# Removing Shadows from Images of Documents

Steve Bako[1]([✉]), Soheil Darabi[2], Eli Shechtman[2], Jue Wang[2],
Kalyan Sunkavalli[2], and Pradeep Sen[1]

[1] University of California, Santa Barbara, Santa Barbara, CA, USA
stevebako@umail.ucsb.edu
[2] Adobe Research, Seattle, WA, USA

**Abstract.** In this work, we automatically detect and remove distracting shadows from photographs of documents and other text-based items. Documents typically have a constant colored background; based on this observation, we propose a technique to estimate background and text color in local image blocks. We match these local background color estimates to a global reference to generate a *shadow map*. Correcting the image with this shadow map produces the final unshadowed output. We demonstrate that our algorithm is robust and produces high-quality results, qualitatively and quantitatively, in both controlled and real-world settings containing large regions of significant shadow.

## 1 Introduction

Images of documents, receipts, menus, books, newspapers, flyers, signs, and other text are frequently captured. Whether we are sending a page from a textbook highlighting important information, taking a picture of an ancient engraving while on vacation, or just saving an illustration from a loved one, such images show up in our everyday lives. However, these images are highly susceptible to shadows due to occlusions of ambient light by the photographer or other objects in the environment. These shadows cause distracting artifacts that can make an image difficult to interpret or use.

In this paper, we propose an automatic technique to detect and remove shadows from images of documents. Our main observation is that documents typically have a constant colored background; for example, the actual color of the paper typically does not change throughout a document. However, illumination effects like shadows and shading cause changes in observed images intensities. Our technique detects these changes and enforces a consistent background color to produce the unshadowed output. In particular, we estimate text and background colors in local blocks of the image and generate a *shadow map* that uses a per-pixel gain to match these local background estimates to a global reference. We evaluate the robustness of our approach, both quantitatively and qualitatively, on a variety of controlled and real-world examples.

## 2   Previous Work

There are two main categories of work in shadow removal. The first category focuses on removing shadows from general images, such as typical outdoor pictures, that have strong distracting shadows. For example, Guo et al. [1] remove shadows from natural images by finding corresponding shadow and non-shadow regions and performing a per-pixel relighting. Gong et al. [2] demonstrate even more robust results by using manually specified well-lit and shadow regions. A recent method from Gryka et al. [3] uses a learning approach with user-provided brush strokes to relight the shadow regions appropriately. Moreover, there are intrinsic imaging approaches that separate an image into its reflectance and shading components [4–8]. Lastly, there are algorithms that use shadow estimation for a specific application such as video relighting [9] or shape recovery [10]. All these methods tend to have artifacts when applied to document images.

A second category of techniques has been specifically developed to remove shadows from document images. Some such methods [5,11,12], inspired by general intrinsic approaches, correct geometric distortions and can estimate illumination within their framework to address shading artifacts. Our approach is more similar to the state-of-the-art method of Oliveira et al. [13], where a constant color for the document background is assumed to generate a gain map. However, they detect background-only regions and interpolate the remaining areas of the gain map. Therefore, their method can fail to remove shadows when excessive interpolation is required to fill in the holes of the gain map. Our approach, on the other hand, can effectively estimate the gain map in text and background regions, so it does not suffer from interpolation inaccuracies. Finally, Adobe Acrobat uses an "Enhance" feature [14] on images of documents that is typically used to brighten dark images. However, since this applies a global correction, it fails to remove local shadow regions and leaves residual shadows. Our approach performs analysis on small overlapping blocks throughout the document and is thus able to correct localized shadows.

Finally, there are image binarization methods that segment an image into black and white and discard all color information [15–22]. We note that, although related to our work, these approaches have the fundamentally different goal of creating a binary image that is more effective for optical character recognition (OCR) applications. On the other hand, we aim to improve documents by removing their shadows while still keeping the same color and tone as the original. Thus, we view this field of work as orthogonal to ours. In fact, in Sect. 5, we demonstrate how our method can be used as a pre-process to improve binarization techniques and thus the OCR applications that use them.

