https://doi.org/10.2352/ISSN.2470-1173.2017.12.IQSP-234


# Detection of streaks caused by dust in the sheetfed scanners

***Daulet Kenzhebalin*** [a], ***Xing Liu*** [a], ***Ni Yan*** [a], ***Peter Bauer*** [b], ***Jerry Wagner*** [b], ***Jan Allebach*** [a]
[a] ***School of Electrical and Computer Engineering, Purdue University; West Lafayette, IN***
[b] ***HP Inc.; Boise, ID***

## Abstract

*Dust in the scanner may cause vertical streaks in the scanned image since it reflects some part of the incident light. In this paper, we propose a method for detecting streaks in the scanned images that are a direct result of dust on the scanner glass. This lets customers resolve the issue without calling the maintenance. The solution includes denoising, conversion to opponent color space, calculation of $\Delta E'$, calculation of features, and classification. We denoise the image in order to remove halftones in case the image was halftoned. Opponent color space lets us look separately at luminance channel and chrominance channels. We have developed three features that use the data in luminance and chrominance channels. Eventually, we will use these features to detect streaks, and distinguish them from content.*

## 1. Introduction

Sheet-fed scanners allow customers to obtain scans of stacks of pages a lot faster and with less interaction compared to flatbed scanner. In sheet-fed scanners the page is fed automatically while the scan head remains stationary. Dust, that sticks to the scan head glass, causes vertical streaks in the scanned image, which are not part of the original paper. An example of an image with dust streak obtained using sheet-fed scanner is shown in Fig. 1. Figure 2 shows a zoom of the region of the image with dust streak.

This work builds on recent image quality work focused on printer and scanner products that was conducted in our laboratory, and which addressed assessment of page non-uniformity [1]-[6], fine-pitching banding [7]-[11], ghosting [12], local defects [13],[14], fading [15],[16], scanner MTF [17], and scanner motion quality [18].

Rosario et al. proposed an algorithm for detecting streaks in printed images [19]. However, they assume that the scanned image does not have content on it. In our case, we are designing an algorithm to work on any image.

Given part of the image, some of the dust streaks are very hard to distinguish from the content streaks. People can connect image components and figure out whether the streak was caused by dust or was intended to be in the image. There are research studies that perform object mapping for an image. However, these algorithms have high complexity and cannot be adopted to the hardware implementation. Therefore, currently the dust identification makes errors in prediction.

We categorize these errors into two types: misses and false alarms. Miss occurs when the detection algorithm does not find a real streak caused by dust. False alarm occurs when the detection algorithm finds a streak, which was not caused by dust. The aim is to have a small number of false alarms while detecting most of the real dust streaks.

***Figure 1.*** *Scan of an image with dust streak*

In this paper[1], we propose a solution to finding vertical streaks in the image that were caused by dust in the scanner. Next sections will cover the proposed procedure, the results and conclusion.

## 2. Procedure

### *2.1. Preprocessing*

We start with a scanned image that is stored in linear RGB color space. If the image is gamma corrected, then gamma uncorrection is done first, and followed by the procedure below.

First, the image is converted from RGB color space to CIE L*a*b* color space. CIE L*a*b* is an opponent color space that stores luminance and chrominance in separate channels. After conversion, we need to apply a descreening algorithm to remove halftones. Since dust causes only vertical streaks in the scanned image, a one-dimensional vertical Gaussian filter of size 9 and $\sigma = 3$ was used for descreening. Figure 3a shows an original

[1]Research supported by HP Inc.

