An automated image prescreening tool for a printer qualification process

Du-Yong Ng [†] ***and Jan P. Allebach*** [††] ***;*** [†] ***Lexmark International Inc., 740 West New Circle Road Lexington, KY, 40550;*** [††] ***School of Electrical and Computer Engineering Purdue University West Lafayette, IN, 47907-2035***

Abstract

Printer qualification is an important process to ensure that a newly designed printer works accordingly to specifications. Traditionally, image screening is a part of the process that is rather labor intensive. In this work, we develop an automated prescreening tool, along with an image fidelity metric to reduce the workload of expert observers involved in screening the softcopy test images. This tool works for a wide range of image types and content.

Introduction

Electrophotographic (EP) printers are very complex image reproduction devices. When an image is sent to a printer, the data are processed through an imaging pipeline and finally turned into patterns of dots on a piece of paper. For a newly designed printer, it is extremely important to ensure that the entire imaging pipeline is working accordingly to specification. Tests are designed to verify the correctness of the output at various stages of the imaging pipeline. The output can be a digital image or a hardcopy image. The output image is then compared to a reference image to assess the correct functioning of the new printer. Often the reference image is obtained from an earlier product that is known to function correctly. We will refer to the reference image and the output image as the *master image* and *the current image*, respectively, throughout this paper.

In the process of developing a follow-on printer model or an upgrade to an existing model, many changes may have been made to the image rendering pipeline. The fonts, the halftone screen, the color tables, and the page settings may all be different. Any one of these changes, as well as numerous others, will result in differences at many pixel locations between the current and master images. Traditionally, the judgement of a human observer has been needed to determine whether or not the current image is a qualitatively acceptable match to the master image. If there is a mismatch, the developers will investigate the causes of the differences so that the problem can be resolved.

The qualification process requires many trained (expert) observers to compare thousands of softcopy and hardcopy master-current image pairs. The developers will only investigate master-current pairs with visually significant differences that the expert observers have identified. This is a very time-consuming and costly process. Generally, only a fraction of the existing tens of thousands of master images are tested before a product is rolled out due to relatively short development time. It is highly desirable to have an automated tool that will offload part of the work of the expert observers while increasing the number of images that can be tested before the product is released. This is feasible particularly for cases in which softcopy master-current image pairs are compared. In this paper, we will focus on the digital master-current pairs in the qualification process. Consequently, we will exclude print quality defects such as banding, color plane misregistration, repetitive marks, and streaks, which are related to the print mechanism and marking technology in the print engine [1].

Image Fidelity

A current image may be rendered in an orientation different from that of its master, i.e portrait versus landscape or vice versa. This will be the case when the master page is generated using a product that prints pages in an orientation that is different from the current product. The content of the current image may also be spatially shifted slightly from that of the master image. An expert observer will ignore the difference in orientation, minor spatial shifts in content, and error pixels that are sporadically scattered around text, synthetic graphics, and halftone graphics when comparing the image pairs. To mimic the judgement of an observer on the differences between two images, a human visual system model-based metric is needed.

The topic of image quality has been extensively investigated in the literature [2, 3, 4, 5], and image quality is certainly an important factor in judging the success of a new printer. However, the goal of our application is to develop a tool that judges image fidelity instead of image quality. Since there is a rather weak relation between image fidelity and image quality [6], we will only focus on image fidelity from this point onwards.

Many metrics have been introduced over the years to gauge image fidelity. These metrics can be classified into two different categories. Metrics in the first category do not take into consideration the spatial interaction of pixels. Root mean square error, peak signal to noise ratio (PSNR), $\Delta E_{a^*b^*}$, and Kullback-Leibler gain [7] are examples of metrics that fall into this category. While being very simple and appealing computationally, these metrics do not account for spatial processing taking place in the human visual system. The difference between the current and master image pair is inevitably averaged over all the pixels in the images. Consequently, these metrics will fail to distinguish an image pair with a small but visually significant cluster of pixels in error from an image pair with the same number of pixels in error but with these pixels scattered uniformly throughout the image. The latter case may not be noticeable at all even if the magnitude of the average error is the same for both cases. On the other hand, human visual system model-based fidelity predictors such as Daly's visible differences predictor [8], Taylor's image fidelity assessor [9] and Wu's [10] color image fidelity assessor belong to the second category. These algorithms account for several different aspects of information processing in the human visual system, including spatial processing. Taylor's [9] and Wu's [10] results were also

validated with psychophysical experiments. These models are effective; but they are also computationally expensive. The output of these models is a probability map instead of a single number that is much more suitable for our application. Furthermore, these models were not tested with very high resolution images. It is not clear how scalable they are nor how well they will work with the high spatial frequency content that is crucial for our application. The structural similarity image metric (SSIM) also takes spatial interaction of pixels into account but does not apply any human visual system model directly [11]. Instead, the SSIM applies the first and second order statistics of the luminance channel information in a local neighborhood to compare images. The overall image quality is evaluated using the mean of the SSIMs (MSSIM) in all the local neighborhoods. It is not evident in this metric if the size of the local neighborhood should vary accordingly to the resolution of the image pairs. Luminance information alone also may not be sufficient to pick up differences in hue and chroma. In addition, the MSSIM also suffers from the same averaging effect as the metrics in the first category. Weight assignment is not obvious if a weighted average model is used.

