https://doi.org/10.2352/ISSN.2470-1173.2019.10.IQSP-302

Cost-function-based repetitive interval estimation method with synthetic missing bands for periodic bands in electrophotographic printer

Wan-Eih Huang [a], Eric Maggard [b], Renee Jessome [b], Yousun Bang [c], Minki Cho [c], and 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 Printing Korea Co., Ltd.**

Abstract

The repetitive interval is a very crucial feature of bands in print quality assessment, because any irregularity on the surface of a rotating component localized in the circumference will incur repetitive defects on the output of printer [1] [2] [3]. Hence, the repetitive interval can help us diagnose the issues. In previous work, a cost function method provides a robust algorithm to predict the repetitive interval on less noisy samples. However, if the samples contain more aperiodic bands and noise, the estimation will become a challenge. Moreover, the missing periodic bands will decrease the probability of correct prediction. In this paper, we proposes a novel cost-function-based repetitive interval estimation method for periodic bands. By adding synthetic missing bands, we re-evaluate the cost function values to check whether it has a better result. We also show the improvement of accuracy on the print samples with our proposed algorithm.[1]

1. Introduction

Electrophotographic printers have been widely used in the world. There are many print quality (PQ) issues shown in different types of defects, like bands, streaks, and gray spots. Since the electrophotographic process involves multiple delicate components, different appearance of a certain defect might be caused by different components. Hence, the intent of this work is to extract the features of defects so that it could help the diagnosis of the failing components.

One of the most common printing defects in the electrophotographic process is bands, which is the one we want to address in this paper. It occurs along the scan direction and repeats along the process direction. There are some related works analyzed the problem of halftone banding [4] [5] [6]. In addition, some works addressed isolated large pitch bands [7] [8] [9]. However, we focus on sharp roller bands in this work. Figure 1 shows an example of this kind of bands defect.

Our bands detection is based on the work of Zhang et al. [10]. Some fixed threshold values from observations are used to identify the bands. In this work, we want to explore a new method to identify the bands using machine learning method, logistic regression, to classify the potential defects.

Moreover, the repetitive interval of periodic bands is a very important feature to diagnose the root cause components. There are a couple of methods to deal with it. One is the histogram method [10], using the histogram of the intervals between neighboring bands. However, if there are some aperiodic bands or noise between two periodic bands, the correct interval will not be chosen. Another method is the cost function method [3], with an exhaustive search for the best solution by evaluating the cost function value. Nevertheless, if the samples are more noisy or corrupted, this method is not able to estimate the repetitive interval correctly. For example, there may be multiple equally spaced bands sequences, and some of the true periodic bands may be missing. Therefore, we introduce synthetic missing bands in our work to improve the accuracy. We will discuss more details in later sections.

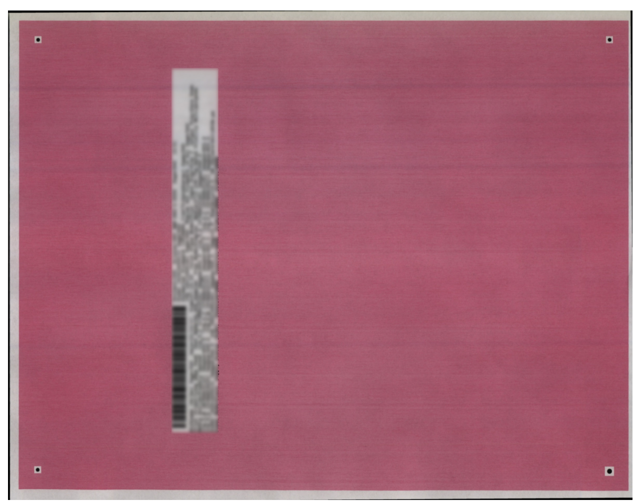


Figure 1: Example of bands defect.