***Figure 2.*** *Crop of an image with dust streak*

image and Fig.3b shows a descreened image. Here is the equation for descreening:

$$g(x,y,c) = \sum_{k=-4}^{4} f(x,y+k,c)gaus(k), \quad (1)$$

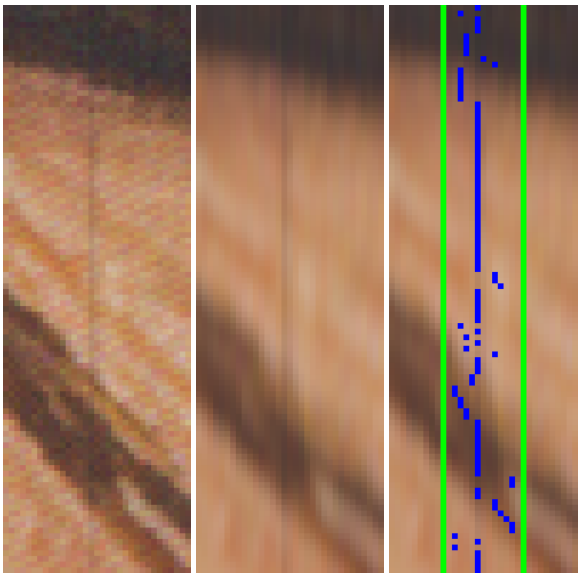
where $f(x,y,c)$ is an input image in L*a*b*, $gaus(k)$ is a gaussian window of size 9, and $g(x,y,c)$ is a descreened image.

The next step is to obtain a two-dimensional $\Delta E'$. $\Delta E'$ is a measure of how different a given pixel is from neighbor pixels in the horizontal direction. We used median function to find the representative of the local window, which is our *baseline*. We then subtract the *baseline* from the image to compute $\Delta E'$. The *baseline* and $\Delta E'$ are defined in Eqs. 2 and 3. Most of the streaks caused by dust are very narrow, 1 to 3 pixels wide. We used only luminance ($L^*$) channel for computing $\Delta E'$ because it contains most of the information about the content of the image. Also, dust streaks are usually light or dark shade of the neighboring pixels.

$$baseline(x,y,c) = median(g(x-5,y,c),...,g(x+5,y,c)) \quad (2)$$

$$\Delta E'(x,y) = g(x,y,L^*) - baseline(x,y,L^*) \quad (3)$$

Next, the image is split into vertical column strips with width $w = 13$ pixels and an overlap $o = 6$ pixels. Splitting the image into column strips allows us to look at small local changes in $\Delta E'$. Column strip size of 13 pixels lets us process the image quickly while still being able to find local distortions. An example of a column strip row is given in Fig. 4.



(a) Original image (b) Descreened image (c) Descreened image with annotatiions

***Figure 3.*** *Crop of an image on Fig. 1*

## 2.2. Features

The next step is the extraction of features. We have developed 3 features that enable us distinguish dust streaks from content.

### 2.2.1. Feature 1. Locally averaged peak location derivative magnitude (PLD)

This feature looks for vertical lines and streaks in the image. For a given column strip, we find locations of local peaks and valleys in each row. In this paper, we don't distinguish between a valley and a peak. We only look at what the magnitude of the $\Delta E'$ is at a particular point. However, we are not using absolute value of $\Delta E'$ because that way we will not be able to find correct edges of the peaks or valleys. Figures 4 and 5 show the same column strip row using $\Delta E'$ and $|\Delta E'|$, respectively. In Fig. 4 there are two peaks at locations 6 and 9 and two valleys at locations 2 and 7. However, when using absolute value, there is only one peak at location 6 with width of 5 pixels. This is the reason for not using absolute value of $\Delta E'$.

Figure 3c shows a crop of a descreened image with annotations. The green lines are the edges of the column strip. The blue marks are the peak locations in each row. In the areas where background is smooth, the peak locations form a vertical line. In the area where the dust streak goes through hair, the blue dots are not vertically aligned.