In this paper, we will present a fidelity metric that can be trained to mimic the decisions of expert observers based on examination of a set of master-current pairs accompanied by observers' decisions. Due to the constraint on processing speed, this metric is designed only to screen out images with visually insignificant, or at the other extreme, highly objectionable differences between the master and the current images.

Prescreening tool

The automated prescreening tool assesses the fidelity of a current image with respect to the corresponding master image. It will categorize the assessment into three categories: passed, failed, and further evaluation required. Both the expert observers and the developers will never have to examine the image pairs rated as ''passed", since the differences are visually insignificant. For the image pairs flagged as 'failed', the differences are highly objectionable. These images will go directly to the developers to investigate the causes for the differences. Image pairs that fall in between the 'passed' and the 'failed' category are classified as 'further evaluation required'. The expert observers will screen these image pairs manually, and categorize the image pairs into the 'passed' and 'failed' groups based on the visual significance of the differences. The developers will study only the failed cases. With this tool, the expert observers need only focus on the master-current pairs that are flagged as 'further evaluation required'.

In this paper, we will assume that the master-current image pairs have been preprocessed and prepared for comparison. The current image $\mathbf{C}$ is assumed to have the same orientation as the master image $\mathbf{M}$, and the spatial shift between the entire content of the master image and that of the current image is assumed to have been corrected. The corrected current image is denoted as $\widetilde{\mathbf{C}}$. The term 'current image' now will be understood to mean the corrected current image. In addition, the size, resolution, and image type (color, indexed color, grayscale, or bilevel) of the current-master image pairs are assumed to be available during the comparison.

In the following subsection, we will present a fidelity metric we have developed to compare these images. This metric produces a single number to indicate the degree of difference between the master and the corresponding current image. This number is further thresholded to classify the images into the three categories mentioned earlier. This fidelity metric will be referred to as the *error metric* from this point onwards.

In the rest of this paper, we will represent vectors and images using boldface lowercase letters and boldface uppercase letters, respectively. The superscript $(\bullet)^T$ is used to indicate the transpose of a vector or a matrix.

Error metric

Using the pixels of the master image as the reference, if the color of a pixel in the current image differs from that of the pixel at the same spatial location in the master image, the difference can be interpreted as an error. Thus, for each master-current image pair, a 2D error map $\mathbf{E}$ can be obtained using the following equation

$$\mathbf{E}(i,j) = \begin{cases} 1, & \left\| g(\mathbf{M}(i,j)) - g(\widetilde{\mathbf{C}}(i,j)) \right\|_2 \geq T \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Here $g(\bullet)$ can be an operator that transforms the color value to a more perceptually uniform color space, such as the CIE $L^*a^*b^*$ color space. The threshold T may be selected so that only perceptible color differences are considered to be errors. Kuehni *et al* reported that on average, the 50% level of acceptability for small color differences in the CIE $L^*a^*b^*$ space is about one unit of ΔE [12]. To be conservative, we choose $T = 0.6\,\Delta E$.

While computing the error map, the number of pixels in error is also counted. If this number is smaller than a threshold, these differences may be ignored and no further processing is required. In this case, the prescreening tool passes the image pair. However, caution must be exercised given the fact that this threshold will depend on the resolution of the images. If the differences are not negligible, then the error map is clustered based on spatial proximity of the pixels in error so that we can analyze the error of each cluster locally. From this point onward, $\mathbf{E}$ is assumed to contain clusters of pixels whose corresponding locations in the current image are the pixels in error. These error pixels are labeled according to the clusters to which they belong.

Bilevel images

Let the pixel at the spatial location (i,j) with $\mathbf{E}(i,j) = k$ be the l^{th} pixel of the k^{th} cluster in the error map. The corresponding pixel in the master image, $\mathbf{M}(i',j')$ is first located. For notational simplicity, we assume (i,j) to be the same as (i',j'). Next, an $S \times S$ contrast sensitivity function (CSF) filter $\mathbf{F}_{CSF}$ is centered at $\mathbf{M}(i,j)$. This linear filter is used to account for part of the human visual system spatial response. The term CSF refers to both the luminance [13, 14] and chrominance [15] models throughout this paper.