## 3   Algorithm

Our main observation is that documents tend to have a constant colored background throughout, so the unshadowed output should have this property as well. We propose to apply a factor, $\boldsymbol{\alpha}_i$, as determined by our computed shadow map

at each of the $i$ pixels in the input image in order to match the local background color intensities with the global reference color. Specifically, we calculate:

$$\tilde{\mathbf{c}}_i = \frac{\mathbf{c}_i}{\boldsymbol{\alpha}_i}, \tag{1}$$

where $\mathbf{c}_i$ and $\tilde{\mathbf{c}}_i$ are the RGB color intensities of the shadowed input and the unshadowed output at pixel $i$, respectively.

We find the local and global background intensities in the document and normalize the local background intensities by the global reference color to generate the per-pixel RGB shadow map, $\boldsymbol{\alpha}$. Applying this shadow map (Eq. 1) to the input image produces the final result.

**Local and Global Background Colors:** To find the local background colors, we begin by dividing the input image into small overlapping blocks. In each of these blocks, we cluster the pixel intensities into two clusters that we label as either the paper background or text. For clustering both the local and global data, we used Gaussian mixture models (GMM) fit with Expectation-Maximization (EM) and initialized with k-means clustering. In general, documents typically have dark colored text on a bright background. Based on this, we assign the cluster center with the higher mean as the local background RGB color, $\boldsymbol{\ell}_i$, where $i$ denotes the pixel at the center of the current block. In the case of a constant colored block (i.e., all background), the clusters have almost identical averages so selecting the higher one is still valid.

Next, we find the global reference background color. We take the pixel intensities from the entire input, rather than a local region, and cluster them into two categories of paper and text, as before. Again, the cluster mean that has a higher value is labeled as the background intensity. Finally, we search all of the intensities in the original input and assign the one closest to the background cluster mean as the final global background RGB reference, $\mathbf{g}$. Note that although the cluster mean could be used in this step instead of the closest intensity, we empirically found that this approach slightly improves results.

**Computing Shadow Map:** We expect the global background color to be the true background color. The local background color deviates from this because of illumination effects such as shadows and shading. In order to remove the influence of illumination, we compute the ratio of the local and global background colors to generate the shadow map as:

$$\boldsymbol{\alpha}_i = \frac{\boldsymbol{\ell}_i}{\mathbf{g}}, \tag{2}$$

where $\boldsymbol{\ell}_i$ is the local background intensity at pixel $i$ and $\mathbf{g}$ is the global background reference for all the pixels. Moreover, $\boldsymbol{\alpha}_i$ maps each input pixel to the reference background color and, when applied to the input image (Eq. 1), produces the final unshadowed result.

Figure 1 shows the general framework of our proposed approach along with an example of a shadow map, which accurately detects regions of shadow and their relative intensities.
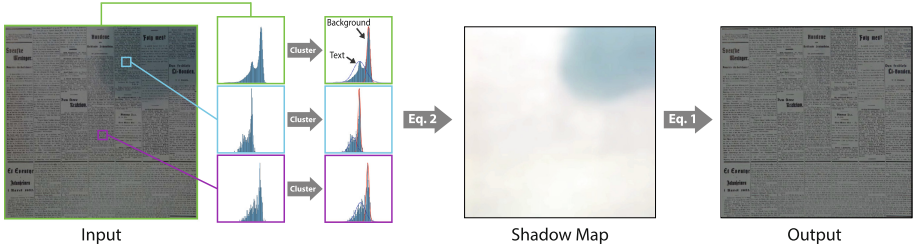
**Fig. 1.** Our approach takes an input document image that contains shadow and produces a high-quality shadow-free result. We analyze image intensities in local blocks (light blue and purple) and partition them into two clusters – background and text. We also do this for the whole image (green) to estimate a global background reference. Applying Eq. 2 to the background mean clusters results in a shadow map that is used in Eq. 1 to produce the final unshadowed result (Color figure online).