2. Methodology

In this work, the pipeline we used can be divided into two parts: bands detection and repetitive bands analysis shown in Fig. 2. Our algorithms are specifically tailored to the test page shown in Fig. 1

2.1 Bands detection

In order to extract the features of bands, we need to identify the bands first. The steps in this part include pre-processing, bands profile extraction, and bands identification. We follow previous work in pre-processing and bands profile extraction. The details are described in [10].

[1] Research supported by HP Inc., Boise, ID 83714

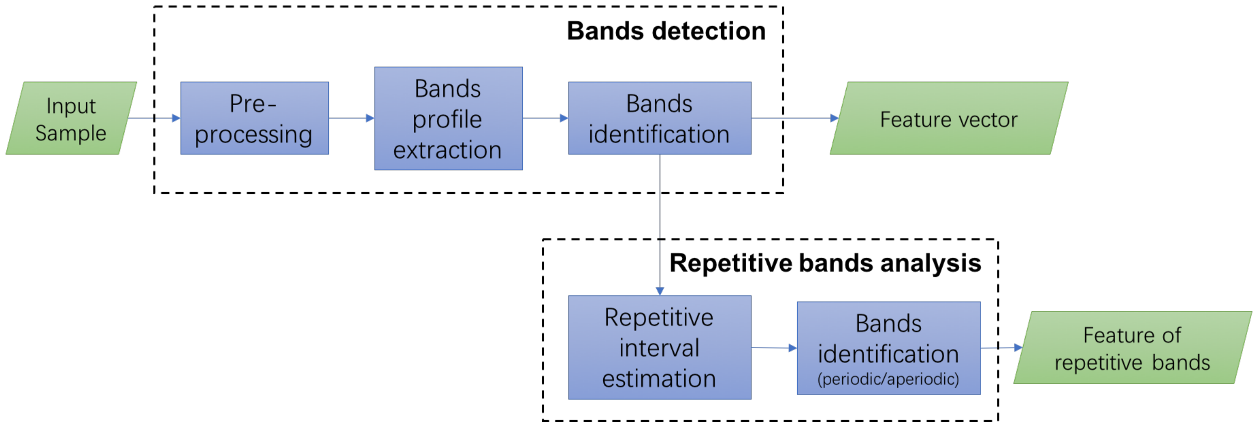


Figure 2: Overall pipeline of proposed algorithm.

Pre-processing

First, we de-screen the input sample to remove halftone patterns and then mask fiducial dots, the non-printable area, and the bar code which contains confidential information. Because the bands might fade along the scan direction, we partition the sample into three regions, left region, center region, and right region [10] so that we can analyze each region independently.

Bands profile extraction

In the bands profile extraction step, the input is one region of the image from pre-processing. The image is converted from the sRGB color space to the CIE 1931 XYZ color space. Then we use spatial projection of the 2-D image onto the process direction by computing the mean value of each line along the scan direction [10]. Afterwards, we perform color space conversion from CIE XYZ to CIE 1976 L*a*b* on the 1-D projection data. To better distinguish the bands from the background, we subtract the baseline from the 1-D projection data for each channel. The baseline is obtained from the filtered 1-D projection data. Finally, we can compute ΔE defined in Eq. (1) and combine it with the sign of the baseline removed L^* channel projection data as our bands profile. The sign value represents whether the ΔE is lighter than the background or darker than the background.

$$\Delta E = \sqrt{(L^*_{proj} - L^*_{base})^2 + (a^*_{proj} - a^*_{base})^2 + (b^*_{proj} - b^*_{base})^2} \quad (1)$$

Bands identification

After the 1-D bands profile extraction process, the small fluctuations of ΔE are eliminated by the threshold value. We use the mean value of ΔE plus the standard deviation of ΔE as the threshold value. Next, we locate the edges of peaks. Moreover, if there are positive and negative values of ΔE within a peak, we separate it into a light peak and a dark peak. In other words, each peak is either light or dark. After that, we extract features by computing the center, height, width, area, and sharpness (transition width) for each peak.