After finding the peak locations we find the magnitude of the peak location derivative in the vertical direction. Then, we average it across 21 rows. The PLD is defined as $f_1(x,y)$ and is calculated the following way:

$$f_1(x,y) = \frac{\sum_{k=-10}^{9} |pl(x,y+k) - pl(x,y+k-1)|}{20}, \quad (4)$$

where $pl(x,y)$ is a peak location in column strip $x$ and row $y$. If this feature has a low value, then it means that there is a vertical


***Figure 4.*** *Example of a column strip row*

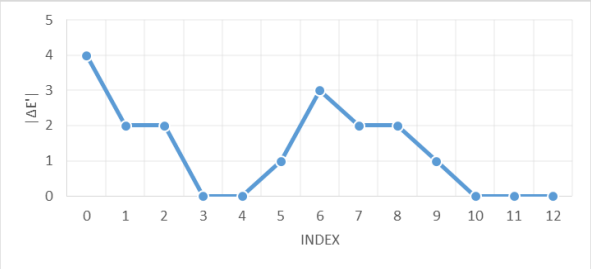



***Figure 5.*** *Example of a column strip row using absolute value of* $\Delta E$

streak in this column strip.

#### 2.2.2. Feature 2. Peaking factor (PF)

This feature shows the relative strength of the peak in the column strip row. If the peaking factor has a very high value, then it is most likely to be a content, such as a line. If the peaking factor is low, then the peak is very weak and, consequently, not noticeable. We define peaking factor as a ratio of sum of constant value $c$ and 3 highest values in a column strip row to a sum of $c$ and 3 lowest values in a column strip row, as given in Eqs. 5-7:

$$A_{i+1}(x,y) = A_i(x,y) - max(A_i(x,y)), \tag{5}$$

$$B_{i+1}(x,y) = B_i(x,y) - min(B_i(x,y)), \tag{6}$$

$$f_2(x,y) = \frac{c + \sum_{i=1}^{3} max(A_i(x,y))}{c + \sum_{i=1}^{3} min(B_i(x,y))}, \tag{7}$$

where both $A_1(x,y)$ and $B_1(x,y)$ are sets of values in the row $y$ and column strip $x$. Difference on a set is defined as removing an element from a set. Hence, $A_2(x,y)$ is a set of values in the row $y$ and column strip $x$ without the highest value and $max(A_2(x,y))$ is the second highest value in column strip $x$ and row $y$. The range of values that peaking factor can take is from 1 to $255*3+1 = 766$.

#### 2.2.3. Feature 3. Side Difference (SD)

Sometimes the change in color background results in a small peak in $\Delta E'$, which resembles the peak of dust streaks. In order to distinguish change in content from a dust streak, we introduce the third feature - side difference. Side difference shows difference in color between the left side of the peak and the right side of the peak. First, we find the edges of the peak. We define peak edge as the index closest to the peak index, where the value drops below 20% of the peak value. Second, we take the average of 3 pixels following the peak edge on both sides. We call average value of the left side and average value of the right side as $color_l$ and $color_r$, respectively. Third, we compute the side difference by using Eq. 8:

$$f_3(x,y) = \sqrt{\sum_c (color_l(x,y,c) - color_r(x,y,c))^2}, \tag{8}$$

where $x$ is columnstrip index, $y$ is row index, $c$ is channel index, and $l$ and $r$ are relative sides of the peak.

### 2.3. Defective mask

After extracting features, we can decide whether the image has dust streaks. This is done by thresholding the features. The thresholds for features can be trained. If PLD is small, then the peak location is not changing in the vertical direction. This means that there is a vertical line in the image. Therefore, if the PLD is smaller than threshold ($PLDmax$), then there is a vertical line. The PF shows the strength of the peak relative to the columnstrip. If PF is small, then the disturbance is very small and probably not noticeable to the user. On the other hand, if it is very big, then it is very likely to be a content, such as frame border or line. Therefore, if PF is larger than threshold ($PFmin$) and smaller than threshold ($PFmax$), then it might be a line caused by dust. The SD shows whether the background color is changing. Therefore, if the SD is smaller than threshold ($SDmax$), then the streak is a vertical line.

The defective mask is formed by thresholding each feature with corresponding thresholds. For a column strip row to be defective, all 3 features has to pass thresholding. For example, if a column strip row has a PLD, which is smaller than $PLDmax$, and a PF, which is smaller than $PFmin$, then this column strip row is not defective. Therefore, the defective column strip row has a PLD value smaller than $PLDmax$, PF value smaller than $PFmax$ and larger than $PFmin$, and SD value smaller than $SDmax$.