The parameter S is an odd integer that varies with image resolution; and its choice will be discussed later. The weighted average $\alpha_{kl}^{\mathbf{M}}$ at $\mathbf{M}(i,j)$ within the CSF filter window is computed as

$$\alpha_{kl}^{\mathbf{M}} = \sum_{\bar{i}=-\frac{S-1}{2}}^{\frac{S-1}{2}} \sum_{\bar{j}=-\frac{S-1}{2}}^{\frac{S-1}{2}} \mathbf{M}(\bar{i},\bar{j})\mathbf{F}_{CSF}(\bar{i},\bar{j}), \quad (2)$$

where $(\bar{i},\bar{j}) \in S \times S$ window centered at (i,j). This process is repeated for every $\{(i,j) : \mathbf{E}(i,j) = k\}$. The statistical mean $\overline{\alpha_k^{\mathbf{M}}}$

for the entire k^{th} cluster is given by

$$\overline{\alpha_k^{\mathbf{M}}} = \frac{1}{N_k}\sum_{l=1}^{N_k}\alpha_{kl}^{\mathbf{M}}, \quad (3)$$

where N_k is the total number of pixels for which $\mathbf{E}(i,j)=k$.

The value $\overline{\alpha_k^{\mathbf{M}}}$ is converted to CIE $L^*a^*b^*$ tristimulus coordinates and recorded. The same process is repeated for the current image $\widetilde{\mathbf{C}}$. Let $g(\bullet)$ be the transformation from *sRGB* color space to the CIE $L^*a^*b^*$ color space. Then, the color difference $\{\Delta E_{a^*b^*}\}_k^{CSF}$ between $\overline{\alpha_k^{\mathbf{M}}}$ and $\overline{\alpha_k^{\widetilde{\mathbf{C}}}}$ can be calculated as

$$\{\Delta E_{a^*b^*}\}_k^{CSF} = \left\| g(\overline{\alpha_k^{\mathbf{M}}}) - g(\overline{\alpha_k^{\widetilde{\mathbf{C}}}}) \right\|_2. \quad (4)$$

This color difference is computed for every cluster in $\mathbf{E}$ and the value is recorded. The weighted color difference $\{\Delta E_{a^*b^*}\}^{CSF}$ for the master-current pair using the CSF filter is defined as

$$\{\Delta E_{a^*b^*}\}^{CSF} = \frac{1}{N_{tot}}\sum_k N_k \{\Delta E_{a^*b^*}\}_k^{CSF}. \quad (5)$$

Here the sum is taken over all the clusters in the image; and the parameter N_{tot} is the total number of pixels in error.

Using Eqs. 2, 3, 4, and 5, we filter the master and current image around the pixels in error, compute the average filtered values, and calculate $\Delta E_{a^*b^*}$ only once per cluster to reduce the amount of computation. Since $\alpha_l^{\mathbf{M}}$ and $\alpha_l^{\widetilde{\mathbf{C}}}$ are only calculated for the pixels in error, the resulting $\{\Delta E_{a^*b^*}\}^{CSF}$ will be higher than that yielded by a conventional approach where we filter the entire image pair, find the average pixels values for the pixels in error and compute the $\Delta E_{a^*b^*}$. For purposes of illustration, consider an isolated pixel in error at (i,j) and a neighborhood $\Delta(i,j)$ of (i,j), to which the error at (i,j) has spread through the filtering process. The color differences of the pixels in the neighborhood are bounded by

$$\left| g\Big((\mathbf{M} * \mathbf{F}_{CSF})(i',j')\Big) - g\left((\widetilde{\mathbf{C}} * \mathbf{F}_{CSF})(i',j')\right)\right| \leq \alpha_l^{\mathbf{M}} \quad \forall (i',j') \in \Delta(i,j). \quad (6)$$

Here, the symbol $*$ denotes a 2D convolution. The right side of Eq. 6 forms an upper bound for the average color difference in $\Delta(i,j)$. If there are two or more pixels in error in the same neighborhood, the superposition principle can be applied to arrive at the same conclusion. In fact, the average $\Delta E_{a^*b^*}$ over $\Delta(i,j)$ can be at most as large as $\{\Delta E_{a^*b^*}\}^{CSF}$ when the CSF in Eq. 2 is replaced with an unweighted $S \times S$ filter. So, $\{\Delta E_{a^*b^*}\}^{CSF}$ in Eq. 5 is indeed an upper bound of the $\Delta E_{a^*b^*}$ if the $\Delta E_{a^*b^*}$ is calculated after the entire image pairs are filtered with the CSF filter.