However, the human vision system is complicated. Some peaks are invisible among the detected peaks. There are two reasons we want to classify visible and invisible potential defects. First, we obtain better data quality. For example, we lower the false alarm rate. Second, if too many invisible potential defects are included in the repetitive interval estimation in the later process, it would harm the performance and accuracy of the estimation of periodic bands.

In this work, we apply logistic regression to build our classification model. The model is a weighting function to predict the probability of the binary output. That is, visible or invisible. We choose three features: height (maximum ΔE of a peak), minimum sharpness of two sides, and width in the logistic regression algorithm. Then , we apply this model to our bands identification algorithm. The detailed results will be shown in Section 3.

2.2 Repetitive bands analysis

The features are extracted after the bands identification process. Hence, we can use the information of each band to predict which band belongs to a set of periodic bands. In our method, the maximum value of ΔE, light/dark, and center position of each band are used in our repetitive interval estimation algorithm.

Repetitive interval estimation

There are two phases in the proposed estimation algorithm: an initial guess and adding synthetic missing bands. We apply the cost function method [3] to make an initial guess and improve the accuracy by adding synthetic missing bands.

Phase 1: Initial guess. Since the repetitive bands are generated from one single component [3], they should look similar. Thus, we compare the strength of light bands and the strength of dark bands. Here, the strength is the maximum value of ΔE. The larger one determines whether our candidate set of repetitive bands will be dark or light. All following process are operated on the center positions of this candidate set of repetitive bands.

First, we want to present the cost function briefly, and summarize the algorithm steps. In order to introduce the cost function method clearly, we start with the notation. N is the total number of bands in the candidate set and the input data is the positions of candidate bands denoted by $\vec{b} = [b_1, b_2, \ldots, b_N]$. Assume there are p periodic bands and define the membership vector $\vec{m}$ to indicate which band belongs to the set of repetitive bands. That is,

$$\vec{m} = [m_1, m_2, \ldots, m_N], \quad m_i = \begin{cases} 1, & periodic \\ 0, & aperiodic \end{cases} \quad i = 1, 2, \ldots, N.$$

There are two variables in the cost function. One is o which is the position of the first periodic band. The other is the repetitive interval Δb. Therefore, the predicted positions of the periodic bands can be expressed as

$$b'_k = o + (k-1)\Delta b, \quad k = 1, \ldots, p.$$

The cost function is defined as the mean square error between the predicted positions of the periodic bands and the true data. It can be represented as shown in Eq. (2).

$$\phi = \frac{1}{p}\sum_{i=1}^{N} m_i \left(o + \Delta b(\sum_{j=1}^{i} m_j - 1) - b_i \right)^2 \tag{2}$$

To find the optimal solution of o and Δb for a given membership vector, we take the first derivatives of the cost function with respect to two variables, o and Δb, respectively. Then, we apply the first order necessary condition (FONC). The closed-form optimal solution for a fixed membership vector $\vec{m}$ can be obtained by solving the linear equations. The solutions are shown in Eq. (3) and Eq. (4).

$$\hat{o}^{(\vec{m})} = \frac{2(2p-1)}{p(p+1)}\sum_{i=1}^{N} m_i b_i - \frac{6}{p(p+1)}\sum_{i=1}^{N}(\sum_{j=1}^{i} m_j - 1) m_i b_i \tag{3}$$

$$\hat{\Delta b}^{(\vec{m})} = \frac{12}{p(p+1)(p-1)}\sum_{i=1}^{N} m_i(\sum_{j=1}^{i} m_j - 1) b_i - \frac{6}{p(p+1)}\sum_{i=1}^{N} m_i b_i \tag{4}$$

