AN ITERATIVE APPROACH FOR SHADOW REMOVAL IN DOCUMENT IMAGES

Vatsal Shah and Vineet Gandhi

CVIT, Kohli Center on Intelligent Systems, IIIT Hyderabad
vatsal.shah@students.iiit.ac.in, vgandhi@iiit.ac.in

ABSTRACT

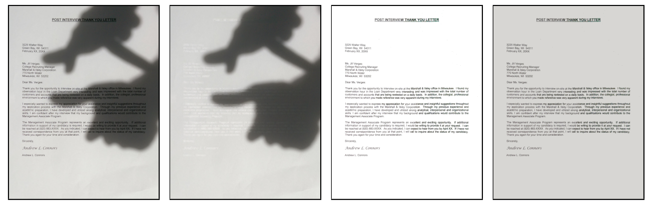
Uneven illumination and shadows in document images cause a challenge for digitization applications and automated work-flows. In this work, we propose a new method to recover unshadowed document images from images with shadows/uneven illumination. We pose this problem as one of estimating the shading and reflectance components of the given original image. Our method first estimates the shading and uses it to compute the reflectance. The output reflectance map is then used to improve the shading and the process is repeated in an iterative manner. The iterative procedure allows for a gradual compensation and allows our algorithm to even handle difficult hard shadows without introducing any artifacts. Experiments over two different datasets demonstrate the efficacy of our algorithm and the low computation complexity makes it suitable for most practical applications.

Index Terms— Shadow Removal; Document Analysis

1. INTRODUCTION

The prolific increase of smart phone usage has put internet and cameras into every pocket and that has led to significant changes in document work-flows. Given the easy access and the ease of sharing, camera phones are slowly replacing scanners as the mainstream document capture devices. However, the uncontrolled nature of the capture gives rise to degradations in the captured document. One of the common forms of degradation is shadows or uneven illumination, which can lead to challenges in automated digital work-flows (for example it can reduce accuracies of OCR algorithms). Even in other general scenarios of document capture for sharing and conservation purposes (newspaper articles, historical documents etc.), shadows may make the image difficult to interpret or use. This motivates the need of an algorithm which can remove the shadows without affecting the relevant details while preserving the natural color and tone of the document.

In this paper, we propose a novel shadow removal algorithm where we iteratively estimate a shadow map and use it to compute the unshadowed document image. The iterative procedure allows our algorithms to compensate the shadows gradually and organically, which is in contrast to single pass algorithms which may lead to artifacts when hard shadows are present. Our algorithm decomposes the image into shading and reflectance components and by applying post processing over the reflectance output, we are able to preserve the natural look of the document (a motivating example is illustrated in Figure 1). This is in contrast to binarization algorithms which lead to a binary image, solely targeting the OCR application. We present thorough qualitative and quantitative experiments to demonstrate the efficacy of our approach, compared to current state of the art.



Input Shading Reflectance Output

Fig. 1. We propose an algorithm for retrieval of unshadowed images. The figure shows a typical input image, the computed shading and reflectance images and the retrieved unshadowed image.

2. RELATED WORK

Early work on shadow removal focused on natural images. Finlayson et al. [1] posed the shadow removal problem as one of finding a special direction in a 2D chromaticity feature space, projecting to which leads to an image which is approximately invariant to intensity and colour of scene illumination. Guo et al. [2] exploit pairwise pixel similarities to first detect and then remove shadows. Interactive algorithms were explored in [3, 4]. It has also been approached as a general problem of intrinsic image decomposition [5, 6, 7, 8] in the wild, which is beautifully described in the work by Bell et al. [7]. However, these algorithms do not generalize well on document images.

Several efforts have been made targeting the specific application of shadow removal in document images. Most of these algorithm pose the shadow removal problem, as of intrinsic image decomposition. Brown et al. [9] assume a constant border around an image and use it to estimate the re-

