Mobile Captured Glass Board Image Enhancement

Boddapati Mahesh, Ajoy Mondal, and C.V. Jawahar

Centre for Visual Information Technology,
International Institute of Information Technology, Hyderabad, India
boddapati.mahesh@research.iiit.ac.in,
{ajoy.mondal,jawahar} @iiit.ac.in

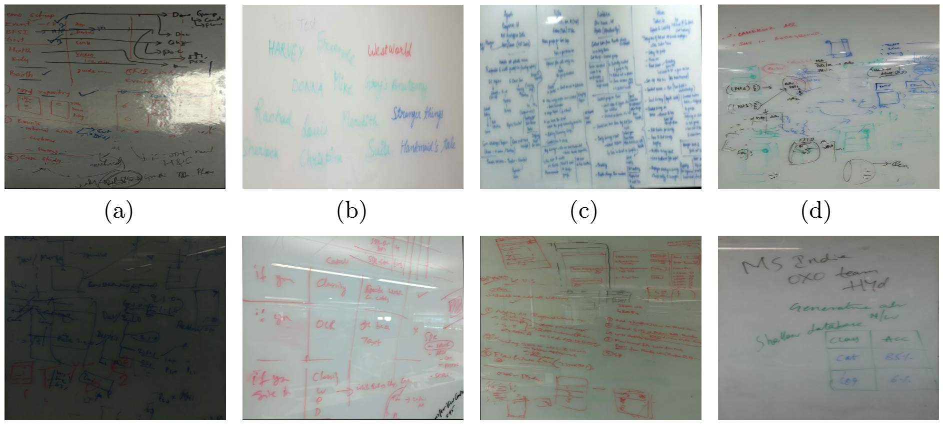
Abstract. Note-taking methods and devices have improved tremendously over the past few decades, and people are finding new ways to write notes and take photos. Automatic extraction, recognition, and retrieval are necessary to process the huge chunk of digitized document data. However, an important step in all of these pipelines is the pre-processing step, mainly image enhancement or clean-up, which enhances the text regions and suppresses the non-text regions. In this article, we look at the problem of image enhancement or clean-up on one such important class of images (i.e., mobile captured glass board images). We present a simple yet efficient algorithm using the concepts of classical image processing techniques to solve the problem, and the obtained results are promising in comparison to the Office Lens.

Keywords: Image enhancement · image binarization · specular highlight removal · glass board document images.

1 Introduction

Recent technological advancements in computer engineering have led to the development of several applications which changed the way we take notes. We moved on from storing the text in hard copies to taking notes on an electronic device. It means that a lot of text documents today are being circulated on the Internet, which is either scanned copies of the documents or electronically handwritten document images. For instance, several people take photos of slides, whiteboards, and glass boards containing text during conferences, class, etc., and scan the handwritten or printed papers directly using mobile.

Given the considerable number of these images, automatic recognition, extraction, and retrieval of text information from these images are significant in today's world. A couple of techniques have been proposed for this purpose. The performance of all these techniques heavily relies on pre-processing stage. This stage includes various tasks: (i) reduce noise, (ii) remove specular highlights, (iii) increase the contrast level between foreground (text) and background regions, (iv) separate text and non-text components, etc. Various techniques have been



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 1. Shows sample images containing (a) highlight text and background regions, (b) non-uniform color in text, (c) smudged text, (d) multi-color text, (e) different light condition, (f) non-uniform background, (g) non-uniform text density and (h) varying stroke of same character.

proposed in the literature to recover textual content (cleaning-up or enhancement) of heavily degraded document images.

Mobile scanned or captured glass board images containing text information make clean-up or enhancement more difficult due to several factors: (i) reflection of the glass board creates highlights on both the background and the handwritten text regions, (ii) due to reflection and improper clean-up of the glass board, text are smudged, (iii) multi-colored (black, red, blue, green, etc.) handwritten text, (iv) non-uniform density of the color in text and stroke of the same character is not constant due to handwriting, (v) non-uniform reflection of the glass board creates non-uniform smudged text, (vi) non-uniform density of text in the image, and (vii) non-uniform background. Fig. 1 highlights the effects of all these factors on mobile captured glass board images.