**Implementation Details:** We implemented our algorithm in C++ and it takes roughly 2 s to process a $1024 \times 1024$ image, with clustering being the most costly sub-process. Thus, for acceleration, we randomly sample 150 pixels in each block ($21 \times 21$) for local clustering and 1000 pixels throughout the entire image for global clustering. For further speed-up, we do not perform local clustering at each pixel. Instead, we only consider pixels at specific strides (i.e., 20 in our implementation) for calculating the local background intensities. Thus, our shadow map is stride times smaller than the input image. To get the full resolution shadow map, we upsample using an $8 \times 8$ Lanczos filter. Note that since we use a stride, the low resolution shadow map can have some slight noise due to differences in cluster means, so we first apply a $3 \times 3$ median filter followed by a Gaussian filter ($\sigma = 2.5$) to smooth it out and avoid small blotchy artifacts in the final result. Since illumination tends to vary smoothly and generates soft shadows in practice, computing a downsampled shadow map does not adversely affect our results. Finally, text can have a lot of variation in intensity values and two clusters is not always sufficient to capture accurate statistics. Thus, we found that using three means for both the local and global clustering worked better than two.

## 4   Results

We begin by validating the correctness of our method by creating a dataset in a controlled environment with a set illumination and obtain a ground truth image for comparison. Specifically, we captured 81 shadow images of 11 documents each with 5–9 variations of shadow intensity and shape using a Canon 5D Mark II DSLR camera, a tripod, a photography lamp, and light-blocking objects. Furthermore, for each document, we captured the ground truth without shadow, where the light-blocking object was removed while illumination was

| | Input | Bell et al. [2014] | Gong et al. [2014] | Acrobat "Enhance" | Oliveira et al. [2013] | Ours | Ground Truth |
| MSE: | | 298.81 | 1445.18 | 252.24 | 51.01 | 44.69 | |

**Fig. 2.** We show an example from a controlled environment used to validate the correctness of our approach relative to the ground truth. We also provide the results and MSE scores of related approaches for comparison. In general, our output has minimal artifacts and is closest to ground truth in terms of average and median MSE (Table 1). See the supplemental materials for additional results from this controlled setting.

**Table 1.** A comparison with previous approaches of average and median MSE on our controlled dataset consisting of 81 images (11 documents and 5–9 variations of shadow).

| Method | Avg. MSE | Median MSE |
| --- | --- | --- |
| Bell et al. [6] | 125.44 | 119.94 |
| Gong et al. [2] | 390.98 | 172.57 |
| Pilu et al. [18] | 67.38 | 53.54 |
| Wagdy et al. [21] | 74.06 | 43.73 |
| Oliveira et al. [13] | 23.08 | 19.01 |
| Ours | 22.26 | 18.45 |

kept constant. In Fig. 2, we demonstrate that we can match ground truth closely with minimal artifacts. Note that to avoid brightness differences we match the average color of our output to that of the ground truth.

Since we have the ground truth image, we can also provide quantitative comparisons. In Fig. 2, we show visual and MSE comparisons for the intrinsic approach of Bell et al. [6], the general shadow removal method of Gong et al. [2], the Adobe Acrobat "Enhance" feature [14], and the state-of-the-art algorithm in document de-shadowing by Oliveira et al. [13] on one specific example. In addition, Table 1 reports the overall performance of each method[1] over our entire dataset. Table 1 also includes statistics for two additional Retinex-based approaches [18,21] that are designed for document binarization rather than our application. In general, we found that these two approaches can leave residual shadow and blur out detail; this can be observed in the results in the supplementary material. As shown, our method performs the best relative to ground truth in terms of MSE. It is worth noting that, since we captured this dataset in a controlled environment with simple conditions (e.g., a single light source, flat documents, etc.), our dataset is not as difficult as the typical scenarios found in

