A Statistical Approach for Multi-frame Shadow Movement Detection and Shadow Removal for Document Capture

Prasenjit Mondal and Ankit Bal
Adobe India
pmondal@adobe.com, ankit@adobe.com

***Abstract*—Shadow detection is an integral part of scanned document image processing. In this paper, we have introduced a new technique for identifying shadow movement regions in documents. Provided a sequence of image frames, our algorithm identifies documents from an image using a boundary detection algorithm followed by computing their respective perspective corrected images. Further, intensity map images are computed from the perspective corrected images. Finally, a statistical analysis is carried out on the different blocks of the intensity maps to estimate the shadow movement regions. The identified blocks along with a shadow removal algorithm are used to remove shadows from the documents. The proposed algorithm has been tested in with the Adobe Scan application and the experimental results show the efficacy of the proposed algorithm. The algorithm has a low computation complexity which makes it suitable to use in different mobile applications.**

***Index Terms*—Shadow movement detection, Shadow removal, Intensity map, Document analysis**

I. Introduction

With the advancement of high-resolution cameras in mobile devices, capturing document images by hand-held devices is becoming very common as compared to the bulky desk scanners. However, the presence of different lighting conditions and uneven illuminations degrade such camera captured documents. Shadow is a common type of degradation and poses significant problems in automatic restoration or preservation of the documents to their original form. The shadows cast in a document can be a hard shadow or a soft shadow. A hard shadow is defined with crisp and sharp shadow edges whereas a soft shadow fades off toward the shadow edges.

In recent years, several efforts have been put for accurate segmentation and removal of shadows in document images using machine learning techniques. In [2], Nagae *et al.* have introduced a Generative Adversarial Network (GAN) based on a shadow model to detect shadows in images. The algorithm uses luminance variations and the illuminance of the shadowed regions (for which the amount of change is large) to estimate the shadow and removes shadows thereafter. Want *et al.* in [5] have proposed a stacked conditional GAN (ST-CGAN) which is composed of stacked CGANs, each with a generator and a discriminator. Their algorithm works with a multi-task perspective to jointly learn both shadow detection and shadow removal as an end-to-end fashion where a shadowed image is fed to the first generator which produces a shadow mask. The input image along with the shadow mask is further fed to the second generator to compute the shadow free image. To remove soft shadows in images, Gryka *et al.* [9] have proposed a supervised regression technique that learns a mapping function for image patches that generates shadow mattes.

Intrinsic image decomposition techniques are also widely used for shadow removal. Zhou *et al.* in [8] have introduced a two-stage learning technique for intrinsic image decomposition. The approach first predicts relative reflectance ordering between image patches followed by image decomposition. In [11], a CRF-based approach has been presented for evaluating intrinsic image decompositions of indoor scenes. Further, the Retinex theory and texture analysis have also been used in [14] for the same. In [15], Yang *et al.* have proposed a technique that first computes a 2D intrinsic image from an input image based on chromaticity. Based on bilateral filtering and the 2D intrinsic image, a 3D intrinsic image is computed which is finally used to get the shadow-free image.

Image processing techniques are very popular for efficient shadow detection and removal. For example, in [6], the authors have proposed an iterative approach that first estimates the shading of an input image which is used to compute a reflectance map. The map is used in an iterative fashion to remove shadows in the image. In [7], a shadow removal technique have been proposed by assuming the constant colored background of documents. The algorithm first estimates background and text color in local image blocks followed by computation of a shadow map which is used to compute shadow-free image. In [10], an interactive approach has been proposed for shadow detection and removal that uses on-the-fly learning approach guided by users to discriminate between