In this article, we propose a pipeline for cleaning up or enhancing glass board images using classical image processing techniques. The proposed pipeline includes three individual tasks: (i) specular highlight detection and removal from glass board images, (ii) segmentation of glass board images into the foreground (text) and background regions, and (iii) color assignment and enhancement of text region.

The first step includes detecting and removing specular highlight regions from the glass board images. A specular highlight is detected using the concept presented in [5]. Then highlight removal is done by proposing a heuristic concept based on local image processing. Therefore, this specular free glass board image is segmented into foreground (text) and background regions, using the adaptive thresholding technique with Gaussian weight. In the final step, the colors of pixels of the specular free glass board image are assigned to the foreground region, and a constant (near to white) color is assigned to the background region of the segmented image. A linear intensity transformation function is considered to enhance the text region. Since the performance analysis of the clean-up task is subjective, we decide to objectively analyze the solution with experts in the loop. The proposed algorithm is tested on 90 mobile captured glass board images, and it provides promising results compared to the existing techniques.

The rest of the article is organized as follows: Section 2 describes the related work. The proposed algorithm is explained in detail in Section 3. Section 4 presents experimental results and analysis. Finally, the conclusive remark is drawn in Section 5.

2 Related Work

Generally, document image enhancement techniques look at the problem as some form of binarization method. The problem of binarization can be solved by global or local processing techniques. The prevalent global thresholding method is Otsu's algorithm [9], which uses a single threshold value that maximizes the inter-class variance and minimizes the intra-class variance. It fails to segment the document images containing illumination variation and non-uniform background (i.e., various local changes in images) due to consideration of global image statistics. In order to address the problems of global thresholding based on document image binarization methods, many pioneering approaches using local image statistics were proposed in the literature [7, 10, 8, 11]. Niblack's [7] method considered local image statistics such as mean and standard deviation within a small image region to determine the local threshold. Under a well-conditioned environment, it obtained highly accurate results. However, its performance is sensitive to local image contrasts. Inspired by Niblack's method, Sauvola modified Niblack's linear decision to the non-linear decision by considering local variance [10]. It provides improved results over Niblack's method when the background contains light texture and non-uniform illuminated documents. In the same direction, Bernsen [8] proposed a modified local adaptive thresholding method based on the mean value of the minimum and maximum intensities of pixels within a small image region. Performance of this method reduces on documents with complex backgrounds. Su *et al.* [11] improve Bernsen's algorithm by introducing a normalization factor with local contrast terms to handle documents with complex backgrounds. It is called the Local Maxima Minima (LMM) method. The limitation of this technique is that it can not handle document images with bright text on bright background.

Disadvantages of the LMM method are overcome by Gatos's method [3]. Its performance does not require any parameter tuning. It contracts the background by interpolating neighboring background intensities. Therefore, the thresholding technique is applied by integrating the estimated background surface with the original image. It is an advantageous method for the binarization of degraded document images. However, it fails to binarize low-resolution images. Similarly, Wolf et al. also modified Niblack's method to provide better results on low-contrast images [13, 6]. Khurshid *et al.* [4] also modified Niblack's method and maned it as NICK to reduce the black noise in results obtained by Niblack's method. It provides good results for low-contrast images but fails for minimal and low-contrast conditions of thin pen stroke text. In a similar direction, Bataineh *et al.* proposed a robust local thresholding approach based on the mean and standard deviation for each current window and the global image [1]. It can able to produce good results for low-contrast images. It also overcomes the challenges of thin pen stroke text in the image. Feng *et al.* [2] also invented a local thresholding technique by considering local image statistics: minimum and standard deviations of two local windows. It is a modified wolf's method and solves the problem of low performance due to sharp variation in the background. Minor changes in parameter values could drastically affect the binarization results is the limitation of this method. Furthermore, Su *et al.* [12] proposed an adaptive contrast technique for binarization of degraded document images. It solves the problem of high inter/intra variation between the document background and the foreground text of different document images by considering both local image contrast and local image gradient. It is simple and robust, and it requires minimum parameter tuning. The performances of these discussed methods are highly dependent on the window parameters. In order to avoid manually adjusting the window size to the content and take advantage of Sauvola's algorithm, Lazzara and Géraud described a multi-scale binarization approach in [12].