However, p and $\vec{m}$ are unknown. In order to find the best solution, this algorithm uses an exhaustive search with these two variables. The flow is described in Fig. 3. The input is the true data $\vec{b}$ and the total number of candidate bands N. The initial value for parameter p is 3. For a given p, we have $\binom{N}{p}$ possible combinations of the membership vector. $M_{possible}$ denotes the set of possible membership vectors. For each membership vector in $M_{possible}$, we compute the optimal position of the first periodic band and the optimal repetitive interval by Eq. (3) and Eq. (4). Then, cost function value can be obtained by these two values for a given membership vector. The relationship is described in Eq. (5).

$$\phi^{(\vec{m})} = \frac{1}{p}\sum_{i=1}^{N} m_i \left(\hat{o}^{(\vec{m})} + \hat{\Delta b}^{(\vec{m})}(\sum_{j=1}^{i} m_j - 1) - b_i \right)^2 \tag{5}$$

After obtaining the cost function values for all $\binom{N}{p}$ possible combinations of the membership vector in $M_{possible}$, the optimal result for the given p is the minimum cost function value ϕ_p. So we save its corresponding estimated repetitive interval $\Delta\hat{b}_p$ and membership vector $\vec{m}_p$. We repeat above process until we finish the computations for $p = 3, 4, \ldots, N$. In the last step, we compute the fitting error defined in Eq. (6), which is the cost function value normalized to its repetitive interval. At the end of this phase, we have the fitting error ε_p and its corresponding repetitive interval and membership vector, for $p = 3, 4, \ldots, N$. How do we choose the best solution from this set of possible solutions? There are two aspects we need to consider. The first is the fitting error. However, we cannot just choose the minimum one. Because when p is small, it is easy to find equally spaced bands. Thus, the fitting error is small. In previous work, the criterion was to choose the maximum p from those with fitting error less than 5% [3].

$$\varepsilon_p = \frac{1}{\Delta\hat{b}_p}\sqrt{\phi_p} \tag{6}$$

Since our samples are more noisy, there is a larger probability to find multiple equally spaced band sequences with different repetitive intervals. In addition, missing periodic bands might affect the estimation because the criterion tends to choose the solution with maximum p. In our previous work, we only determined a single repetitive interval. That is, we selected from the candidate sets of periodic bands with fitting error less than 5% that set with the maximum number p of periodic bands. To solve this problem, we introduce synthetic missing bands and re-evaluate the cost function value to find the best solution. To begin with, we choose candidate sets of periodic bands with fitting error less than 5% as our candidates for Phase 2.

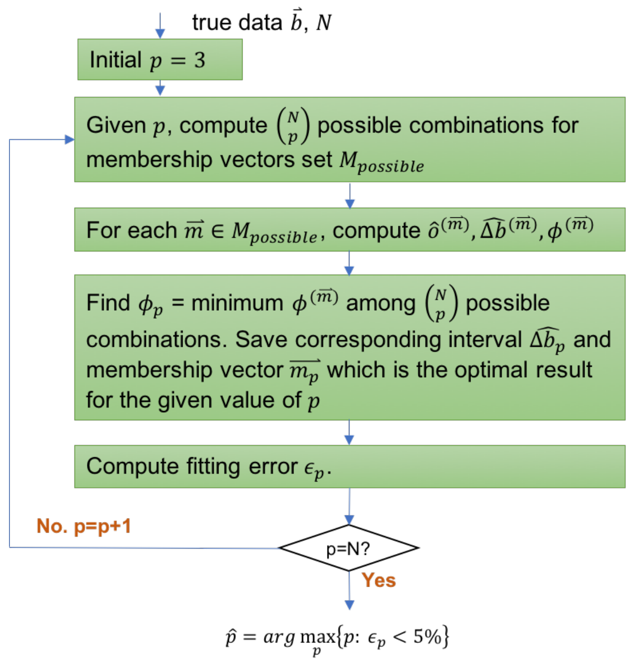


Figure 3: Cost function estimation algorithm.