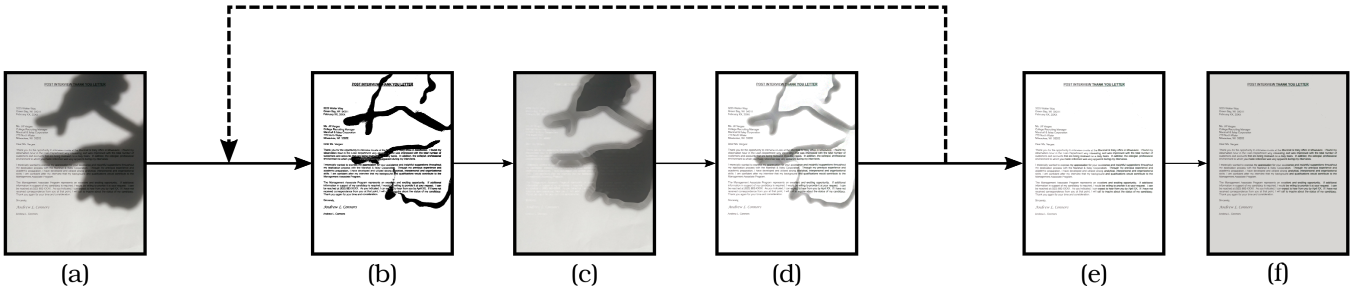


Fig. 2. System work-flow.(a) Input image (b) Image binarized into text and background to obtain M (c) Shading S estimated using M (d) Reflectance image R derived using S (e) Reflectance after completion of 10 iterations (f) Final tone adjusted output

flectance component. Zhang et al. [10], first estimate text regions and then inpaint them using the non text pixels. Later they fit a radial basis function to compensate for local noise. However, the fitting procedure may adversely affect the hard shadows or actual local variations present in the image. The computational load of such an approach is also intractable for most practical applications.

Oliveira et al. [11] replace the radial basis fitting to a natural neighbour interpolation mechanism, to account for pictures or other figures which may be present in the image. However, their method is also computationally heavy and may fail when excessive interpolation is required. More recently, Bako et al. [12] utilize variations in global clustering and local patch level clustering, to compensate for shadow component. By design such a method can not handle variations inside the local patch and is also sensitive to the considered patch size (hence it is also limited against hard shadows).

Finally, some approaches directly segment the image into black and white (text and non text) and discard all color and tone information [13, 14, 15, 16]. These approaches are specifically aimed at the goal of Optical Character Recognition, which is fundamentally different from our approach where the aim is to improve the document by removing the degradation caused by uneven illumination and shadows. However, our approach can be used as a preprocessing step to improve the OCR accuracies.

3. ALGORITHM

The proposed algorithm is iterative, wherein each iteration encompasses: binarization, shading estimation and reflectance estimation. Post the final iteration, the tone in the reflectance image is adjusted to produce the final shadow-less image. The algorithm is illustrated in Fig. 2 with relevant explanation in the following subsections.

3.1. Binarization

The first step of the system binarizes the input image I into two sets foreground(text) $\mathcal{F}$ and background $\mathcal{B}$ represented in Fig. 2(b). We use Bradley et al.'s [17] adaptive thresholding algorithm with a high threshold for foreground pixels. A high threshold ensures that none of the text pixels are misclassified as background. This gives a binary image M where, $M(x,y) = 0, \forall(x,y) \in \mathcal{F}$ and $M(x,y) = 1, \forall(x,y) \in \mathcal{B}$.

3.2. Reflectance and Shading Estimation

Decomposing an image into reflectance and shading is at the core of several image processing problems. The reflectance captures the inherent properties of the objects in an image and the shading captures illumination and shading conditions of the scene. Eliminating the shading component from an image results in an image devoid of any shadows. Note that, the shading map for a document image, looks almost exactly like the input image devoid of any foreground text. Attesting to this, we obtain a shading map S by keeping the background pixels to be the same as the input image and replacing the foreground pixels by interpolating their surrounding background pixels. Formally:

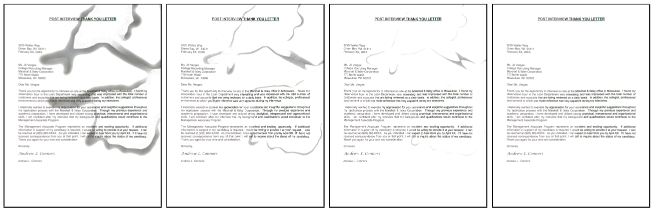
$$S(x,y) = \begin{cases} I(x,y) & \text{if } (x,y) \in \mathcal{B} \\ \dfrac{\sum_{i=x-dx}^{x+dx} \sum_{j=y-dy}^{y+dy} I(i,j) \cdot M(i,j)}{\sum_{i=x-dx}^{x+dx} \sum_{j=y-dy}^{y+dy} M(i,j)} & \text{if } (x,y) \in \mathcal{F} \end{cases} \quad (1)$$

To ensure that the interpolated value is a good estimate, the neighbourhood window of size $dx \times dy$ is adaptively chosen such that there are at least np number of background pixels present in the neighbourhood window. In practice, a value of 25 for np works well.