It is proven that local thresholding techniques work well for images having local image variation, whereas global thresholding performs better for images having global variation. Each of these techniques has its advantages. Several techniques have been proposed in literature to segment document images using both concepts. In this direction, Biswas *et al.* applied global thresholding followed by local thresholding to perform binarization of document images in [6]. The authors constructed edge images using a Canny edge detector. The valley of the two peaks of the histogram of the non-edge pixel is chosen as a global threshold. The pixel values exceeding this threshold are turned as background. The remaining pixels are classified as foreground or background depending upon the local threshold using the average of the highest and lowest gray value of the window.

All methods are designed based on global thresholding, local thresholding, and combining both for binarization of degraded (printed, handwritten and historical) document images. Performances of such approaches reduce the processing of camera captures whiteboard images due to various complex factors: illumination variation, shadows, multi-colored handwritten text, etc. In [14, 15], Zhang and He proposed a technique to enhance the text regions of the whiteboard images. However, whiteboard images are similar to the established document image enhancement techniques.

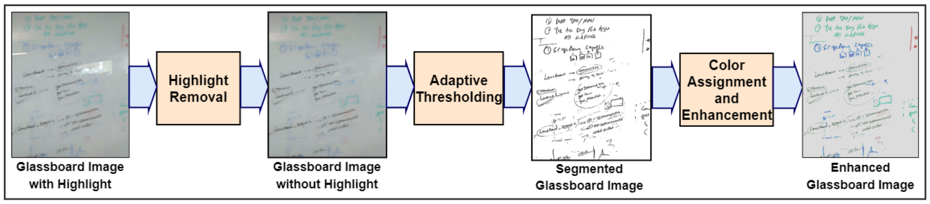


Fig. 2. Present the proposed algorithm for enhancement of camera captured glass board images.

3 The Proposed Algorithm

This section presents the proposed algorithm to enhance the quality of glass board images. Fig. 2 displays the block diagram of the proposed algorithm. The input to our algorithm is a rectified glass board image. Office Lens App[1] does image rectification. It is already mentioned that due to the reflection, the glass board images contain a lot of specular highlight regions. Our first job is to remove those specular regions by using a simple yet efficient specular highlight removal technique. After removing those specular regions, adaptive thresholding is applied to segment images into the foreground (text) and the background regions. Then foreground region is assigned with the color of the corresponding region in the specular free image, and the background region is assigned with constant color (i.e., near to white). Finally, a linear transformation is applied to obtain an enhanced cleaned-up glass board image. Each of these steps is elaborated on in the following subsections.

3.1 Highlight Removal

Segmentation of images containing specular highlights is challenging due to its complex nature. Therefore, removing specular highlights from the input glass board image is an essential task before the segmentation of the glass board image. Inspired by the work presented in [5], the specular free (SF) and modified specular free (MSF) images of the input glass board image are estimated using the following equations.

$$SF_c(x, y) = I_c(x, y) - min\{I_1(x, y), I_2(x, y), I_3(x, y)\}. \tag{1}$$

$$MSF_c(x, y) = SF_c(x, y) - \bar{I}_{min}. \tag{2}$$

$I_c(x, y)$ is the value of c^{th} color channel at $(x, y)^{th}$ pixel position of the original glass board image I. $c \in \{1, 2, 3\}$ indicates Red, Green, and Blue channel; and $\bar{I}_{min} = \frac{1}{MXN} \sum_{x=1}^{M} \sum_{y=1}^{N} min\{I_1(x, y), I_2(x, y), I_3(x, y)\}$. The highlight region

[1] https://play.google.com/store/apps/details?id=com.microsoft.office.officelenshl=en$_I N$

is detected by thresholding the difference image between the original glass board image I and the modified specular free image MSF. Therefore, the highlight detected image (HDI) is obtained by

$$HDI(x,y) = \begin{cases} 1 \ if \ D_c(x,y) \ > \ TH \ \forall \ c \\ 0 \ otherwise, \end{cases}$$

where $D_c(x,y) = I_c(x,y) - MSF_c(x,y)$. In [5], the authors discussed the selection of the threshold value TH. However, this threshold value (obtained using [5]) is not working for our (glass board image) dataset. Therefore, we manually select the value of TH as 20, which is best suited for our dataset.