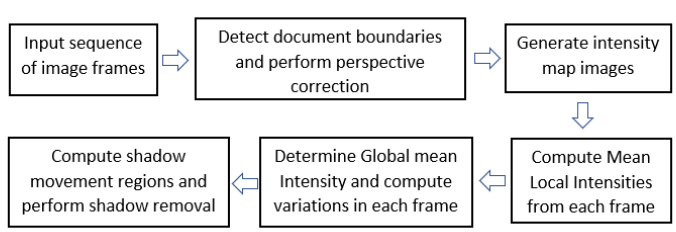
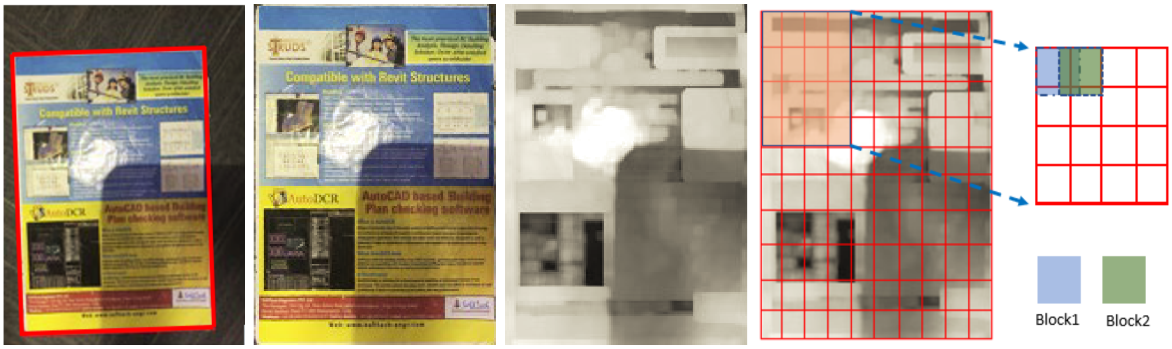


Fig. 1: Block diagram of the proposed algorithm.



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

Fig. 2: Mean local intensity (MLI) computation: (a) Original image with document boundary, (b) Perspective corrected image, (c) Intensity map, and (d) Dividing an intensity map into overlapping image blocks for computation of mean intensity.

shadowed pixels and the lit area which is used to compute shadow boundary intensity changes. Guo *et al.* [12] have proposed a region based approach to predict relative illumination changes between segmented regions. However, the algorithm is well suited for natural scenes. In [13], a shadow removal technique is presented based on the segmented background of a document which are used to estimate the shadow of non-background pixels by means of natural neighbor interpolation technique. The authors in [16] have formulated the shadow removal problem as finding a special direction in a 2D chromaticity feature space which is approximately invariant to intensity and color of scene illumination.

While single frame shadow detection has received significant attention, very fewer multi-frame shadow detection techniques are explored mainly because of lack of datasets with well annotated videos and their respective shadows. In [1], a cooperative video shadow detection network (TVSD-Net) is presented that learns discriminative representations at intra-video and inter-video levels for shadow detection.

Most of the existing shadow segmentation algorithms keep residual shadow during shadow removal. This is mainly because of the inappropriate shadow boundary identification and segmentation. For efficient shadow removal, identification of such shadow (both soft and hard shadows) boundaries is important. In this paper, we have proposed a novel technique based on statistical analysis of local blocks in a sequence of frames for shadow boundary identification. The blocks are further used to effectively remove shadows. A block diagram of the proposed algorithm is presented in Fig. 1.

II. Shadow Movement Analysis

The proposed algorithm computes shadow movement regions based on the following assumptions:

- During document capture using mobile devices, user's hand moves/shakes a little bit. This is validated by applying our algorithm on a dataset captured by anonymous users.
- The relative movement of the document (which is 2D in nature) and the shadow can be determined using hand movement due to the difference of distance between the objects throwing the shadow and the light source, as compared to the document.

The proposed shadow movement region identification algorithm considers a sequence of image frames from mobile camera. Let $F = F_1, F_2, F_3, ..., F_n$ be a sequence on n frames taken at regular interval. The detailed algorithm is discussed in the following subsections.

A. Document Registration

As the algorithm operates on local image blocks in multiple frames, segregation of the document region with respect to the background and registration of the documents are necessary. Registration of all the frames is performed by computing exact document boundary and mapping each document to a same space. Document boundary based image registration may produce pixel level precision error which is taken care by considering image block regions while processing them. In this work, boundary detection is performed by using a contour-based method [3] that computes a boundary based on the contrast changes between the inside and outside of a quadrilateral. Based on the computed boundaries, each frame F_i is extracted from the background and their respective perspective corrected image is computed. Let $I = I_1, I_2, I_3, ..., I_n$ be the perspective corrected images.

B. Intensity Map Computation

The proposed statistical approach operates on intensity map images for shadow movement detection. An intensity map (M_i) is a gray scale image computed from I_i where each pixel is set to the highest luminance level in a configuration window surrounding the pixel in I_i. To compute M_i, we consider any suitable measure of luminance, such as the Y-channel in the YC_bC_r color space or the L-channel in the $CIE-LAB$ color space and accumulate the values under consideration and finally choose the highest luminance value. The intensity maps are further smoothen using Gaussian filter to remove artifacts.