### 2.4. Postprocessing

The output defective mask that we get after thresholding 3 features sometimes has false alarms. Number of these false alarms can be greatly reduced by looking at a post processing feature(PPF) feature that is related to the number of defective pixels in a columnstrip. This feature is not related to the actual image data, but only to the output of thresholding 3 features introduced in the previous section.

First, we obtain the lengths of all the defective segments in each columnstrip. We call a run of consecutive defective rows in a columnstrip a defective segment. The table of segments in one columnstrip is given in Fig. 6a. Then we plot the number of segments against the length in Fig. 6b. Next, we multiply the curve by its length and obtain the plot in Fig. 6c. Eventually, we integrate the new curve and get a PPF feature.

## 3. Training

### 3.1. Ground truth

The ground truth consists of a set of 128 images which were scanned and provided to us by HP inc. Each image was manually evaluated and scored based on the strength of the streak. Each streak on each image was given a score from 1 to 9, where 1 is the least visible and 9 is the most visible.

### 3.2. Thresholds Training

We have three features and four thresholds associated with them. We trained these thresholds on the ground truth. We know that currently the algorithm cannot find all of the dust streaks without finding streaks that were not caused by dust. Hence, we will have some misses and some false alarms. Miss occurs when the algorithm does not find the dust streak. False alarm occurs when the algorithm finds a streak, which is not a dust streak. There is a tradeoff between the miss rate and false alarm rate. Miss rate and false alarm rates are computed as in Eqs. 9,10, respectively.

$$missrate = 1 - \frac{tp}{tp+fn}, \tag{9}$$

$$falsealarmrate = \frac{fp}{fp+tn}, \tag{10}$$

where $tp$ is true positive, $fp$ is false positive, $tn$ is true negative, and $fn$ is false negative.

We needed to find optimal thresholds that would find most of the dust streaks, while finding very little false alarms. The miss rate has much smaller denominator compared to false alarm rate. The denominator in miss rate is number of defective column strip rows, whereas the denominator in false alarm rate is the number of non-defective column strip rows. Therefore, we decided to set the upper bound for the miss rate at 10% and looked for combination of feature thresholds that would achieve the lowest false alarm rate.

## 4. Results

We ran the training on the dust streaks that have a score of 4 and above. After training the algorithm, the minimum false alarm rate was found to be 0.19% at miss rate of 9.47%. That is 110680 false alarm column strip rows and 8094 missed column strip rows. The example images are given in Figs. 7-9. The ground truth images have two colored lines superimposed on the original image. The yellow means that there is a dust streak with a score $< 4$ and red means that the dust streak has a score $\geq 4$

The feature thresholds were optimized to the following values:

$$PLDmax = 0.15, \tag{11}$$

$$PFmin = 3.9, \tag{12}$$

$$PFmax = 33, \tag{13}$$

$$SDmax = 5, \tag{14}$$

$$PPFmax = 1000. \tag{15}$$

Also, the detection algorithm breaks up the dust streaks because of content. Some of the misses were due to dust streaks being masked by the content. Overall, the algorithm found most of the dust streaks.

We have images with tables and graphs in our training database, which are currently identified as dust streaks by the algorithm. These false alarms can be reduced in the future works. Also, images that have vertical lines cause false alarm detection by the algorithm. It is very hard to distinguish content lines that were meant to be in the paper by a customer from those that were produced as a result of a dust in the scanner.

## 5. Conclusion

In colncusion, we developed an algorithm for identifying vertical streaks in scanned images that were caused by dust. Using these detections, one can clean the scanner and remove the dust. The algorithm was able to find most of the dust streaks. However, the algorithm also detects some content lines as dust streaks. In the future we are going to work on finding content streaks, so that they are not detected by the algorithm.