---

[1] We exclude Acrobat "Enhance" from the table as it requires manual interaction for each image.
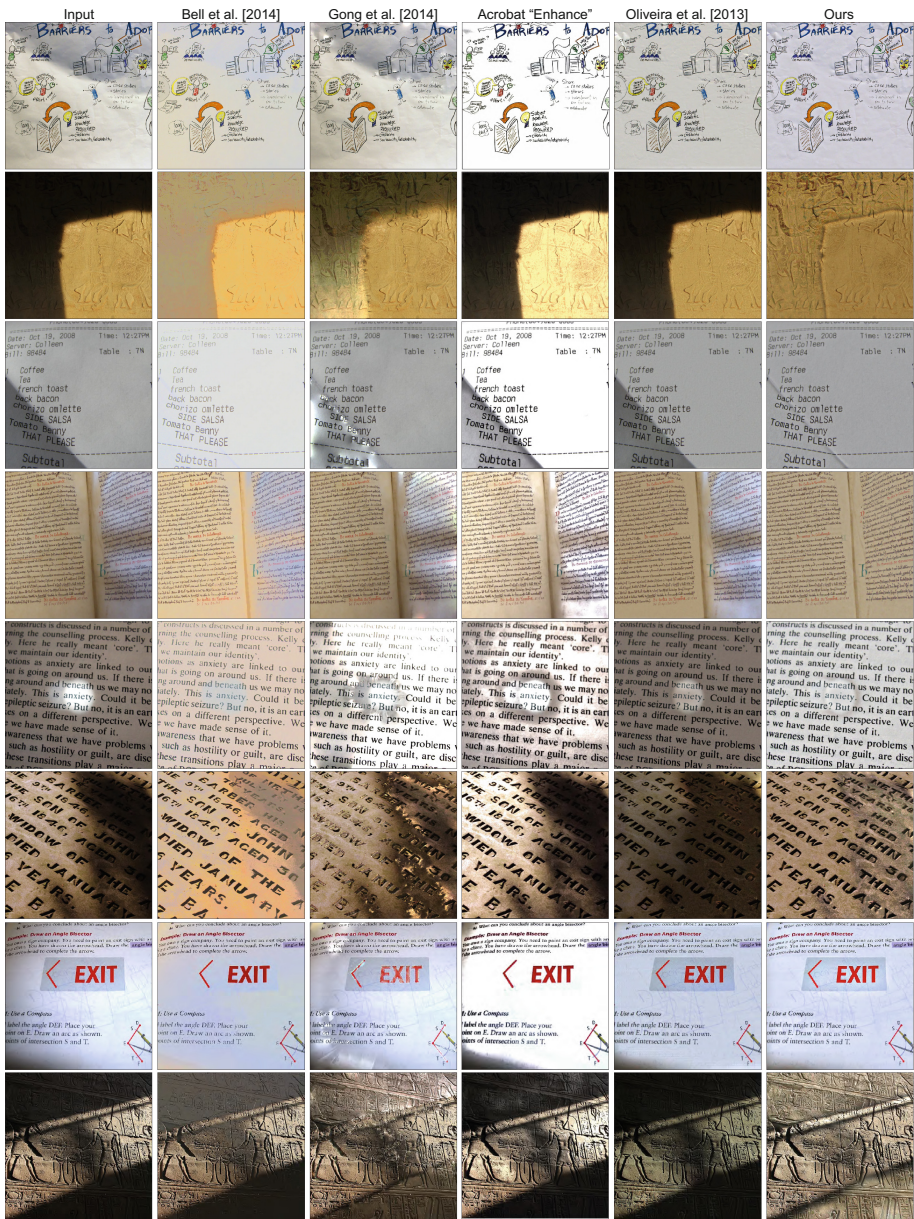
**Fig. 3.** We provide comparisons on challenging real-world examples from Flickr. Our method consistently produces high-quality results where other approaches fail. See the supplemental materials for full resolution comparisons on our complete real-world dataset.

practice. Thus, although we significantly outperform other approaches, we have only slightly better MSE relative to the method of Oliveira et al. [13].

In Fig. 3, we show comparisons for a subset of our more challenging real-world dataset obtained from Flickr. Note the shadows appear as is in the original image and were not altered in any way. As shown, all of the previous methods produce significant artifacts for these examples. The reflectance image from Bell et al. [6] loses some of the text to the shading layer and still has residual shadows. The general method from Gong et al. [2] cannot accurately estimate the statistics of the shadow region, despite the user interaction, and introduces saturation artifacts. The Adobe Acrobat "Enhance" feature [14] applies a global transformation on the image, so it fails to remove the local shadows. Finally, Oliveira et al.'s method [13] relies on interpolation so it cannot always remove the shadows. On the other hand, our approach works robustly across a wide range of difficult examples and generates high-quality results. Note that comparisons for all 16 images of our real-world dataset can be found in the supplemental materials along with additional comparisons to Retinex methods [18,21].