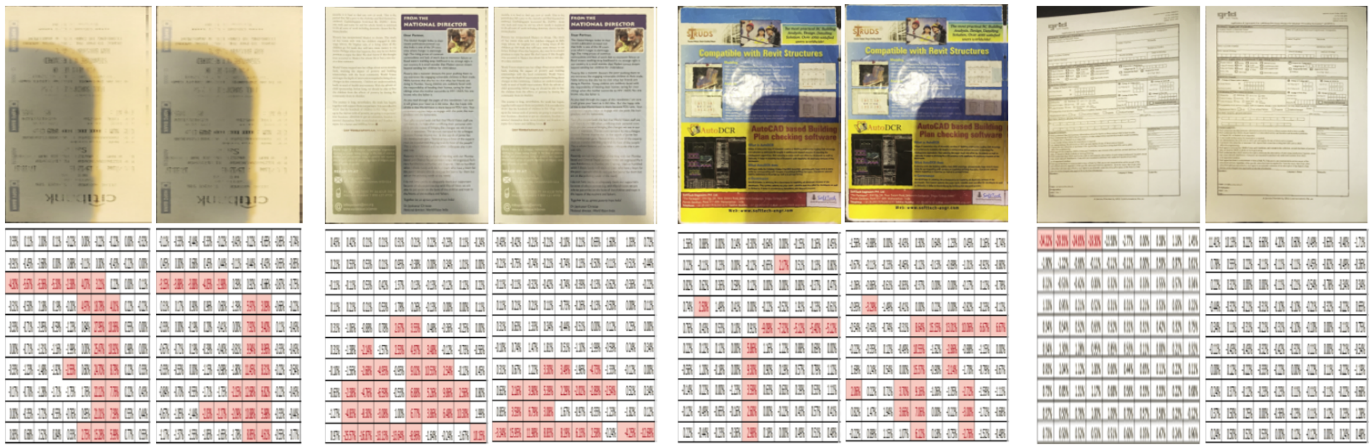


Fig. 3: Results of the shadow movement detection algorithm. Top row: Some sample frames are presented. Bottom row: Their corresponding shadow movement image blocks on the intensity variation matrix V are highlighted with red color.

C. Statistical Approach

Let $M = M_1, M_2, M_3, ..., M_n$ be n intensity maps computed from I. First, we divide each intensity map image into a number of local blocks (for example $p \times q$). Mean local intensity (MLI) is computed for each image block in $p \times q$ using Eq 1. To reduce the intensity variations, MLI is computed on overlapping image blocks of $(p+\Delta)$ and $(q+\Delta)$. A total of n numbers of MLI matrices are computed each with dimension of $p \times q$. A pictorial representation of computing MLI is shown in Fig. 2(d).

$$MLI_i(r, c) = \frac{\sum M_i(u, v)}{(\frac{image\ height}{p+\Delta} \times \frac{image\ width}{q+\Delta})} \quad (1)$$
$$\forall\ 1 \le i \le n, 1 \le r \le p\ and\ 1 \le c \le q$$

where $\sum M_i(u, v)$ is the sum of all the pixels from the overlapping block in i^{th} image.

To estimate the intensity variations on the entire image sequence M, we compute global mean intensity (GMI) variations using MLI as shown in Eq. 2.

$$GMI(r, c) = \frac{\sum_{i=1}^{n} MLI_i(r, c)}{n} \quad (2)$$

Once we have the MLI for each image frame and the GMI of all the frames, the intensity variations (V) of each block corresponding to each image frame is computed using Eq. 3. From the analysis, a threshold value on V is determined that denotes a probable shadow movement image block region in each frame. The proposed shadow movement detection algorithm is patented [4] by Adobe.

$$V_i(r, c) = \frac{MLI_i(r, c) - GMI(r, c)}{GMI(r, c)} \times 100 \quad (3)$$

D. Shadow Removal

Shadow removal of an image is done based on computing a shadow map of the image and the intensity variation V computed from the image sequence. Shadow map computation, based on local and global background colors, is a well known technique in literature [7]. Let α be the computed shadow map. Given an input shadowed image I, the shadow removed image $\hat{I}$ is computed in Eq. 4.

$$\hat{I}(i, j) = \frac{I(i, j)}{\alpha(i, j) \times \lambda(i, j)} \quad (4)$$

While removing shadow for $(i, j)^{th}$ pixel, the value of $\lambda(i, j)$ is set to 1 if $(i, j)^{th}$ pixel does not belong to a shadow movement block; otherwise $\lambda(i, j) = c \times V(i, j)$ where c is a constant. The value of c is proportional to the value of $V(i, j)$ which determines how much shadows to be removed from an image block. A higher value of c is used for more aggressive shadow removal.