## References


[1] X. Jing, S. Astling, R. Jessome, E. Maggard, T. Nelson, M. Q. Shaw, and J. P. Allebach, "A General Approach for Assessment of Print Quality," Image Quality and System Performance X, SPIE Vol. 8653, P. D. Burns and S. Triantaphillidou, Eds. San Francisco, CA, 3-7 February 2013.

[2] X. Liu, G. Overall, T. Riggs, R. Silveston-Keith, J. Whitney, G. T. C. Chiu, and J. P. Allebach, "Wavelet-Based Figure of Merit for Macrouniformity," Image Quality and System Performance X, SPIE Vol. 8653, P. D. Burns and S. Triantaphillidou, Eds. San Francisco, CA, 3-7 February 2013.

[3] W. Wang, G. Overall, T. Riggs, R. Silveston-Keith, J. Whitney, G. T. C. Chiu, and J. P. Allebach, "Figure of Merit for Macrouniformity Based on Image Quality Ruler Evaluation and Machine Learning Framework," Image Quality and System Performance X, SPIE Vol. 8653, P. D. Burns and S. Triantaphillidou, Eds. San Francisco, CA, 3-7 February 2013.

[4] M. Q. Nguyen, S. Astling, R. Jessome, E. Maggard, T. Nelson, M. Q. Shaw, and J. P. Allebach, "Perceptual Metrics and Visualization Tools for Evaluation of Page Uniformity," Image Quality and System Performance XI, SPIE Vol. 9016, S. Triantaphillidou and M.-C. Larabi, Eds. San Francisco, CA, 3-5 February 2014.

[5] M. Q. Nguyen and J. P. Allebach, "Controlling Misses and False Alarms in a Machine Learning Framework," Image Quality and System Performance XII, SPIE Vol. 9396, M.-C. Larabi and S. Triantaphillidou, Eds. San Francisco, CA, 8- 12 February 2015.

[6] Weibao Wang, Y. Guo, and J. P. Allebach, "Image Quality Evaluation Using Image Quality Ruler and Graphical Model," Proceedings of ICIP-2015 IEEE International Conference on Image Processing, Quebec City, Canada, 27-30 September 2015

[7] S. Hu, H. Nachlieli, D. Shaked, S. Shiffman, and J. P. Allebach, "Color- Dependent Banding Characterization and Simulation on Natural Document Images," Color Imaging XVII: Displaying, Processing, Hardcopy, and Applications, SPIE Vol. 8292, R. Eschbach, G. Marcu, and A. Rizzi, Eds., San Francisco, CA, 23-26 January 2012.

[8] X. Jing, H. Nachlieli, D. Shaked, S. Shiffman, and J. P. Allebach, "Masking Mediated Print Defect Visibility Predictor," Image Quality and System Performance IX, SPIE Vol. 8293, F. Gaykema and P. D. Burns, Eds, San Francisco, CA, 23-26 January 2012.

[9] J. Zhang, H. Nachlieli, D. Shaked, S. Shiffman, and J. P. Allebach, "Psychophysical Evaluation of Banding Visibility in the Presence of Print Content," Image Quality and System Performance IX, SPIE Vol. 8293, F. Gaykema and P. D. Burns, Eds, San Francisco, CA, 23-26 January 2012.

[10] J. Zhang, S. Astling, R. Jessome, E. Maggard, T. Nelson, M. Q. Shaw, and J. P. Allebach, "Assessment of Presence of Isolated Periodic and Aperiodic Bands in Laser Electrophotographic Printer Output," Image Quality and System Performance X, SPIE Vol. 8653, P. D. Burns and S. Triantaphillidou, Eds. San Francisco, CA, 3-7 February 2013.

[11] J. Zhang and J. P. Allebach, "Estimation of Repetitive Interval of Periodic Bands in Laser Electrophotographic Printer Output," Image Quality and System Performance XII, SPIE Vol. 9396, M.-C. Larabi and S. Triantaphillidou, Eds. San Francisco, CA, 8-12 February 2015.