These computations are repeated with the visual acuity filter (VAF), a second but a smaller linear filter which will pick up finer details. The choice of the VAF will be described in detail later. When the VAF is placed over $\mathbf{M}$ and $\widetilde{\mathbf{C}}$, the errors are taken into account only if the entire region within the VAF window of either $\mathbf{M}$ or $\widetilde{\mathbf{C}}$ is completely black or white because these errors are more noticeable. Similar to the case of the CSF, the error $\{\Delta E_{a^*b^*}\}^{VAF}$ picked up by the VAF is given by

$$\{\Delta E_{a^*b^*}\}^{VAF} = \frac{1}{N_{tot}}\sum_k N_k \{\Delta E_{a^*b^*}\}_k^{VAF}. \quad (7)$$

The quantities $\{\Delta E_{a^*b^*}\}^{CSF}$ and the $\{\Delta E_{a^*b^*}\}^{VAF}$ are measures designed to capture objectionable errors at two different scales. These measures are combined using a variable power Minkowski metric as follows

$$\{\Delta E_{a^*b^*}\}^{CSF+VAF} = \left[\left[\{\Delta E_{a^*b^*}\}^{VAF}\right]^{N_p} + \left[\{\Delta E_{a^*b^*}\}^{CSF}\right]^{N_p}\right]^{\frac{1}{N_p}} \quad (8)$$

where

$$N_p = 1 + 2tanh\left(max\Big(\{\Delta E_{a^*b^*}\}^{VAF}, \{\Delta E_{a^*b^*}\}^{CSF}\Big)\right). \quad (9)$$

Since $\{\Delta E_{a^*b^*}\}^{VAF} \geq 0$, $\{\Delta E_{a^*b^*}\}^{CSF} \geq 0$, and $1 \leq N_p \leq 3$, the $\{\Delta E_{a^*b^*}\}^{VAF}$ and the $\{\Delta E_{a^*b^*}\}^{CSF}$ combine additively when both of them are small. If one of them is much greater than the other, the greater measure will dominate the overall metric. This combination was inspired by Keelan's work in predicting overall image quality from individual attributes [3].

The overall error metric ε that measures the fidelity of the current image with respect to the master image is obtained by adjusting $\Delta E_{a^*b^*}^{CSF+VAF}$ according to the ratio of the number of pixels N_{tot} in error to the total number of pixels N in $\mathbf{E}$:

$$\varepsilon = \left[\{\Delta E_{a^*b^*}\}^{CSF+VAF}\right]^{N_f}, \quad (10)$$

where $N_f = 1 + \frac{N_{tot}}{N}$. As the size of the region in error grows, ε also increases.

Indexed color images

Indexed color images have three channels and each channel is bilevel. So, halftone images also belong to this category. The error metric ε for this class of images is a straightforward extension from that for the bilevel images. Each of the three channels ($\mathbf{M}_R$, $\mathbf{M}_G$, and $\mathbf{M}_B$) of the master image is first filtered independently using the CSF filter by applying Eq. 2. Let $\alpha_{kl}^{\mathbf{M}} = \left[\alpha_{Rkl}^{\mathbf{M}}, \quad \alpha_{Gkl}^{\mathbf{M}}, \quad \alpha_{Bkl}^{\mathbf{M}}\right]^T$ be a vector containing the filtered outputs of the red (R), green (G), and blue (B) channels, respectively, of the pixel in the master image corresponding to the l^{th} pixel of the k^{th} cluster in the error map. By applying Eq. 3 to the R, G, and B channels independently, the mean color vector $\overline{\alpha_k^{\mathbf{M}}}$ for the k^{th} cluster in the master image can be obtained

$$\overline{\alpha_k^{\mathbf{M}}} = \left[\overline{\alpha_{Rk}^{\mathbf{M}}}, \quad \overline{\alpha_{Gk}^{\mathbf{M}}}, \quad \overline{\alpha_{Bk}^{\mathbf{M}}}\right]^T. \quad (11)$$

The computations are repeated with the current image to produce the vector $\overline{\alpha_k^{\widetilde{\mathbf{C}}}}$. The error $\{\Delta E_{a^*b^*}\}_k^{CSF}$ between the master and current images is simply the Euclidean norm in the CIE $L^*a^*b^*$ space

$$\{\Delta E_{a^*b^*}\}_k^{CSF} = \left\| g(\overline{\alpha_k^{\mathbf{M}}}) - g(\overline{\alpha_k^{\widetilde{\mathbf{C}}}}) \right\|_2, \quad (12)$$

computed according to Eq. 5. The error picked up by the visual acuity filter is obtained in a similar way. In this case, the errors are taken into account only if the entire region within the VAF window of either $\mathbf{M}$ or $\mathbf{C}$ is completely colored for any of the three channels or white. In other words, the VAF only picks up the error under two different circumstances. First, the error will

be picked up if there are no white pixels within the VAF window of the master or the current images. Second, if the pixels within the VAF window of either **M** or **C** are completely white, the error will also be taken into account. Otherwise, the error will not be considered as sufficiently strong and will be ignored. The ε are calculated using Eqs. 8 and 10.