Having obtained the shading map S, the reflectance R(Fig. 2(d)) of the image is obtained as:

$$R(x,y) = \frac{I(x,y)}{S(x,y)} \quad (2)$$

For dark shadows, we observe that after a single iteration of the aforementioned steps, residual shadows still persist in the reflectance images. To resolve this, the reflectance image is iteratively passed through the pipeline, until either there are



(a) (b) (c) (d)

Fig. 3. Illustrating the improvement in obtained reflectance across iterations. Reflectance images after (a) 1 iteration (b) 3 iterations (c) 5 iterations (d) 10 iterations

no changes between the reflectance images in two consecutive iterations, or if the iteration limit is reached. The gradual improvement in the reflectance image is showcased in Fig. 3

3.3. Tone Adjustment

Upon completion of the iterative procedure, we have 3 images: R, S and M. Background pixels have values close to 1 in R. To maintain the tone of the original image, we apply a post processing step. We globally threshold the shading image S using Otsu's [13] method to get a shadow mask SM, and calculate a mean global background color as follows :

$$gm = \frac{\sum_{i=1}^{r}\sum_{j=1}^{c} I(i,j)\cdot M(i,j)\cdot SM(i,j)}{\sum_{i=1}^{r}\sum_{j=1}^{c} M(i,j)\cdot SM(i,j)} \tag{3}$$

The final shadow-less image $\hat{I}$ is obtained by scaling the reflectance image by gm as follows:

$$\hat{I}(x,y) = R(x,y)\times gm$$

It is worth noting that $\forall(x,y)\in\mathcal{B}$, $R(x,y)=1$. To ensure that none of the pixels belonging to text are lost to the background, the threshold for initial binarization is kept high, and the binary image M is dilated with a disc of radius r (we use a value of r = 5). However, as a result, major portions of shadows also get classified as foreground pixels in M and appear in the reflectance image R, as illustrated in Fig. 2(a). This is precisely why the iterative nature of our algorithm is crucial. Across successive iterations, the shadows in the reflectance images get lighter and narrower, and eventually get removed. This is illustrated in Fig. 3.

4. EXPERIMENTS

In this section, we present results obtained using our novel algorithm over two datasets. We also present thorough comparisons with the current state-of-the-art.

	Average MSE
Bell et al. [7]	125.44
Gong et al. [3]	390.98
Wagdy et al. [15]	74.06
Pilu et al. [18]	67.38
Bako et al. [12]	24.60
Proposed Approach	**24.37**

Table 1. Results of different approaches on Adobe Dataset

	Average MSE
Bako et al.	76.28
Proposed Approach	**35.56**

Table 2. Comparison of Bako et al's [12] algorithm with our approach on proposed HS dataset