[12] X. Jing, S. Astling, R. Jessome, E. Maggard, T. Nelson, M. Shaw, and J. P. Allebach, "Electrophotographic Ghosting Detection and Evaluation," Proceedings of NIP-31 IS&Ts 2015 Conference on Digital Fabrication and Digital Printing, Portland, OR, 27 September 1 October 2015.

[13] J. Wang, T. Nelson, R. Jessome, S. Astling, E. Maggard, M. Shaw, and J. Allebach, "Local Defect Detection and Print Quality Assessment," International Congress of Imaging Science (ICIS), Tel Aviv, Israel, 12-14 May 2014.

[14] J. Wang, T. Nelson, R. Jessome, S. Astling, E. Maggard, M. Q. Shaw, and J. P. Allebach, "Local Defect Detection and Print Quality Assessment," Image Quality and System Performance XIII (Part of IS&T Electronic Imaging 2016), R. Jenkin and M.-C. Larabi, Eds. San Francisco, CA, 14-18 February 2016.

[15] N. Yan, E. Maggard, R. Fothergill, R. J. Jessome, and J. P. Allebach, "Autonomous Detection of ISO Fade Point with Color Laser Printers," Image Quality and System Performance XII, SPIE Vol. 9396, M.-C. Larabi and S. Triantaphillidou, Eds. San Francisco, CA, 8-12 February 2015.

[16] Y. Ju, E. Maggard, R. J. Jessome, and J. P. Allebach, "Autonomous Detection of Text Fade Point with Color Laser Printers," Image Quality and System Performance XII, SPIE Vol. 9396, M.-C. Larabi and S. Triantaphillidou, Eds. San Francisco, CA, 8-12 February 2015.

[17] W. Wang, P. Bauer, J. K. Wagner, and J. P. Allebach, "MFP Scanner Diagnostics Using Self-Printed Target to Measure the Modulation Transfer Function," Image Quality and System Performance XI, SPIE Vol. 9016, S. Triantaphillidou and M.- C. Larabi, Eds. San Francisco, CA, 3-5 February 2014.

[18] M. Kim, J. P. Allebach, P. Bauer, and J. K. Wagner, "MFP Scanner Motion Characterization Using Self-Printed Target," Image Quality and System Performance XII, SPIE Vol. 9396, M.-C. Larabi and S. Triantaphillidou, Eds. San Francisco, CA, 8-12 February 2015.

[19] H. S. Rosario, E. Saber, W. Wu, and K. Chandu, "Streak detection in mottled and noisy images," Journal of Electronic Imaging 16(4), 043005 (2007).

## Author Biography

*Daulet Kenzhebalin received his BS in computer engineering from Purdue University (2015). Currently he is pursuing PhD in electrical engineering at Purdue University. His primary area of research has been image processing.*

*Xing Liu received her BS in physics from the University of Science and Technology of China in 2008, and her MS in physics from Purdue University in 2011. She is working toward the Ph.D. degree in image processing at Purdue University. She has been working as a visiting researcher at Yonsei University, Seoul, Korea focusing on pattern recognition since Sep. 2015. Her main research interests include image processing, computer vision and their applications in intelligence systems.*

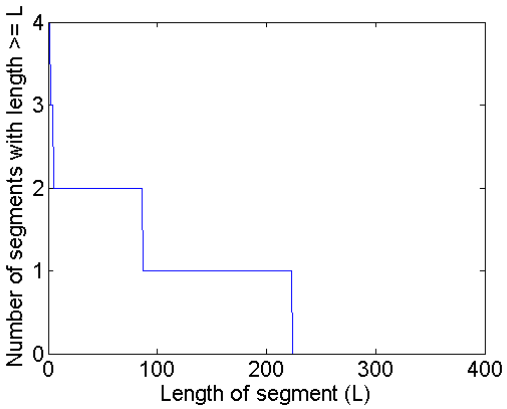
*Ni Yan received her BS in Electrical Engineering from Beijing University of Posts and Telecommunications in 2012, Beijing, China. Since then she is pursuing her PhD degree in Purdue University, U.S. Her research is focused on Imaging Processing and Machine Learning. Her advisor is Jan P. Allebach.*