Grayscale images

Unlike bilevel images, grayscale images have 254 intermediate gray levels between white and black. In a given cluster, the magnitude of the perceived error at the l^{th} pixel location $\left|g(\alpha_{kl}^{\mathbf{M}}) - g(\alpha_{kl}^{\widetilde{\mathbf{C}}})\right|$ may vary significantly. If the number of pixels with very large errors is much smaller than those with smaller errors in the same cluster, these very large and localized errors will be attenuated by the averaging effect when Eq. 3 is applied. To reduce this effect, for each pixel in error in each cluster, we first compute the $\Delta E_{a^*b^*}$ between **M** and $\widetilde{\mathbf{C}}$. Pixels with $\Delta E_{a^*b^*}$ greater than a threshold T_c are averaged separately from those below the threshold. This threshold is chosen to be $T_c = TM^2$ where T is the threshold mentioned earlier in the Error Metric section and M^2 is the total area of the VAF in pixel units. A pixel with error less than T units of ΔE will be ignored. The threshold T_c incorporates the spatial summation effect of HVS [16]. The HVS accumulates quanta of light over a certain area. This area is often referred to as the critical area. According to Ricco's Law, within this critical area, the product of the stimulus area and the accumulated luminous energy must be at least as large as the threshold for the detection of the stimulus [16]. A single pixel with error as large as T_c units of ΔE has the same effect as an area of the error in $M \times M$ pixels with T units of ΔE. Individual pixels with error between T and T_c units of ΔE are less significant in terms of visibility, and are thus separated from those with error larger than T_c units of ΔE. For the k^{th} cluster, the total number of pixels $\{N_a\}_k$ with error larger than T_c units of ΔE, and the total number of pixels $\{N_b\}_k$ with error below the threshold are recorded. It is important to note that $N_{tot} = N_a + N_b$, where $N_a = \sum_k \{N_a\}_k$ and $N_b = \sum_k \{N_b\}_k$.

The overall $\Delta E_{a^*b^*}$ above and below the threshold for all the clusters $\{\Delta E_{a^*b^*}\}_a^{CSF}$ and $\{\Delta E_{a^*b^*}\}_b^{CSF}$ are given, respectively, by

$$\{\Delta E_{a^*b^*}\}_a^{CSF} = \frac{1}{N_a}\sum_k \{N_a\}_k \left\{\{\Delta E_{a^*b^*}\}_a^{CSF}\right\}_k, \quad (13)$$

$$\{\Delta E_{a^*b^*}\}_b^{CSF} = \frac{1}{N_b}\sum_k \{N_b\}_k \left\{\{\Delta E_{a^*b^*}\}_b^{CSF}\right\}_k. \quad (14)$$

Similar to Eq. 8, $\{\Delta E_{a^*b^*}\}^{CSF}$ is obtained by combining $\{\Delta E_{a^*b^*}\}_a^{CSF}$ and $\{\Delta E_{a^*b^*}\}_b^{CSF}$ as follows

$$\{\Delta E_{a^*b^*}\}^{CSF} = \left[\left[\{\Delta E_{a^*b^*}\}_a^{CSF}\right]^{N_p^{CSF}} + \left[\{\Delta E_{a^*b^*}\}_b^{CSF}\right]^{N_p^{CSF}}\right]^{\frac{1}{N_p^{CSF}}}, \quad (15)$$

where $N_p^{CSF} = 1 + 2tanh\left(max\left(\{\Delta E_{a^*b^*}\}_a^{CSF}, \{\Delta E_{a^*b^*}\}_b^{CSF}\right)\right)$.

This process is repeated for the VAF. In this case, the errors are taken into account only if the color of the entire region in either **M** or $\widetilde{\mathbf{C}}$ under VAF is uniform. The error $\{\Delta E_{a^*b^*}\}^{VAF}$ is given by

$$\{\Delta E_{a^*b^*}\}^{VAF} = \left[\left[\{\Delta E_{a^*b^*}\}_a^{VAF}\right]^{N_p^{VAF}} + \left[\{\Delta E_{a^*b^*}\}_b^{VAF}\right]^{N_p^{VAF}}\right]^{\frac{1}{N_p^{VAF}}}, \quad (16)$$

where $N_p^{VAF} = 1 + 2tanh\left(max\left(\{\Delta E_{a^*b^*}\}_a^{VAF}, \{\Delta E_{a^*b^*}\}_b^{VAF}\right)\right)$.