## 5   Discussion

In Fig. 4, we demonstrate that our algorithm can also be used as a pre-process for image binarization, which is typically used for OCR applications. Here, we compare the binarization method of Su et al. [22] with and without our shadow removal on an example from the Document Image Binarization Contest 2013 (DIBCO 2013) [23] dataset. When using our approach as a pre-process, there is a clear improvement in both visual quality and PSNR[2]. Note this binarization
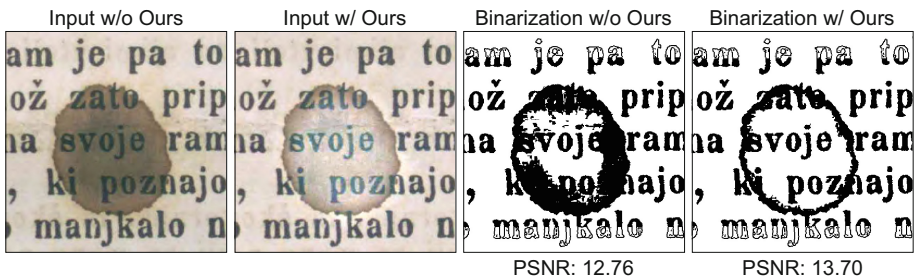


**Fig. 4.** Our approach can be used as a pre-process to improve the performance of OCR applications. We take a single example from the DIBCO 2013 [23] dataset containing a shadow (leftmost image) and run our shadow removal algorithm on it (center left image). We see a clear improvement both perceptually and in PSNR when applying the image binarization method of Su et al. [22] on these inputs (two rightmost images).

---

[2] The DIBCO 2013 dataset provides ground truth images to evaluate the performance of binarization algorithms. Furthermore, almost all of the images in this dataset are already shadow-free, so we report PSNR on just a single example containing shadow.
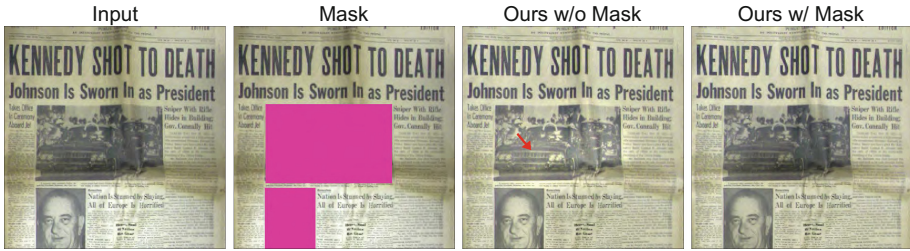
| Input | Mask | Ours w/o Mask | Ours w/ Mask |
|---|---|---|---|



**Fig. 5.** Our algorithm is designed for document images and can introduce artifacts when non-text elements (e.g., figures, charts, pictures, etc.) are present. However, we are able to handle such cases by taking as input a user-defined mask that specifies such regions. We ignore these areas when clustering and perform interpolation to fill the holes and obtain our final shadow map. We show how this mask helps reduce artifacts in this real-world example.

method is already robust against illumination changes like shadow, yet there is still a benefit from running our shadow removal algorithm. This improvement will help the performance of OCR applications that take binary images as input.

One limitation of our method is that we assume that the input documents contain only intensities that correspond to either paper or text (i.e., no figures, pictures, graphs, etc.) so that we can cluster local and global pixel intensities into two categories: paper background and text. If the pixel intensities correspond to additional regions (e.g., a picture in the document), then the cluster means will be biased, cause incorrect background estimates, and generate overly bright or overly dark regions in the final result.

However, figures and pictures typically only account for a small region of the document. Thus, we can take a user-defined mask specifying these regions and ignore them when performing our shadow map calculation. After we have our shadow map, we can use interpolation (e.g., natural neighbor interpolation) to fill in the holes left by the mask and use it to generate the result. As shown in Fig. 5, using a mask allows us to improve results for such cases.