*Jerry Wagner received his BS and MS in Electrical Engineering from Pennsylvania State University (1978 and 1980) and his MS in Computer Science from Rochester Institute of Technology (1992). He has worked in the Research Division at Eastman Kodak Company, and he has been employed in product development by HP Inc. for the last 16 years. His work has focused on image processing for scanners and multi-function peripherals.*

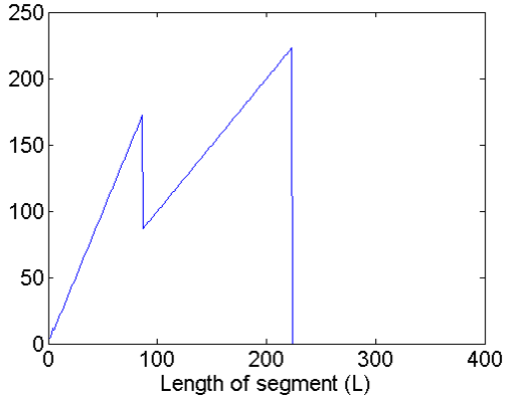
*Jan P. Allebach is Hewlett-Packard Distinguished Professor of Electrical and Computer Engineering at Purdue University. Allebach is a Fellow of the IEEE, the National Academy of Inventors, the Society for Imaging Science and Technology (IS&T), and SPIE. He was named Electronic Imaging Scientist of the 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 and the IS&T/OSA Edwin Land Medal, and is a member of the National Academy of Engineering.*