The combined error $\{\Delta E_{a^*b^*}\}^{CSF+VAF}$ and the error metric ε are computed using Eqs. 8 and 10.

Full color images

The error metric for full color images is simply an extension from that for the grayscale images. The color difference formulas are modified in exactly the same way as we did when extending the metric model from bilevel images to indexed color images. The $\Delta E_{a^*b^*}$ between **M** and $\widetilde{\mathbf{C}}$ is calculated as in the grayscale case. Pixels with $\Delta E_{a^*b^*}$ greater than a threshold T_c are averaged separately from those below the threshold. The three channels are filtered independently using the CSF and VAF windows. The $\{\Delta E_{a^*b^*}\}^{VAF}$ and $\{\Delta E_{a^*b^*}\}^{CSF}$ errors are obtained using Eqs. 11, 12, and then 5 with the CSF and VAF filters, respectively. The remaining computations are exactly the same as those for grayscale images. It is interesting to note that the same error in a black and white content region for all these image types will produce the same value of ε as that of a bilevel image pair.

Experiment

The prescreening tool was implemented in C language. The image set used in the experiment were made up of bilevel, grayscale, indexed color and full color images. This image set consists of 147 image pairs. The image pairs are letter-size images with one of the 3 resolutions: 300 dpi, 600 dpi, or 1200 dpi. The images contain a wide variety of content such as text, synthetic graphics, natural scenes, people, and color ramps. The types of errors between the master and the current image are also very diverse. Spatial shifts of portions of content, hue shift, incorrect fonts, sizes and cases, missing or extra content, different halftone screens, different shade, incorrect content such as wrong words or characters are among the examples.

For a 600 dpi image, the highest spatial frequency at a viewing distance of one foot is approximately 63 cycles per degree. The contrast sensitivity function (CSF) envelope (approximated using a circularly symmetric envelope) is truncated below this cutoff frequency and sampled in the spatial frequency domain at N equally spaced points. An N-point inverse discrete Fourier transform (IDFT) is taken to obtain the corresponding spatial domain CSF filter. We found that $N = 23$ yields a good CSF spatial window with reasonably small filter coefficients near the boundary. The visual acuity filter is added to pick up errors for the details at the normal visual acuity limit (approximately 1 minute of arc). To speed up the computation, the CSF spatial filter and visual acuity filter (VAF) for 600 dpi images were approximated with simple averaging windows of size 23×23 and 5×5, respectively. The sizes of the windows were scaled accordingly for 300 dpi and 1200 dpi images. At a viewing distance of 12 inches, a pixel of a 600 dpi image subtends slightly less than 0.5 minutes of arc. At this acuity limit, an observer is unable to resolve two to three

pixels that are next to one another. Hence, a 5×5 window with the pixel in error at the center is a reasonable choice for VAF. We have also conducted a separate test to see whether or not the VAF window could be excluded. The tool fails to pick up small (in terms of size) but visually significant errors as expected. This test justifies the need for the VAF in our error metric model.

These image pairs were screened by 6 expert observers and classified into 3 categories: passed (acceptable), further evaluation (require further evaluation) and failed (objectionable). These image pairs are then screened using the prescreening tools with configuration page identification, and time stamp detection options turned on. The PSNR and the MSSIM metrics for the image set is also computed for comparison. To compute the MSSIM value for a current-master image pair, we applied the MSSIM algorithm to the grayscale version of the image pair since this algorithm [11] considers only information in the luminance channel. This algorithm is implemented as described in Wang *et al*'s work [11].

Results

The error metric value ε predicted by the prescreening tool for each master-current image pair is plotted in Fig. 1. One of the image pairs with very severe error (approximately 470) is not shown in the figure since it is the only one with ε greater than 200. The slice of Fig. 1 with ε between 0 and 10 is plotted in Fig. 2 to show the results in the lower range more clearly. The PSNR metric values for the same image set are shown in Fig. 3.