After detecting highlight regions in the input image, our goal is to remove those specular (highlight) parts from the image using the concept of local image processing. Color values of a pixel with highlight are replaced by the average color value of all pixels without highlights within a small region in the input image around this particular pixel with highlight. Thus, a highlight-free glass board image ($\bar{I}$) is obtained using the following equation.

$$\bar{I}_c(x,y) = \begin{cases} I_c(x,y) \ if \ HDI(x,y) = 0 \ \forall \ c \\ I_c^{avg}(x,y) \ if \ HDI(x,y) = 1 \ \forall \ c, \end{cases}$$

where $I_c^{age}(x,y)$ is average color of $(x,y)^{th}$ position in the input image (I). I_c^{avg} is obtained using

$$I_c^{avg}(x,y) = \frac{\sum_{(i,j)\in w} I_c(i,j)}{l}, \ \forall \ (i, \ j), \ HDI(i,j) \ = \ 0, \tag{3}$$

where w is a small region around the center pixel (x,y) and l is many pixels without highlights within this small region. Here, it is noted that the small region w is selected in such a way that it contains at least one highlight-free pixel. The highlight-free image is considered as an input for adaptive thresholding to segment the image into two possible regions: foreground (text) and background.

3.2 Image Segmentation

The next step of the proposed algorithm is to segment specular highlight-free glass board images into two different regions: foreground (text) and background. Various algorithms have been developed to segment images in the literature. However, thresholding-based approaches are quite popular for segmenting camera-based document images due to their simplicity and good performance. In this work, we have explored the adaptive thresholding technique for segmenting specular free glass board images. For this purpose, the specular free color image glass board image ($\bar{I}$) is transformed into a grayscale specular free image using the following equation

$$\begin{aligned} I_G(x,y) = 0.2989 * \bar{I}_1(x,y) + 0.5870 * \bar{I}_2(x,y) \\ +0.1140 * \bar{I}_3(x,y). \end{aligned} \tag{4}$$

This grayscale specular free glass board image is considered as an input for the thresholding technique. Instead of the global thresholding technique, adaptive thresholding with Gaussian weight is considered to segment images into two regions: foreground (text) and background. For this purpose, each image is divided into many smaller regions. Therefore, pixels of each of these small regions are classified into foreground or background according to the equation

$$I_w(x,y) = \begin{cases} 1 \; if \; I_w(x,y) \; > \; TH_w \\ 0 \; otherwise. \end{cases}$$

Here, I_w is the small image patch, and TH_w is the local threshold value corresponding to this small region w. The threshold value TH_w is calculated as

$$TH_w = \frac{\sum_x \sum_y W(x,y) * I_w(x,y)}{Z}, \tag{5}$$

where $z = \sum_x \sum_y W(x,y)$ and $W(x,y)$ is the weight of a pixel at location (x,y) based on Gaussian distribution. We get different threshold values for different regions of an image. The adaptive thresholding technique provides better segmentation results for the glass board images containing various complex factors.

Furthermore, the results of adaptive thresholding depend on the size of the image patch (i.e., the window). Different windows provide different segmentation results as it contains varying image information. The window also varies from image to image. Therefore, selecting a proper window is crucial for the adaptive thresholding technique. In the subsequent subsection, we will discuss the selection of a proper window for adaptive thresholding.

3.3 Selection of Proper Window

The proper window is selected based on the statistical feature (variance) of the small image region. We first calculate the variance of an initial window. Then, we increase the window size in every step and select the best window for which variance exhibits a maximum local behavior. It can be mathematically expressed as

$$w_{best} = max_i\{V_{w_i}\} \; \forall \; i, \tag{6}$$

where V_{w_i} is the variance of the image patch corresponding to window w_i. The variance of a window will be maximum if it contains both the foreground and background information in a good ratio. It is also noted that we may find multiple windows having equal (maximum) variance in some cases. In such a case, we consider the window as the average of all these windows. It can be mathematically expressed as

$$w_{best} = g\left(int\left[\frac{\sum_{i=1}^{l} w_i}{l}\right]\right), \tag{7}$$

where l is the number of windows having equal (maximum) variance, and g is defined as

$$g(x) = \begin{cases} x+1 \ , if \ x \ is \ even \\ x \ , if \ x \ is \ odd. \end{cases}$$

Sometimes, it happens that there is no window having a maximum variance. For such a case, we chose the default window as 81×81.