After shadow removal, frequently it is observed that the text quality becomes poor. To fix this, we have computed a document mask using an integral image based adaptive thresholding technique [17]. Given an input image, an integral image is a function that maps from pixels to real numbers, and this is computed as the sum over a rectangular region of the image. The document mask is used to darken the pixels in shadow removed image. A few sample images and their shadow removed results are shown in Fig. 4. The results show that the proposed algorithm successfully removes shadows from images.

III. Experimental Results

The section presents the experimental results for shadow movement detection, the datasets used for the same and comparison with other techniques for shadow removal.

A. Datasets

To create dataset for shadow movement detection, we have developed a mobile application to capture document images. Using the application, members in our team are asked to capture documents (both using Android and iOS mobile phone cameras) in a traditional way without letting them know the assumptions used in the proposed shadow movement

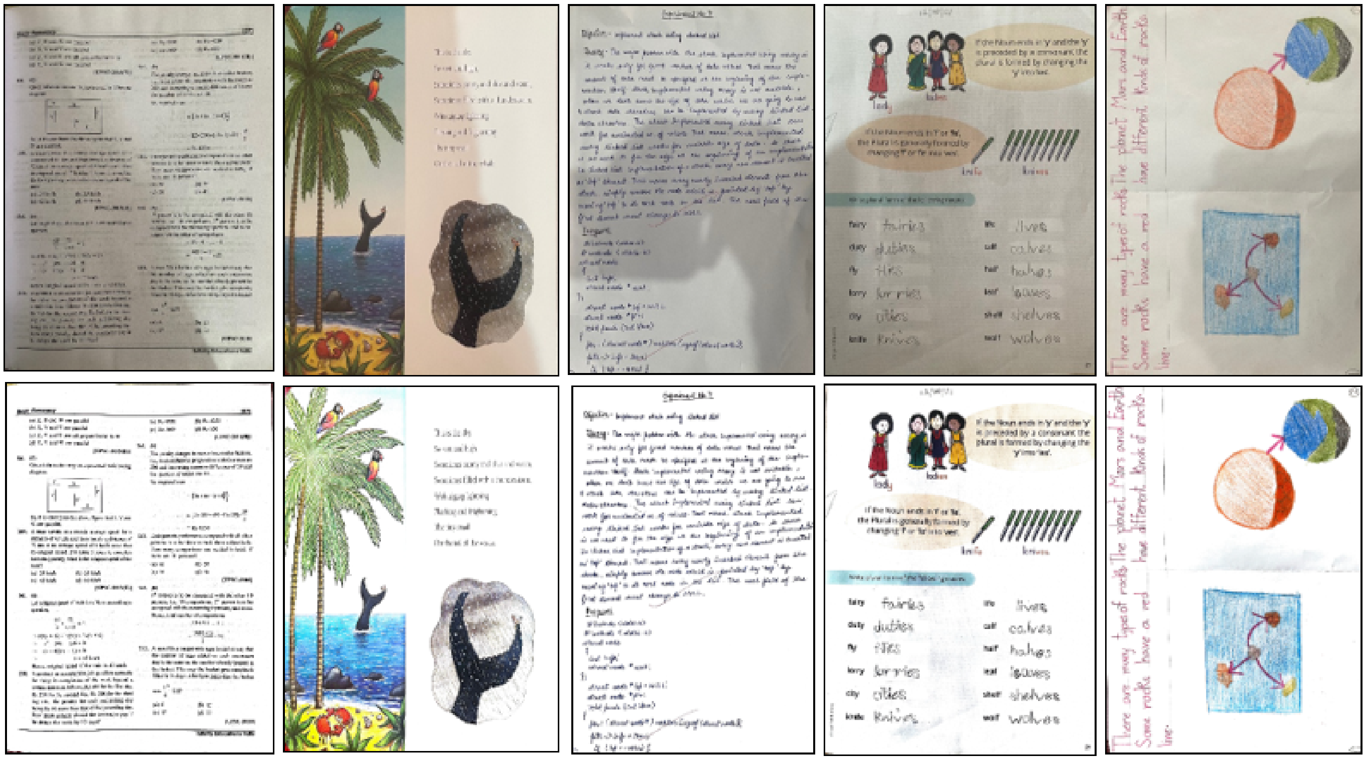


Fig. 4: Shadow removal results. Top row: Input images with different shadow patterns. Bottom row: Their corresponding shadow removal output using the proposed algorithm.

technology. Meanwhile, we recorded a sequence of frames in the application background corresponding to each capture. To capture documents, we have considered a number of PDF files and they are printed using a standard printer. Such printed documents are used for capturing datasets. Furthermore, we have generated the corresponding ground truth of each document by converting the PDF files to images using loss-less image compression techniques in Acrobat. A total of 200 datasets (each with 5 frames in sequence) are selected with varying intensities and shapes of shadows (both soft and hard) and used for the proposed invention.