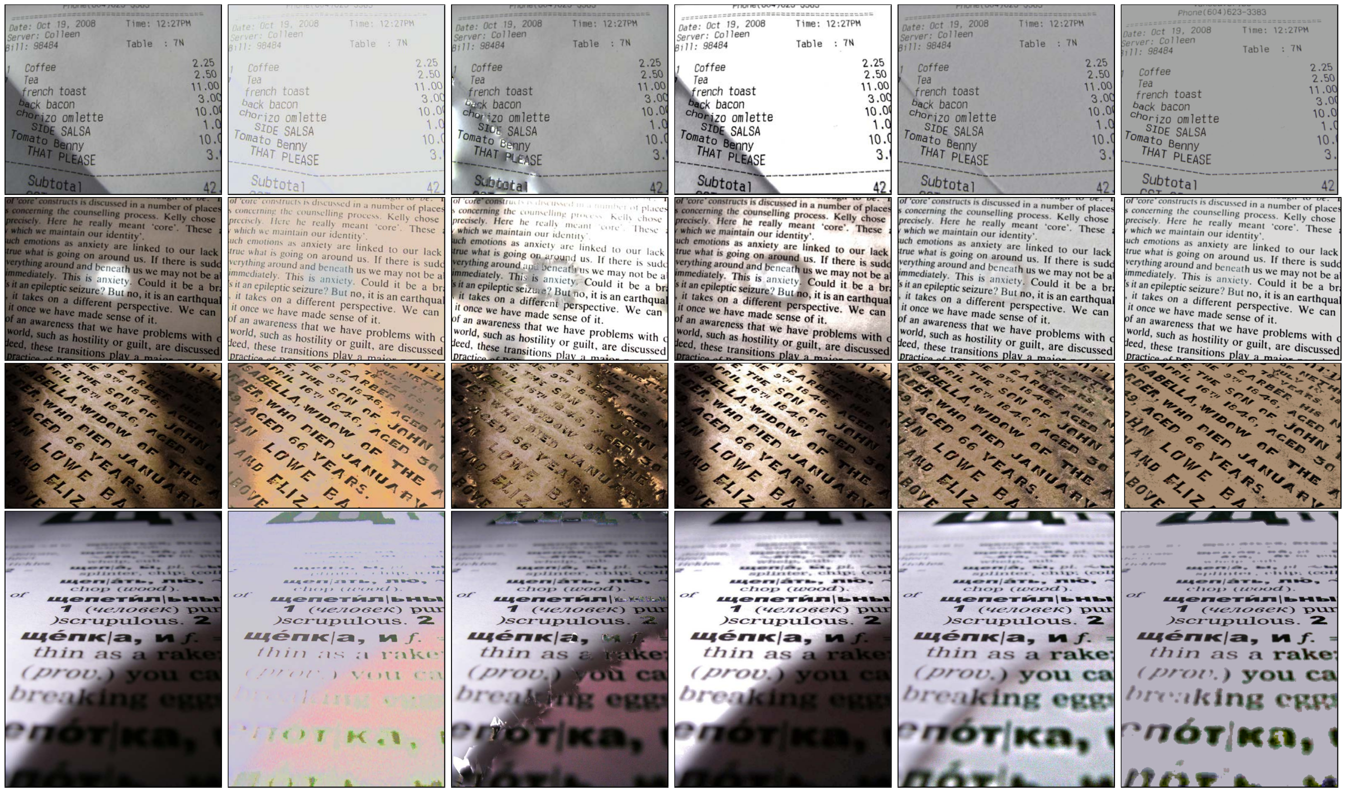
4.1. Datasets

Adobe Dataset: The Adobe dataset has been proposed and used in the work by Bako et al. [12]. This dataset consists of 81 images taken using 11 documents. The images are taken in a controlled setup with varying degrees of intensities and shapes of shadows. A corresponding unshadowed ground truth image is provided for each shadowed image.

HS dataset: The Adobe dataset is limited in terms of range of shadow intensities in an image. Hence, we propose a more challenging dataset with harder shadows with sharp transition between shadowed and non-shadowed regions (calling it HS dataset). We fix the capturing device over a tripod and take multiple images of a document using obstructions of varying shapes. Additionally, we also vary the distances between obstructions and document, resulting in images with a broad range of shadow intensities. The image captured without any shadows is considered as the ground truth for the document, in consistency with the previous dataset.The dataset comprises of 100 images of 10 documents, 10 per each document, captured using a Moto G5 Plus camera. To simulate varying use case scenarios, 5 images per document are captured in sunlight, and 5 under an LED lamp.

4.2. Quantitative Evaluation

We present a quantitative comparison of our approach over both the datasets using average Mean Squared Error(MSE) to get a score correlating with similarity between retrieved and ground truth images. We compare the scores generated using our algorithm with the scores of the algorithms proposed in [7], [3], [15], [18] and [12]. The MSE scores for these algorithms along with our approach on Adobe dataset are presented in Table 1. We have matched the background color of retrieved image and ground truth image to avoid brightness



(a) Input (b) Bell et al. [7] (c) Gong et al. [3] (d) Acrobat Enhance (e) Bako et al. [12] (f) Our method

Fig. 4. Comparisons on challenging real-world images from Flickr. Our method consistently produces results free of any residual shadows and artifacts near the transition regions.

differences. The results show that the proposed algorithm performs comparable to state-of-the-art [12]. However, as mentioned previously, this dataset does not accurately encompass the range of shadow intensities possible. Therefore, we also present MSE scores on the proposed HS dataset in Table 2 for a better analysis and comparison with the closely performing algorithm [12]. The results indicate that the proposed algorithm clearly outperforms [12], especially in the presence of challenging hard shadows.

4.3. Qualitative Evaluation

In this section, we present qualitative results of our approach and their comparison with various algorithms as seen in Fig 4. It can be observed that all previous approaches either fail in eliminating shadows completely or introduce some other artifacts in the process. The algorithm proposed by Bell et al. [7] loses a considerable amount of text in the output image, along with generating some reflectance artifacts. Similarly, the approach used in [3] fails to estimate the shadow statistics, even after being provided some cues by the user. The algorithm used in the Acrobat 'Enhance' tool is a global approach, thus it is not able remove the local shadows generated in real scenes. The state-of-the-art algorithm proposed in [12] removes the shadows but leaves traces near the sharp transition regions. This may be due to the presence of transitions inside a local patch. This effect tends to be even more pronounced in images with stronger shadows. Our algorithms overcomes this limitation iteratively, reducing the traces across iterations.