Phase 2: Adding synthetic missing bands. For each candidate set, we check whether there is an existing band at twice the interval away from either end of the set of periodic bands. If no such band exists, we check the next candidate set. On the contrary, if such a band at twice the interval does exist on either end of the candidate set of bands, we add that band and a new synthetic band that is equally spaced between the band at the end of the set and the new band to update this candidate set of periodic bands. Note that it is possible that additional bands are located at the repetitive interval, beyond the new band that is added to the candidate set. In this case, these bands will also be added to the candidate set. In fact, we may end up merging two candidate sets of periodic bands. After that, we have a new set of fitting errors corresponding to our updated sets of periodic bands. If the smallest fitting error is smaller than our best-so-far fitting error, then we update the best-so-far solution. We repeat this process until all candidate sets have been checked. Finally, we apply the same criterion as in the previous work [3] to select the best one. That is, choose the set with maximum p from those sets for which the fitting error is less than 5%. The overall repetitive interval estimation algorithm is described in Fig. 4.

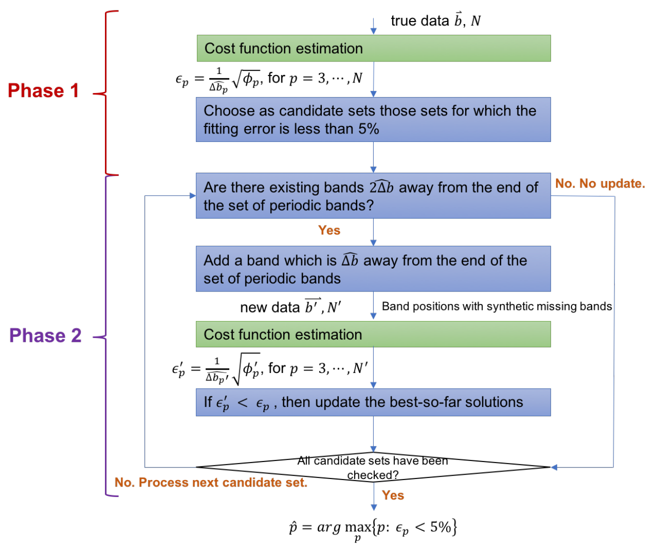


Figure 4: Proposed repetitive interval estimation algorithm.

In our previous work, we also found missing bands. However, we only used one candidate set. We choose the one best result from the first cost function estimation and then continue adding bands one interval away until the set of periodic bands expands to fill the whole page. The purpose for our previous method is to find as many periodic bands as possible. However, here our target is to improve the accuracy. The key idea of our proposed method is searching more possibilities and using less synthetic data. We use more candidates from Phase 1 initial guess and only add one or two synthetic bands in Phase 2 for each candidate set.

Bands identification for periodic and aperiodic bands

After the estimated repetitive interval is determined, the corresponding membership vector is used to identify the periodic and aperiodic bands. Thus, we can collect the statistic values, like maximum, average strength of periodic bands for our feature vector.

3. Experimental Results

This section shows the experimental results of logistic regression and applying our algorithm on the test pages.

Maximum ΔE, minimum sharpness of two sides, and width are our selected features in the logistic regression algorithm. We have total of 18 sample pages and 1693 labeled bands on those sample pages. We apply K-fold cross validation; K is 9 in this work. First, we partition the data into 9 groups randomly; 8 groups are used for training and 1 group is used for testing. We repeat for all 9 groups. The result is shown in Table 1. And then we reshuffle the data 100 times. From Fig. 5, we can see the accuracy is very stable. The average accuracy is 93.4%.

In this project, the test pages are softcopies of constant tone printed from a color laser electrophotographic printer and scanned at 600 dpi. We are only interested in the smooth area since the bands defect in smooth areas is more obvious for our perception [11].

Figure 6 is an example of detection result. The yellow lines separate the three regions. The blue lines are the projection data in signed ΔE. The black lines are the threshold values we use to find the peaks. The red bars are periodic bands; and the green bars are aperiodic bands.