3.4 Color Assignment and Enhancement

In the final step of the proposed algorithm, the colors of the pixels of the specular free glass board image, corresponding to the foreground region of the segmented image, are assigned to pixels in the foreground region of the segmented image. Background regions of the segmented image are assigned with unique (i.e., near to white) color. It is mathematically represented as

$$J_c(x,y) = \begin{cases} I_c(x,y) \ , if \ I_{seg}(x,y) = 1 \\ 220 \ , if \ I(x,y) = 0, \ \forall \ c, \end{cases}$$

where I_{seg} is the segmented (binary) image of the specular free glass board image $\bar{I}$. Finally, a linear transformation function is applied to the saturation component of the color assigned (RGB) image (J) to enhance the contrast between the foreground and the background regions. For this purpose, RGB color image J is converted into HSV color image J^{hsv} using standard RGB to HSV conversion. Thus

$$J^{hsv} = F(J) \ and \ \bar{J}^{hsv}_S(x,y) = T\left[J^{hsv}_S(x,y)\right], \tag{8}$$

where F is the RGB to HSV conversion function, $J^{hsv}_S(x,y)$ is the saturation value of the $(x,y)^{th}$ pixel of the HSV color image (J^{hsv}), $\bar{J}^{hsv}_S(x,y)$ is the enhanced saturation value and T is the enhancement function. Therefore, an enhanced HSV image is obtained using

$$\bar{J}^{hsv}_c(x,y) = \begin{cases} J^{hsv}_c(x,y) \ , if \ c \in \{H,V\} \\ \bar{J}^{hsv}_S(x,y) \ , if \ c \in S. \end{cases}$$

Finally, this enhanced HSV image $\bar{J}^{hsv}$ is converted to enhanced RGB image $\bar{J}^{rgb}$ using standard HSV to RGB conversion function. Therefore,

$$\bar{J}^{rgb} = F\left(\bar{J}^{hsv}\right), \tag{9}$$

where $\bar{J}^{hsv}$ is enhanced HSV image, $\bar{J}^{rgb}$ is the enhanced cleaned-up RGB glass-board image and F is standard HSV to RGB conversion function.

4 Experiments

4.1 Datasets

For experimental purposes, a dataset containing 90 glass board images is collected. A few of the sample images are displayed in Fig. 3. We have captured these glass board images by a mobile camera by varying zoom, angle of the camera, and light conditions. These factors increase the complexity of the dataset for the clean-up problem. Each image contains handwritten text with multi-color (black, red, blue, and green) font. The density of the color is not uniform for all text, and the stroke of the same character is not constant due to handwriting. The density of the text is also non-uniform for the images. Due to the reflection of the glass board and improper clean-up of the glass board, the text is smudged. Non-uniform reflection of the glass board creates non-uniform smudged text in the images. Reflection of the glass board also creates highlight regions on both the background and handwritten text regions. All of these factors make the clean-up task more difficult for the glass board images. Each of these images in Fig. 3 illustrates the complexity level of the clean-up task for the mobile captured glass board images.

4.2 Evaluation Measure

Since the performance analysis of the clean-up task is subjective, we decided to analyze the solution with experts in the loop objectively. Since the images are large to display on monitors and for experts to see comfortably, images are split into smaller patches. A total of 582 patches are selected randomly, with all the important parts of the 90 images being covered.

For this purpose, a simple approach for evaluating the performance of the method is proposed by asking questions to the experts. Each expert is asked a simple question, "Compare the outputs obtained by two different algorithms, A and B, for a given input," and the possible answers to this question are: (i) A is better than B, (ii) A is arguably better than B, (iii) A is similar to B, and (iv) A is less than B. The objective score of the algorithm is calculated as the ratio of the total number of positive votes (images) for each answer over the total number of images. Respective normalization is done to make sure that there is no bias in the expert's evaluation. Mathematically, it can be expressed as

$$S_{h(i)} = \frac{\sum_i g(h(i))}{n}, \tag{10}$$

where $h(i) \in \{r1, r2, r3, r4\}$, $r1 =$ “A is better than B", $r2 =$ “A is arguably better than B", $r3 =$ “A is similar to B", $r4 =$ “A is less than B" and $g(h(i) = r1) = 1$ if $h(i) \in r1$ otherwise 0.