| Segment # | Length |
|---|---|
| 1 | 1 |
| 2 | 4 |
| 3 | 86 |
| 4 | 223 |

(a) Table of segments in one columnstrip




(b) Number of segments with length $\geq$ L as a function of L




(c) Curve on Fig. 6b multiplied by its length

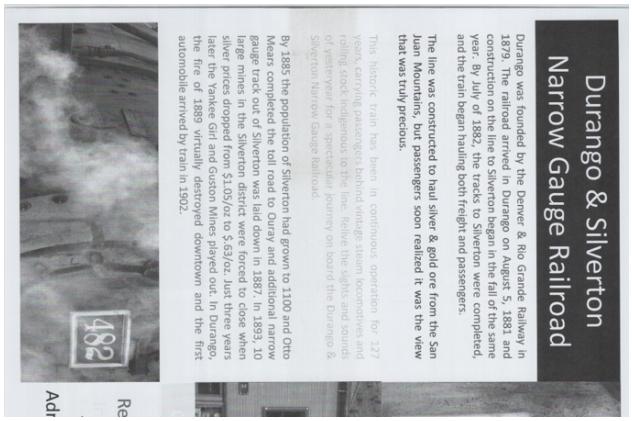
***Figure 6.*** *PPF feature*

(a) Original image

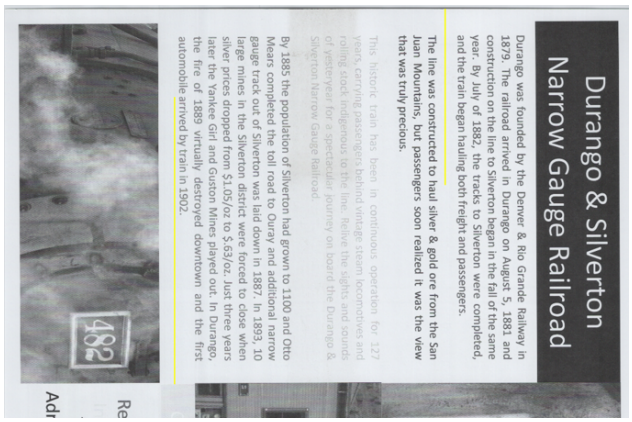
(b) Image with groundtruth superimposed

(c) Image with detections superimposed

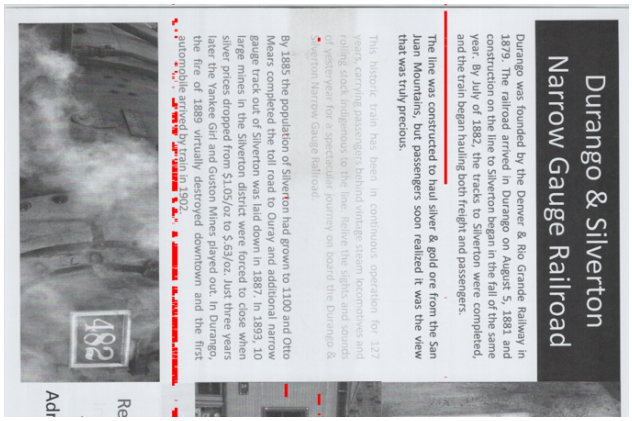
***Figure 7.*** *Example of algorithm detection*




(a) Original image

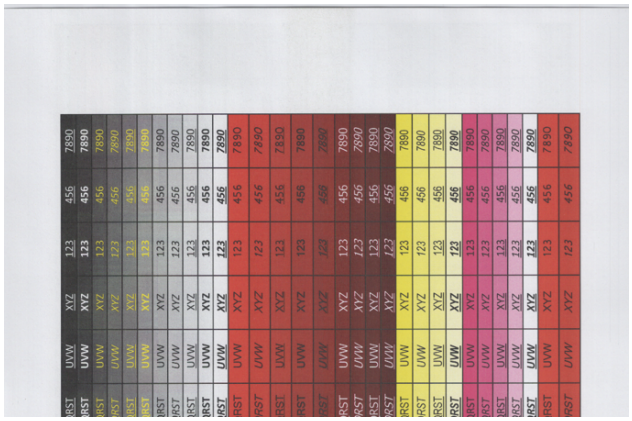



(b) Image with groundtruth superimposed

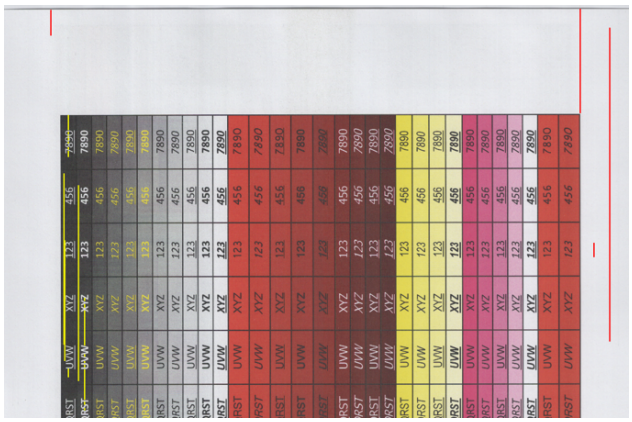



(c) Image with detections superimposed

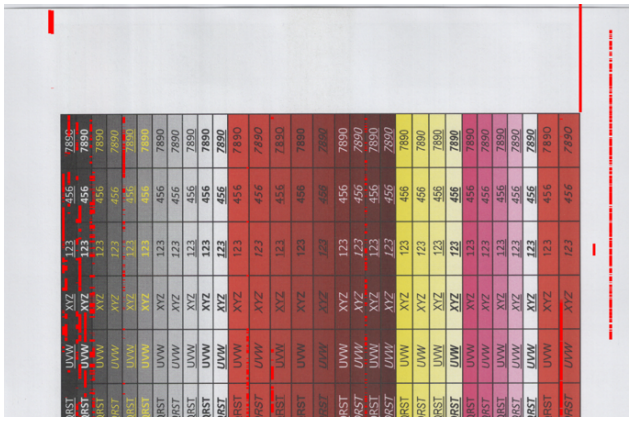
***Figure 8.*** *Example of algorithm detection*



(a) Original image



(b) Image with groundtruth superimposed



(c) Image with detections superimposed

***Figure 9.*** *Example of algorithm detection*