Matching by Normalized Cross-Correlation," *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, Toulouse, France, 14-19 May 2006.
- [14] M. A. Fischler, and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Communications of the ACM*, Toulouse, France, 1 June 1981.
- [15] D. Liu, and J. Yu, "Otsu Method and K-means," *IEEE, Ninth International Conference on Hybrid Intelligent Systems (Vol. 1, pp. 344-349)*, August, 2009.
- [16] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, New York: Springer Series in Statistics, August, 2009.
- [17] P. Chen, and S. Liu, "An Improved DAG-SVM for Multiclass Classification," *2009 Fifth International Conference on Natural Computation*, Tianjian, China, 14-16 August. 2009
- [18] M. W. Browne, "Cross-Validation Methods," *Journal of Mathematical Psychology*, 44(1), pp.108-132, Mar, 2000.

Author Biography

Runzhe Zhang received his B.S. degree in Mechanical Engineering from Qingdao University of Technology (2013), Shandong, China and received the M.S. degree in Mechanical Engineering from Boston University, MA, USA. Currently, he is pursuing the Ph.D. in Electrical and Computer Engineering at Purdue University. His research areas include digital image processing, computer vision, and machine learning.

Eric Maggard received his B.S. degree in Physics from Northwest Nazarene University, Nampa, Idaho in 1991 and the M.S. degree in Computer Science specializing in image analysis and processing from Walden University in 2006. He is an Expert Imaging Scientist in the LaserJet Hardware

Division and has developed programs and image quality algorithms for the last 15 years that are used in the testing of LaserJet print and scan image quality. His interests include machine vision, object recognition, machine learning, robot control, and navigation.

Yousun Bang is a manager of Image Quality Part in the Imaging Lab at HP Printing Korea Co. Ltd. She received her BS and MS in mathematics from Ewha Womans University, Seoul, Korea in 1994 and 1996, and her Ph.D. in the School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana in 2005. She worked for Samsung Advanced Institute of Technology and Samsung Electronics Company from 2004 to 2017. Her current research interests include image quality diagnosis and metrics and ML/DL based prediction for smart device services.

Minki Cho is an engineer with HP Printing Korea. He received B.S.(1997) and M.S.(1999) in electrical engineering from the Inha University, Korea. From 2003 to 2017, he worked for Samsung Electronics and Samsung Advanced Institute of Technology. His research areas are print image processing, print image quality diagnosis and calibration. Recently, he is researching the above interests using ML and DL.

Mark Shaw received his B.S. in Printing and Publishing Technology from University of Herfordshire in 1997 and his M.S. in Color Science from Rochester Institute of Technology in 1999. Then he went to Boise State University to take Electrical Engineering classes to meet the pre-requisites to study under Prof. Jan Allebach at Purdue. Finally he received his Ph.D. in Electrical and Computer Engineering from Purdue University in 2015. He is now a principal color and imaging master architect at HP Inc.

Jan P. Allebach is Hewlett-Packard Distinguished Professor of Electrical and Computer Engineering at Purdue University. Jan P. Allebach is a Fellow of the IEEE, the National Academy of Inventors, the Society for Imaging Science and Technology (IS&T), and SPIE. He was named Electronic Imaging Scientist of the Year by IS&T and SPIE and was named Honorary Member of IS&T, the highest award that IS&T bestows. He has received the IEEE Daniel E. Noble Award, the IS&T/OSA Edwin Land Medal, the IS&T Johann Gutenberg Prize, and is a member of the National Academy of Engineering.

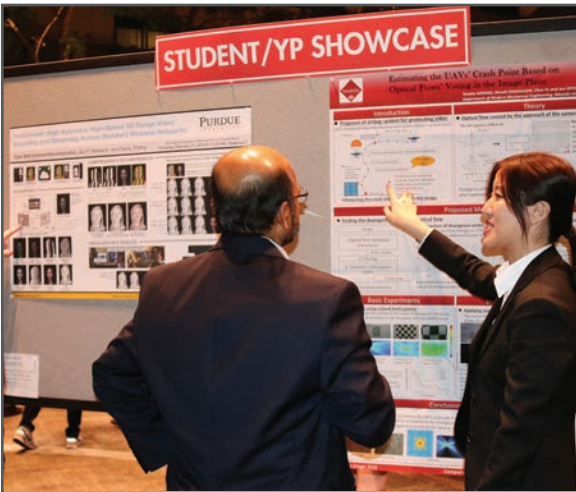
JOIN US AT THE NEXT EI!

IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!



- **SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •**
- **INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •**

www.electronicimaging.org