Therefore, success rate of algorithm A and B are obtained using

$$\begin{aligned} S^A_{rate} &= S_{h(r1)} + S_{h(r2)} \\ S^B_{rate} &= S_{h(r4)}. \end{aligned} \tag{11}$$

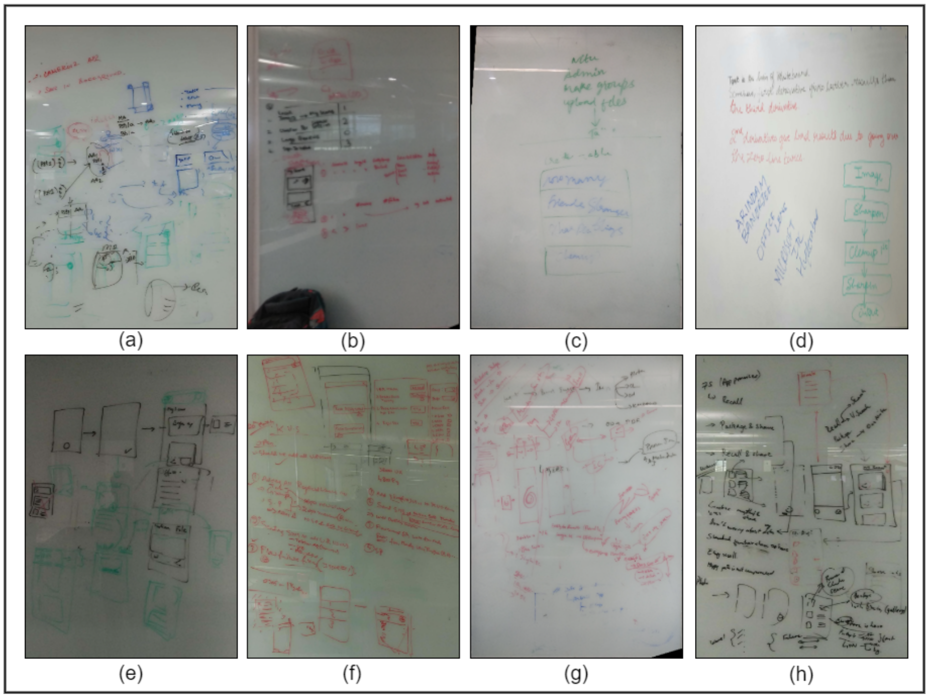


Fig. 3. Show few sample images containing several factors like (a) highlight text and background regions, (b) non-uniform color in text, (c) smudged text, (d) multi color text, (e) different light condition, (f) non-uniform background, (g) non-uniform text density, and (h) varying stroke of same character.

Therefore, gain of algorithm A over B is obtained using

$$G_A = \frac{S^A_{rate}}{S^B_{rate}}. \tag{12}$$

In our experiments, “A" is the proposed algorithm and “B" is any existing algorithm.

Name of B	A is better than B	A is arguably better than B	A is similar to B	A is less than B
Office Lens	$117/582 = 20\%$	$295/582 = 51\%$	$124/582 = 21\%$	$46/582 = 8\%$

Table 1. Objective comparison of the proposed method with the Office Lens with human in the loop.

Fig. 4. Shows original glassboard image and enhanced glassboard images obtained by Office Lens (OL) and Ours, respective.

4.3 Quantitative Results Analysis

Cleaned-up glass board images obtained by the proposed algorithm are compared with Office Lens' existing algorithm. The quantitative result obtained using this algorithm is given in Table 1. The table shows that among 582 image patches, the proposed algorithm produced better results for 117 and arguably better results for 295 image patches than Office Lens. In contrast, Office Lens produced better results only for 46 image patches. This numeric value underlined the superiority of the proposed algorithm over Office Lens. Here A is the proposed algorithm. Therefore, a gain of the proposed algorithm over the Office Lens is $412/46 = 8.96$.