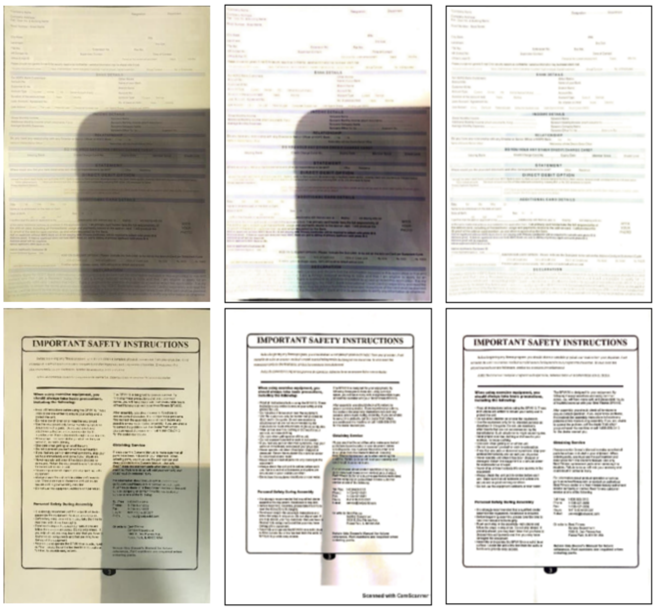
B. Results and Discussion

1) Shadow Movement Detection: While analyzing V_i for $1 \leq i \leq n$ computed from a sequence of n frames, we have identified that the values in V_i in i^{th} frame which are greater than $+2\%$ or less than -2% represents the image blocks where there are shadow movements and the values which are greater than $+15\%$ or less than -15% depicts that there are artifacts which are not present in few of the frames in the image sequence under consideration and mainly caused due to misalignment. Experiments shows that these threshold values are consistent in datasets captured using both iOS and Android device cameras. The result of applying the proposed method for shadow movement detection is shown in Fig. 3 where few sample frames and their variation matrix V are shown. The image blocks in the variation matrix which satisfies the threshold values are highlighted in red color. The variation matrix is reshaped according to the input frames for a better visual understanding of shadow movement regions. In the figure, first three sets of frames shows shadow movements on different types of documents and the last set shows variation due to image registration failure. The figure shows that the proposed algorithm is able to compute the shadow movement regions very efficiently and can be used to detect artifacts in a sequence of frames.

During the experimentation, we have also observed the presence of few isolated shadow movement blocks (whose variations satisfy threshold criteria). Such isolated blocks are removed with the assumption that shadows can not be isolated on a document and it should touch the document boundary. Also, on both iOS and Android devices, the average running time to detect the shadow movement blocks on a sequence of 5 frames (each with a dimension of 256×256) is recorded as 6ms which allows the algorithm suitable for mobile device usage. There are several advantages of the proposed statistical approach as can be used in the following:

- Real-time analysis of presence of shadows.
- For effective detection of artifacts in image sequence.
- Helps in shadow removal around shadow boundaries.
- Shadow detection on pictures/images inside a document.

To the best of our knowledge, the proposed shadow movement detection algorithm is the first statistical technique used for shadow boundary detection and also there are no standard video dataset available for documents and their annotated



(a) (b) (c)

Fig. 5: Shadow removal result before and after using the proposed algorithm. (a) Sample input images, (b) Shadow removed result without using the proposed shadow detection technique, and (c) Results with the proposed algorithm.

ground truths. Therefore, we can not compare our technique with other algorithms available for the same.

2) Shadow Removal: Due to the unavailability of annotated document video dataset, the proposed algorithm can not be directly compared with the existing shadow removal algorithms in the literature. However, to validate the effectiveness of the proposed shadow removal method, we have presented the results with and without the proposed algorithm in Fig. 5. From the figure, it can be observed that the traditional approach introduces artifacts around shadow regions or loses text quality or have shadow residuals, whereas the proposed technique has successfully removes shadows from different documents with minimum artifacts.