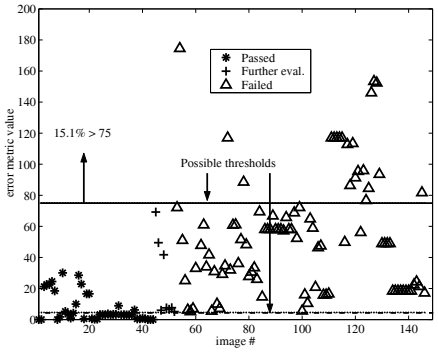

Figure 1. *The error ε for the 147 test image pairs in 0-200 range. The ε of one image pair that exceeds 200 is not shown in the figure.*

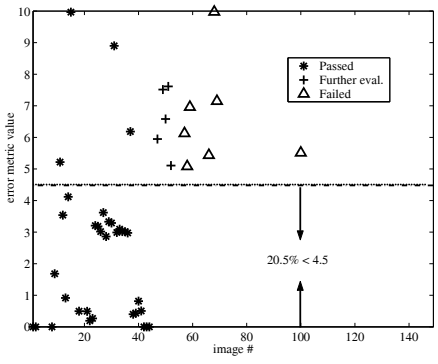

Figure 2. *The error ε for the 147 test image pairs in 0-10 range.*

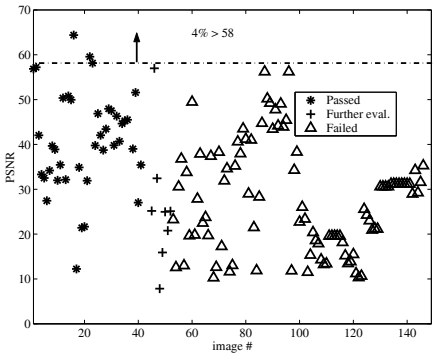

Figure 3. *PSNR values for the 147 test image pairs in 0-70 range. The PSNR values of three image pairs that exceed 70 are not shown in the figure.*

The MSSIM values for the same image set are shown in Figs. 4 and 5.

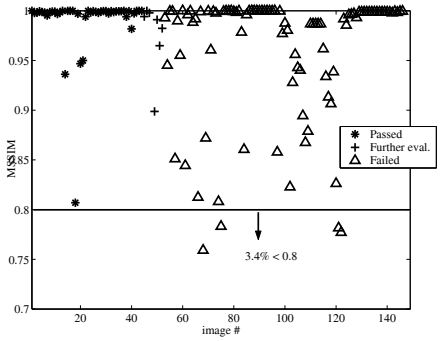

Figure 4. *MSSIM values for the 147 test image pairs. A perfect match yields a value of one.*

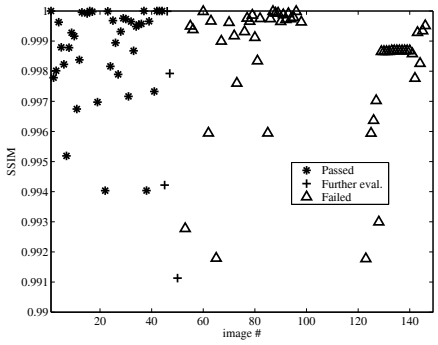

Figure 5. *MSSIM values for the 147 test image pairs in 0.99-1.00 range. A perfect match yields a value of one.*

Discussion

The results in Fig. 1 and 2 indicate that the prescreening tool is capable of screening out as many as 35% of the acceptable and objectionable differences for this image set when the lower and upper thresholds of ε are set to 4.5 and 75 respectively. For the image pairs below the lower threshold and above the upper threshold, the tool produces the same decisions as those of the expert observers. As many as 20% of the image pairs with acceptable

differences can be screened out so that the observers will never have to inspect these image pairs at all. Knowing that approximately 15% of the image pairs have very objectionable errors, the observers will not need to screen these image pairs manually at all. So, the observers only need to concentrate on the remaining 65% of the image pairs to determine if the errors are acceptable or not. When more conservative thresholds (1 for the lower threshold and 100 for the upper threshold) are used for ε, the tool is still able to screen out approximately 19% of the image pairs.

On the other hand, the results for the PSNR metric indicate that this metric is only able to screen out 4%. These are the image pairs with acceptable differences. No reasonable threshold exists for the objectionable case. The results of MSSIM exhibit a trend similar to that of the PSNR. MSSIM values range from zero to one. A perfect match between the current and the master images yields a unity MSSIM. The MSSIM metric only screens out 3.4% of the images. These are the failed image pairs. The PSNR and the MSSIM metrics are obviously not a good error metric for this application.

Conclusion

We have developed an automated prescreening tool, along with an image fidelity metric for a printer qualification process. This tool works for a wide range of image types and content. It is able to reduce the workload of expert observers by a substantial amount.

Acknowledgments

The authors would like to thank Hewlett Packard Company for supporting this research. This work was completed when Du-Yong Ng was a student at Purdue University.

References

[1] W. Jang and J. P. Allebach, "Simulation of Print Quality Defects," *Journal of Imaging Science and Technology*, **49**, pp. 1-18, Jan./Feb. 2005.