Figure 6a and Figure 6b are the same test page, but estimated by different methods. Usually, the bands defect is most obvious in the center part. Thus, we check the center part only. In Figure 6a, the estimated repetitive interval is 12.2 mm by the cost function method without adding synthetic missing bands. However, the estimated repetitive interval is 33.62 mm by our proposed algorithm, which is the cost function method with added synthetic missing bands. The blue bar shown in Figure 6b is the position where we add the synthetic missing band. The ground truth is 34 mm. Similarly, Fig. 7 is another example for the two estimation methods applied on the same test page. Cost function method predicted the repetitive interval to be 42.64 mm on this sample page. But the repetitive interval estimated by our proposed method is 33.66 mm. The ground truth is the same: 34 mm. Therefore, our method can estimate the repetitive interval better than the previous method on these test pages.

The total number of test pages with obvious periodic bands we have in this project is 15. We apply our proposed method to these test pages and the results are shown in Table 2.

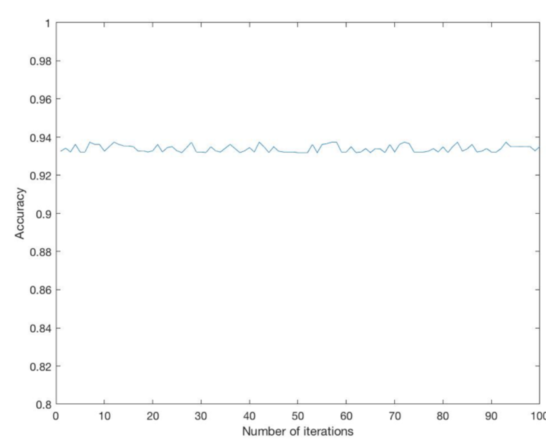


Figure 5: 9-fold cross validation when the data is reshuffled 100 times.

4. Conclusion

In this paper, we build a classification model by logistic regression to determine whether the potential defects are visible or invisible. The average result achieves 93.4% accuracy. In addition, we proposed a new cost-function-based repetitive interval estimation method. We re-evaluate the cost function values by combining true data with synthetic missing bands to improve the accuracy on noisy and corrupted test sample pages.

References

[1] W. Jang and J. P. Allebach, "Simulation of print quality defects," *Journal of Imaging Science and Technology*, vol. 49, no. 1, pp. 1–18, 2005.

[2] R. Kumontoy, K. Low, M. Ortiz, C. Kim, P. Choe, S. Leman, K. Oldenburger, M. Lehto, X. Lehto, H. Santos-Villalobos, H. Park, and J. Allebach, "Web-based diagnosis tool for customers to self-solve print quality issues," *Journal of Imaging Science and Technology*, vol. 54, no. 4, pp. 40503–1, 2010.

[3] J. Zhang and J. P. Allebach, "Estimation of repetitive interval of periodic bands in laser electrophotographic printer output," in *Image Quality and System Performance XII, SPIE*, vol. 9396, 2015.

[4] S. Hu, H. Nachlieli, D. Shaked, S. Shiffman, and J. P. Allebach, "Color-dependent banding characterization and simulation on natural document images," in *Color Imaging XVII: Displaying, Processing, Hardcopy, and Applications*, vol. 8292, p. 82920W, International Society for Optics and Photonics, 2012.

[5] X. Jing, H. Nachlieli, D. Shaked, S. Shiffman, and J. P. Allebach, "Masking mediated print defect visibility predictor," in *Image Quality and System Performance IX*, vol. 8293, p. 82930R, International Society for Optics and Photonics, 2012.

[6] A. H. Eid, M. N. Ahmed, B. E. Cooper, and E. E. Rippetoe, "Characterization of electrophotographic print artifacts: Banding, jitter, and ghosting," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1313–1326, 2011.