Adobe Scan is a very popular application as a portable document scanner to generate high quality documents. Due to the robustness and efficiency, the proposed shadow movement detection and the shadow removal algorithms have been integrated with Adobe Scan application.

IV. Conclusion

This paper presents a novel multi-frame statistical approach to detect shadow movement regions in image frames. Intensity variations along the shadow movement regions are computed and further used for efficient removal of shadows. The qualitative comparisons are carried out between the proposed algorithm and a traditional shadow removal algorithm. Evaluation results show the superiority of our method as compared to other algorithms. During experimentation, is has been observed that our algorithm works well on various document types with different lighting conditions and various shadow patterns. Due to the low time complexity, our algorithm can be used with different mobile applications for real-time processing of shadows. The proposed approach has wide range of applications including but not limited to the detection of artifacts in image frames and real-time shadow detection. However, the proposed method has issues with color documents and natural scenes. Our future direction of research involves addressing these issues.

Acknowledgment

The authors would like to thank Ashutosh Mehra and Sachin Soni of Adobe for their review comments and suggestions in making this work possible.

References

[1] Z. Chen, L. Wan, L Zhu, J. Shen, H. Fu, W. Liu, and J Qin, “Triple-cooperative video shadow detection,“ IEEE Conference on Computer Vision and Pattern Recognition, pp. 2715-2724, 2021.

[2] T. Nagae, R. Abiko, T. Yamaguchi, and M. Ikehara, “Shadow Detection and Removal Using GAN,“ 28th European Signal Processing Conference, pp. 630-634, 2021.

[3] D. V. Tropin, S. A. Ilyuhin, D. P. Nikolaev, and V. V. Arlazarov, “Approach for document detection by contours and contrasts,“ 25th International Conference on Pattern Recognition, pp. 9689-9695, 2021.

[4] P. Mondal, A. Shara and A. Bal, “Image shadow detection using multiple images,“ U.S. Patent 10997453, May 4, 2021.

[5] J. Wang, X. Li, and J. Yang, “Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal,“ IEEE Conference on Computer Vision and Pattern Recognition, pp. 1788-1797, 2018.

[6] V. Shah and V. Gandhi, “An iterative approach for shadow removal in document images,“ IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1892-1896, 2018.

[7] S. Bako, S. Darabi, E. Shechtman, J. Wang, K. Sunkavalli, and P. Sen, “Removing shadows from images of documents,“ Asian Conference on Computer Vision. Springer, pp. 173–183, 2016.

[8] T. Zhou, P. Krahenbuhl, and A. A. Efros, “Learning data-driven reflectance priors for intrinsic image decomposition,“ Proceedings of the IEEE International Conference on Computer Vision, pp. 3469–3477, 2015.

[9] M. Gryka, M. Terry, and G. J. Brostow, “Learning to remove soft shadows,“ ACM Transactions on Graphics, vol. 34, no. 5, pp. 1-15, 2015.

[10] H. Gong and D. P. Cosker, “Interactive shadow removal and ground truth for variable scene categories,“ BMVC 2014-Proceedings of the British Machine Vision Conference, 2014.

[11] S. Bell, K. Bala, and N. Snavely, “Intrinsic images in the wild,“ ACM Transactions on Graphics, vol. 33, no. 4, pp. 1-12, 2014.

[12] R. Guo, Q. Dai, and D. Hoiem, “Paired regions for shadow detection and removal,“ IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 12, pp. 2956–2967, 2013.

[13] D. M. Oliveira, R. D. Lins, and G. Silva, “Shading removal of illustrated documents,“ International Conference Image Analysis and Recognition, pp. 308–317, 2013.

[14] Q. Zhao, P. Tan, Q. Dai, L. Shen, E. Wu, and S. 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.

[15] Q. Yang, K. Tan, and N. Ahuja, “Shadow removal using bilateral filtering,“ IEEE Transactions on Image processing, vol. 21, no. 10, pp. 4361–4368, 2012.

[16] G. D. Finlayson, M. S. Drew, and C. Lu, “Entropy minimization for shadow removal,“ International Journal of Computer Vision, vol. 85, no. 1, pp. 35–57, 2009.

[17] D. Bradley and G. Roth, “Adaptive thresholding using the integral image,“ Journal of Graphics Tools, vol. 12, no. 2, pp. 13-21, 2007.