If the document background color changes throughout the document then the global reference background intensity that we find is unreliable and the final result will have brightness or color artifacts, as shown in Fig. 6(a). Finally, with intense hard shadows, there can be artifacts at the shadow boundaries since local blocks straddle regions with both well-illuminated and shadowed text and background. In such areas our cluster means are inaccurate creating a visible boundary between the well-illuminated and unshadowed regions (Fig. 6(b)). See the supplemental materials for full resolution examples of the various limitations.
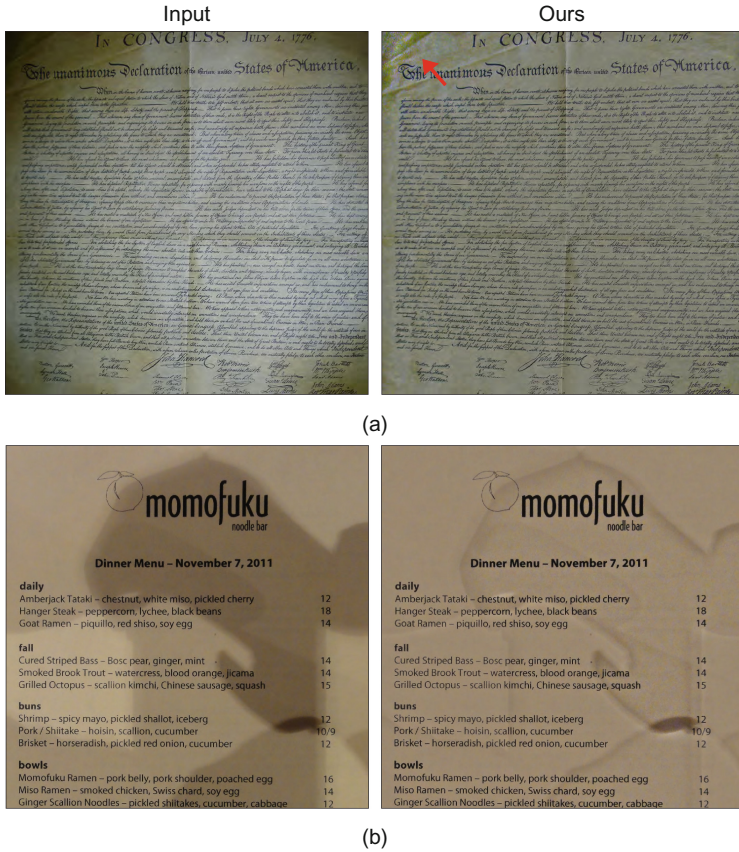
Input                                          Ours



(a)



(b)

**Fig. 6.** Our approach has some limitations. For example, in (a), the background color changes throughout the document, so we produce slight color and brightness artifacts. In (b), we show that in cases with intense hard shadows, our method leaves small residual artifacts at the shadow boundaries. These artifacts result from unreliable clustering at either the global (a) and local (b) levels.

# 6    Conclusion

We have presented an approach for removing shadows from images of text (e.g., documents, menus, receipts) by generating a shadow map, or per-pixel scaling, that matches local background colors to a global reference. Our approach works robustly, as demonstrated qualitatively and quantitatively, on a wide range of examples containing large amounts of strong shadow in both controlled and real-world settings.

Commons 2.0 License without modification. The title and photographers from Flickr (unless otherwise noted) in order of appearance are: *Open Textbook Summit 2014* by BCcampus_News, *Army in the Shadows*, *Army in the Light* by Cuzco84, *That Please* by Kimli, *Medieval text in the Christ Church Archive* by -JvL-, *Untitled* by Jacek.NL, *Cartmel Priory* by Rosscophoto, *Transfer Damaged Textbook* by Enokson, *Untitled* by Colin Manuel, *find* by PHIL, Declaration of Independence photo by taliesin at Morguefile.com (Morguefile License), and *Momofuku - Menu w/ Shadow puppets* by Lawrence. Please see supplementary materials for links to the images and license.