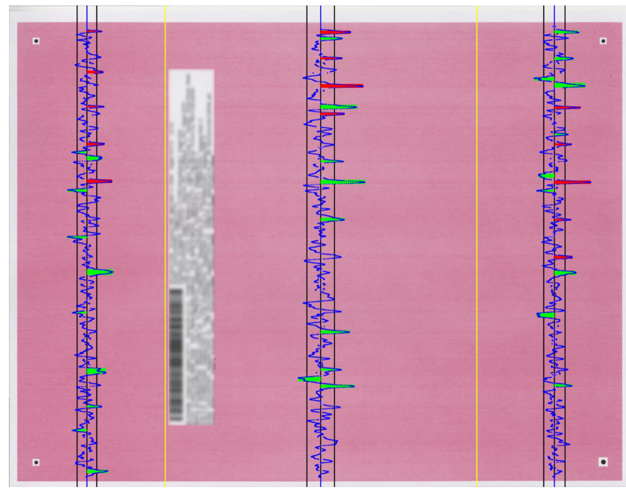
[7] R. Rasmussen, E. N. Dalal, and K. Hoffman, "Measurement of macro-uniformity: Streaks, bands, mottle and chromatic variations," in *PICS*, pp. 90–95, 2001.

[8] D. R. Rasmussen, K. D. Donohue, Y. S. Ng, W. C. Kress, F. Gaykema, and S. Zoltner, "Iso 19751 macro-uniformity," in *Image Quality and System Performance III*, vol. 6059, p. 60590K, International Society for Optics and Photonics, 2006.

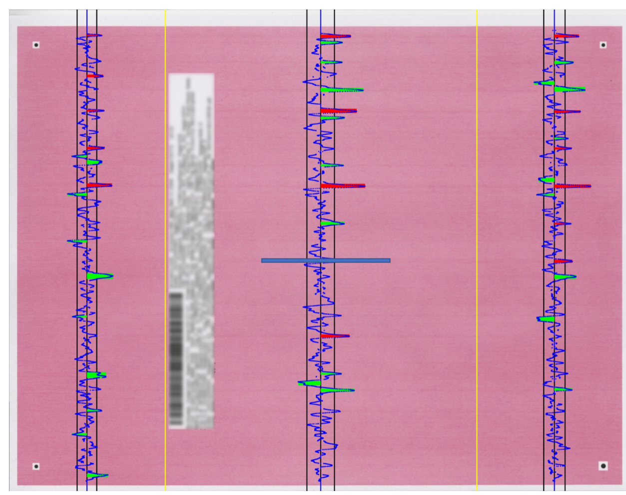
[9] D. R. Rasmussen, "Tent-pole spatial defect pooling for prediction of subjective quality assessment of streaks and bands in color printing," *Journal of Electronic Imaging*, vol. 19, no. 1, p. 011017, 2010.

[10] J. Zhang, S. Astling, R. Jessome, E. Maggard, T. Nelson, M. Shaw, and J. P. Allebach, "Assessment of presence of isolated periodic and aperiodic bands in laser electrophotographic printer output," in *Image Quality and System Performance X, SPIE*, vol. 8653, 2013.

[11] J. Zhang, H. Nachlieli, D. Shaked, S. Shiffman, and J. P. Allebach, "Psychophysical evaluation of banding visibility in the presence of print content," in *Image Quality and System Performance IX, SPIE*, vol. 8293, 2012.

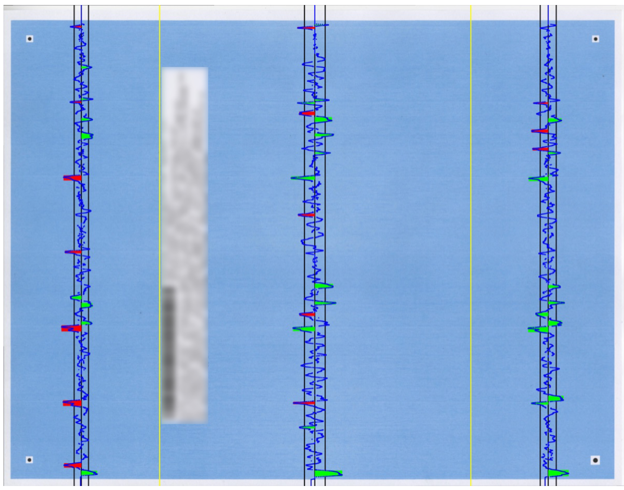


(a)