[2] P. G. Engeldrum, "Image quality modeling: where are we?," in *Proceedings of the IS&T's PICS Conference*, Savannah, Georgia, pp. 251–255, April 1999.

[3] B. W. Keelan, *Handbook of image quality characterization and prediction*. Marcel Dekker, Inc, New York, NY, pp. 149-168, 2002.

[4] Z. Wang, A. C. Bovik and L. Lu, "Why is image quality assessment so difficult?," in *Proceedings (ICASSP '02) IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 3313-3316, May 13-17, 2002.

[5] J. B. Martens, "Multidimensional modeling of image quality," *Proceedings of the IEEE*, **90**, pp. 133-153, 2002.

[6] D. A. Silverstein and J. E. Farrell, "The relationship between image fidelity and image quality," in *Proceedings IEEE International Conference on Image Processing*, Lausanne, Switzerland, **1**, pp. 881-884, Sept. 16-19, 1996.

[7] J. A. Garcia, J. Fdez-Valdivia, R. Rodriquez-Sanchez, and X. R. Fdez-Vidal, "Performance of the Kullback-Leibler information gain for predicting image fidelity," in *IEEE Proceedings 16th International Conference on Pattern Recognition*, Quebec City, Que., Canada, **3**, pp. 843-848, Aug. 11-15, 2002.

[8] S. Daly, "The visible differences predictor: An algorithm for the assessment of image fidelity," in *Digital Images and Human Vision*, A. B.Watson, Editor, pp. 179-205, MIT Press, Cambridge, MA, 1993.

[9] C. C. Taylor, "Image quality assessment based on human visual system model," Ph.D Dissertation, School of Electrical Engineering, Purdue University, West Lafayette, IN, 1998.

[10] W. Wu, "Two problems in digital color imaging: Colorimetry and image fidelity assessor," Ph.D Dissertation, School of Electrical Engineering, Purdue University, West Lafayette, IN, 2000.

[11] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transaction on Image Processing*, **13**, pp. 600-612, April, 2004.

[12] R. G. Kuehni and R. T. Marcus, "An experiment in visual scaling of small color differences," *Color Research and Application*, **4**, pp. 83-91, 1979.

[13] C. W. Thomas, G. C. Gilmore, and F. L. Royer, "Models of contrast sensitivity in human vision," in *IEEE Transaction on Systems, Man, and Cybernetics*, **23**, pp. 857-864, May/June, 1993.

[14] S. H. Kim and J. P. Allebach, "Impact of HVS models on model-based halftoning," *IEEE Transaction on Image Processing*, **11**, pp. 258-269, March, 2002.

[15] T. J. Flohr, B. W. Kolpatnik, R. Balasubramanian, D. A. Carrara, C. A. Bouman, and J. P. Allebach, "Model-based color image quantization," in *Proc. SPIE Human Vision, Visual Processing, and Digital Display IV*, San Jose, CA, **1913**, pp. 270-281, Jan. 31-Feb. 4, 1993.

[16] S. H. Schwartz, *Visual Perception: A Clinical Orientation*, pp. 44-47, The McGraw-Hill Companies, Inc, 3rd Edition , NJ, 2004.

Author Biography

Du-Yong Ng received his BSEE, MSEE and Ph.D from Purdue University in 1998, 2001 and 2004 respectively. He has been with Lexmark International Inc. since 2004. His research interests include colorimetry, electronic imaging systems, multi-spectral imaging systems and applications, image quality and analysis and color image processing.

Jan P. Allebach received his BSEE from the University of Delaware in 1972 and his Ph.D. from Princeton University in 1976. He was on the faculty at the University of Delaware from 1976 to 1983. Since 1983, he has been at Purdue University where he is Hewlett-Packard Professor of Electrical and Computer Engineering. His current research interests include image rendering, image quality, color imaging and color measurement, and digital publishing. Prof. Allebach is a member of the IEEE Signal Processing (SP) Society, the Society for Imaging Science and Technology (IS&T), and SPIE. He has been especially active with the IEEE SP Society and IS&T. He has served as Distinguished/Visiting Lecturer for both societies, and has served as an officer and on the Board of Directors of both societies. Prof. Allebach is a past Associate Editor for the IEEE Transactions on Signal Processing and the IEEE Transactions on Image Processing. He is presently Editor for the IS&T/SPIE Journal of Electronic Imaging. He received the Senior (best paper) Award from the IEEE Signal Processing Society and the Bowman Award from IS&T. In 2007, he was named Honorary Member of IS&T - the highest award bestowed by IS&T. He is a Fellow of IEEE, IS&T and SPIE.