# References

1. Guo, R., Dai, Q., Hoiem, D.: Paired regions for shadow detection and removal. IEEE Trans. Pattern Anal. Mach. Intell. **35**, 2956–2967 (2013)
2. Gong, H., Cosker, D.: Interactive shadow removal and ground truth for variable scene categories. In: Proceedings of the British Machine Vision Conference. BMVA Press (2014)
3. Gryka, M., Terry, M., Brostow, G.J.: Learning to remove soft shadows. ACM Trans. Graph. **34**(5), 153:1–153:15 (2015)
4. Zhao, Q., Tan, P., Dai, Q., Shen, L., Wu, E., Lin, S.: A closed-form solution to retinex with non-local texture constraints. PAMI **34**(7), 1437–1444 (2012)
5. Yang, Q., Tan, K.H., Ahuja, N.: Shadow removal using bilateral filtering. IEEE Trans. Image Process. **21**, 4361–4368 (2012)
6. Bell, S., Bala, K., Snavely, N.: Intrinsic images in the wild. ACM Trans. Graph. (SIGGRAPH) **33**(4), 159:1–159:12 (2014)
7. Barron, J.T., Malik, J.: Shape, illumination, and reflectance from shading. TPAMI **37**(8), 1670–1687 (2015)
8. Zhou, T., Krahenbuhl, P., Efros, A.A.: Learning data-driven reflectance priors for intrinsic image decomposition, pp. 3469–3477 (2015)
9. Sunkavalli, K., Matusik, W., Pfister, H., Rusinkiewicz, S.: Factored time-lapse video. ACM Trans. Graph. **26** (2007). Proceedings of the SIGGRAPH
10. Abrams, A., Hawley, C., Pless, R.: Heiometric stereo: shape from sun position. In: European Conference on Computer Vision (ECCV) (2012)
11. Brown, M.S., Tsoi, Y.C.: Geometric and shading correction for images of printed materials using boundary. Trans. Img. Proc. **15**, 1544–1554 (2006)
12. Zhang, L., Yip, A.M., Tan, C.L.: Removing shading distortions in camera-based document images using inpainting and surface fitting with radial basis functions. In: 9th International Conference on Document Analysis and Recognition, ICDAR 2007, 23–26 September, Curitiba, Paraná, Brazil, pp. 984–988 (2007)
13. Oliveira, D.M., Lins, R.D., França Pereira e Silva, G.: Shading removal of illustrated documents. In: Kamel, M., Campilho, A. (eds.) ICIAR 2013. LNCS, vol. 7950, pp. 308–317. Springer, Heidelberg (2013). doi:10.1007/978-3-642-39094-4_35
14. Acrobat, A.: Enhance document photos captured using a mobile camera (2016). https://helpx.adobe.com/acrobat/using/enhance-camera-images.html
15. Otsu, N.: A threshold selection method from gray-level histograms. Automatica **11**, 23–27 (1975)
16. Pal, N.R., Pal, S.K.: A review on image segmentation techniques. Patt. Recogn. **26**, 1277–1294 (1993)
17. Sauvola, J., Pietikäinen, M.: Adaptive document image binarization. Patt. Recogn. **33**, 225–236 (2000)

18. Pilu, M., Pollard, S.: A light-weight text image processing method for handheld embedded cameras (2002)
19. Shi, Z., Govindaraju, V.: Historical document image enhancement using background light intensity normalization. In: 2004 Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, vol. 1, pp. 473–476. IEEE (2004)
20. Gatos, B., Pratikakis, I., Perantonis, S.J.: Adaptive degraded document image binarization. Patt. Recogn. **39**, 317–327 (2006)
21. Wagdy, M., Faye, I., Rohaya, D.: Fast and efficient document image clean up and binarization based on retinex theory. In: 2013 IEEE 9th International Colloquium on Signal Processing and its Applications (CSPA), pp. 58–62. IEEE (2013)
22. Su, B., Lu, S., Tan, C.L.: Robust document image binarization technique for degraded document images. IEEE Trans. Image Process. **22**, 1408–1417 (2013)
23. Pratikakis, I., Gatos, B., Ntirogiannis, K.: ICDAR 2013 document image binarization contest (DIBCO 2013). In: 2013 12th International Conference on Document Analysis and Recognition (ICDAR), pp. 1471–1476. IEEE (2013)