4.4 Qualitative Results

Fig. 4 shows the qualitative results. In the case of images in the 1st row, the glass board image contains a few white spots due to improper cleaning of the board. We also observe that most text is blurred due to taking pictures far away from the board. It is challenging to recognize the text without cleaning the image. The cleaned image obtained by the Office Lens destroys most of the text. Due to that, it is very difficult to recognize the text. For example, "1", "2", are "3" are written at the lower of the board image. The Office Lens fails to preserve those digits while enhancing the board image. Please see the smaller image below the bigger image for better visualization. While our method enhances the board images by preserving text content. The smaller image below the larger image indicates that our method preserves text while enhancing. The user can easily recognize the text "1", "2", and "3" written within a small region. In the Second example, due to the reflection of the glass board, some background and text parts are highlighted. It creates a significant problem for the enhancement of glass board images. The highlight removal module in our proposed method takes care of this issue. The proposed method obtains better results than the Office Lens.

5 Conclusions

We have presented a simple and efficient algorithm for character enhancement in colored glass board images. The specular highlight removal process heavily influences the system's accuracy before the binarization. A roust specular highlight removal can be used here to boost the performance further. Also, given its failures, the next available solution would be to use deep networks and improve the accuracy of the enhancement. However, the amount of data we currently have is not enough. So, assuming that the deep model would scale up to a small dataset. We should also be able to come up with new metrics to measure the algorithm's effectiveness, thereby reducing human intervention in the system.

References

1. Bataineh, B., Abdullah, S.N.H.S., Omar, K.: An adaptive local binarization method for document images based on a novel thresholding method and dynamic windows. Pattern Recognition Letters **32**(14), 1805–1813 (2011)

2. Feng, M.L., Tan, Y.P.: Contrast adaptive binarization of low quality document images. IEICE Electronics Express **1**(16), 501–506 (2004)
3. Gatos, B., Pratikakis, I., Perantonis, S.J.: Adaptive degraded document image binarization. Pattern Recognition **39**(3), 317–327 (2006)
4. Khurshid, K., Siddiqi, I., Faure, C., Vincent, N.: Comparison of niblack inspired binarization methods for ancient documents. In: Document Recognition and Retrieval XVI. vol. 7247, pp. 267–275 (2009)
5. Koirala, P., Hauta-Kasari, M., Parkkinen, J.: Highlight removal from single image. In: International Conference on Advanced Concepts for Intelligent Vision Systems. pp. 176–187 (2009)
6. Lazzara, G., Géraud, T.: Efficient multiscale sauvola's binarization. International Journal on Document Analysis and Recognition (IJDAR) **17**(2), 105–123 (2014)
7. Niblack, W.: An Introduction to Digital Image Processing. Strandberg Publishing Company (1985)
8. Nurhadiyatna, A., Jatmiko, W., Hardjono, B., Wibisono, A., Sina, I., Mursanto, P.: Background subtraction using gaussian mixture model enhanced by hole filling algorithm (gmmhf). In: International Conference on Systems, Man, and Cybernetics. pp. 4006–4011 (2013)
9. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics **9**(1), 62–66 (1979)
10. Sauvola, J., Seppanen, T., Haapakoski, S., Pietikainen, M.: Adaptive document binarization. In: International Conference on Document Analysis and Recognition (ICDAR). vol. 1, pp. 147–152 (1997)
11. Su, B., Lu, S., Tan, C.L.: Binarization of historical document images using the local maximum and minimum. pp. 159–166 (2010)
12. Su, B., Lu, S., Tan, C.L.: Robust document image binarization technique for degraded document images. IEEE Transactions on Image Processing **22**(4), 1408–1417 (2012)
13. Wolf, C., Jolion, J.M.: Extraction and recognition of artificial text in multimedia documents. Formal Pattern Analysis & Applications **6**(4), 309–326 (2004)
14. Zhang, Z., He, L.w.: Note-taking with a camera: whiteboard scanning and image enhancement. In: IEEE International Conference on Acoustics, Speech, and Signal Processing. vol. 3, pp. iii–533 (2004)
15. Zhang, Z., He, L.W.: Whiteboard scanning and image enhancement. Digital Signal Processing **17**(2), 414–432 (2007)