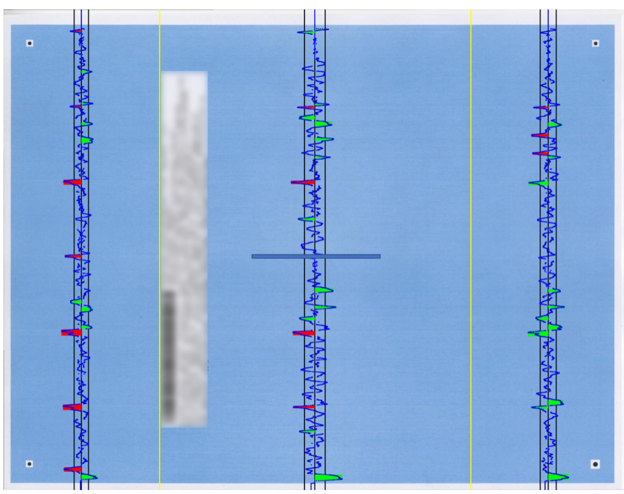


(b)

Figure 6: Comparison of estimated repetitive interval on the same test page. The red bands have been identified as periodic bands. The green bands are aperiodic bands. The ground truth repetitive interval is 34 mm. (a) Estimated repetitive interval is 12.2 mm by cost function method. (b) Estimated repetitive interval is 33.62 mm by cost function method with adding synthetic missing bands. The blue bar is the synthetic missing band



(a)



(b)

Figure 7: Comparison of estimated repetitive interval on the same test page. The red bands have been identified as periodic bands. The green bands are aperiodic bands. The ground truth repetitive interval is 34 mm. (a) Estimated repetitive interval is 42.64 mm by cost function method. (b) Estimated repetitive interval is 33.66 mm by cost function method with adding synthetic missing bands. The blue bar is the synthetic missing band

Fold index	FN	FP	TN	TP	Total	Accuracy	Precision	Specificity	Sensitivity
1	4	7	101	85	197	0.9442	0.9239	0.9352	0.9551
2	10	2	73	78	163	0.9264	0.9750	0.9733	0.8864
3	5	7	88	69	169	0.9290	0.9079	0.9263	0.9324
4	7	6	113	77	203	0.9360	0.9277	0.9496	0.9167
5	11	0	107	60	178	0.9382	1	1	0.8451
6	1	5	115	85	206	0.9709	0.9444	0.9583	0.9884
7	6	9	114	88	217	0.9309	0.9072	0.9268	0.9362
8	10	1	98	70	179	0.9385	0.9859	0.9899	0.8750
9	9	7	110	55	181	0.9116	0.8871	0.9402	0.8594
					Average	0.9362	0.9399	0.9555	0.9105

Table 1: Logistic regression with K-fold cross validation result for classification of visible and invisible potential defects.

Sample ID	Histogram	Cost function	Cost function with adding synthetic missing bands
1	19.81	33.67	33.67
2	32.43	33.91	33.91
3	11.47	33.88	33.88
4	NaN	32.99	32.99
5	28.87	33.76	33.76
6	28.66	31.91	31.91
7	13.38	33.68	33.68
8	30.95	33.97	33.64
9	25.23	33.66	33.68
10	NaN	12.2	33.62
11	33.82	33.83	33.83
12	26.67	33.58	33.58
13	32.72	42.64	33.66
14	9.19	33.97	33.8
15	14.1	33.71	33.71

Table 2: Comparison of repetitive interval estimation result by histogram method, cost function method, and cost function method with adding synthetic missing bands. The ground truth interval for these samples is 34 mm.