5. CONCLUSIONS

In this paper we present a novel computationally efficient algorithm for removal of shadows from document images (current implementation takes about 2 seconds to run for a 1600x1200 image on a Core-i7 processor). Our algorithm iteratively removes shadows and recovers the unshadowed image. We also propose a novel dataset comprising of images with shadows of varying intensities and shapes for a comprehensive analysis of the algorithm over real world document images. Thorough extensive quantitative and qualitative experimentation, demonstrate that the proposed pipeline outperforms the presently used algorithms for shadow removal over a wide range of input images. One limitation of our approach is that it cannot handle presence of pictures/images inside the document, which we plan to address in future work.

6. REFERENCES

[1] Graham D Finlayson, Mark S Drew, and Cheng Lu, "Entropy minimization for shadow removal," *International Journal of Computer Vision*, vol. 85, no. 1, pp. 35–57, 2009.

[2] Ruiqi Guo, Qieyun Dai, and Derek Hoiem, "Paired regions for shadow detection and removal," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 12, pp. 2956–2967, 2013.

[3] Han Gong and DP Cosker, "Interactive shadow removal and ground truth for variable scene categories," in *BMVC 2014-Proceedings of the British Machine Vision Conference 2014*. University of Bath, 2014.

[4] Maciej Gryka, Michael Terry, and Gabriel J Brostow, "Learning to remove soft shadows," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 5, pp. 153, 2015.

[5] Qi Zhao, Ping Tan, Qiang Dai, Li Shen, Enhua Wu, and Stephen Lin, "A closed-form solution to retinex with nonlocal texture constraints," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 7, pp. 1437–1444, 2012.

[6] Qingxiong Yang, Kar-Han Tan, and Narendra Ahuja, "Shadow removal using bilateral filtering," *IEEE Transactions on Image processing*, vol. 21, no. 10, pp. 4361–4368, 2012.

[7] Sean Bell, Kavita Bala, and Noah Snavely, "Intrinsic images in the wild," *ACM Trans. on Graphics (SIGGRAPH)*, vol. 33, no. 4, 2014.

[8] Tinghui Zhou, Philipp Krahenbuhl, and Alexei A Efros, "Learning data-driven reflectance priors for intrinsic image decomposition," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3469–3477.

[9] Michael S Brown and Y-C Tsoi, "Geometric and shading correction for images of printed materials using boundary," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1544–1554, 2006.

[10] Li Zhang, Andy M Yip, and Chew Lim Tan, "Removing shading distortions in camera-based document images using inpainting and surface fitting with radial basis functions," in *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*. IEEE, 2007, vol. 2, pp. 984–988.

[11] Daniel Marques Oliveira, Rafael Dueire Lins, and Gabriel de França Pereira e Silva, "Shading removal of illustrated documents," in *International Conference Image Analysis and Recognition*. Springer, 2013, pp. 308–317.

[12] Steve Bako, Soheil Darabi, Eli Shechtman, Jue Wang, Kalyan Sunkavalli, and Pradeep Sen, "Removing shadows from images of documents," in *Asian Conference on Computer Vision*. Springer, 2016, pp. 173–183.

[13] Nobuyuki Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[14] Basilios Gatos, Ioannis Pratikakis, and Stavros J Perantonis, "Adaptive degraded document image binarization," *Pattern recognition*, vol. 39, no. 3, pp. 317–327, 2006.

[15] Marian Wagdy, Ibrahima Faye, and Dayang Rohaya, "Fast and efficient document image clean up and binarization based on retinex theory," in *Signal Processing and its Applications (CSPA), 2013 IEEE 9th International Colloquium on*. IEEE, 2013, pp. 58–62.

[16] Ioannis Pratikakis, Basilis Gatos, and Konstantinos Ntirogiannis, "Icdar 2013 document image binarization contest (dibco 2013)," in *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*. IEEE, 2013, pp. 1471–1476.

[17] Derek Bradley and Gerhard Roth, "Adaptive thresholding using the integral image," *Journal of Graphics Tools*, vol. 12, no. 2, pp. 13–21, 2007.

[18] M. Pilu and S. Pollard, "A light-weight text image processing method for handheld embedded cameras," in *Proc. BMVC*, 2002, pp. 53.1–53.10, doi:10.5244/C.16.53.