sensors

MDPI

Article

Local Water-Filling Algorithm for Shadow Detection and Removal of Document Images

Bingshu Wang [1] **and C. L. Philip Chen** [2,3,*]

[1] School of Software, Taicang Campus, Northwestern Polytechnical University, Suzhou 215400, China; wangbingshu@nwpu.edu.cn
[2] School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China
[3] Faculty of Science and Technology, University of Macau, Macau 999078, China
[*] Correspondence: philip.chen@ieee.org

Received: 31 October 2020; Accepted: 1 December 2020; Published: 4 December 2020

check for updates

Abstract: Shadow detection and removal is an important task for digitized document applications. It is hard for many methods to distinguish shadow from printed text due to the high darkness similarity. In this paper, we propose a local water-filling method to remove shadows by mapping a document image into a structure of topographic surface. Firstly, we design a local water-filling approach including a flooding and effusing process to estimate the shading map, which can be used to detect umbra and penumbra. Then, the umbra is enhanced using Retinex Theory. For penumbra, we propose a binarized water-filling strategy to correct illumination distortions. Moreover, we build up a dataset called optical shadow removal (OSR dataset), which includes hundreds of shadow images. Experiments performed on OSR dataset show that our method achieves an average *Error Ratio* of 0.685 with a computation time of 0.265 s to process an image size of 960 × 544 pixels on a desktop. The proposed method can remove the shading artifacts and outperform some state-of-the-art methods, especially for the removal of shadow boundaries.

Keywords: local water-filling algorithm; topographic surface; shading map; document image; penumbra removal

1. Introduction

Optical shadows appear out of nowhere in the images captured from camera sensors [1–4]. They are generated when light sources are occluded by static or moving objects [5–8]. In most cases, shadows are regarded to be useless and need to be removed from images. One of the most used engineering applications is to remove optical shadows from document images.

With the increasing use and popularization of smart phones, people are more likely to use them as a mainstream document capture device rather than a conventional scanner. As a result, many document images are captured under various situations and conditions such as indoor and outdoor. Since the occlusion of illumination sources in environments is inevitable, shadows usually appear in the document images [1,9] with different types: weak, moderate, strong or nonuniform [10,11].

When document images are cast by shadows, the occluded regions become darker than before. It is observed that the text is always printed in black on the documents. Specifically, when the darkness of the shadows is similar to that of text, it will generate poor-quality text [9,11]. The shadows may make the perception of documents uncomfortable to the human eye and cause the degradation of text in documents or notes, which will result in difficulties for text binarization and recognition [12,13]. Therefore, removing shadows from document images not only helps generate clear and easy-to-read text [14], but also makes document binarization [15,16] and recognition tasks [17–19] possible.

Over the past decade, shadow removal is playing a growing role in digitized document application and attracting the attention of many researchers. Bradley et al. [20] proposed an adaptive threshold technique for binarization utilizing the integral image that is calculated from the input image. It is sensitive to a slight illumination change but it cannot remove the boundaries of strong shadows. Bako et al. [14] came up with a strategy that estimates local text and background color in a block. They removed shadows by generating a global reference and a shadow map. Shah et al. [21] considered shadow removal as an estimation problem of shading and reflectance components of the input image. An iterative procedure was explored to handle hard shadows. However, the large number of iteratations required too many calculations.

The method proposed by Kligler et al. [11] developed a technique of 3D point cloud transformation for visibility detection. It aims to generate a new representation of an image that can be used in common image processing algorithms such as document binarization [22,23] and shadow removal [14]. However, the transformation process requires huge computational power. The approach proposed by Jung et al. [24] explored a water-filling method to rectify the illumination of digitized documents by converting the input image into a topographic surface. It is implemented based on the $YCbCr$ color space and only takes the luminance component into account. It achieves good performance on weak or medium shadows. However, this method tends to produce degraded color results for scenes with strong shadows.

On one hand, shadows need to be removed. On the other hand, obvious color artifacts should be avoided after shadow removal. Zhang et al. proposed a prior-based [25] method and learning-based [26] method for removing color artifacts. Barron et al. proposed a fast fourier color constancy method [27] and a convolutional color constancy method [28] to recover a white-balanced image and make the image natural-looking. These methods are expected to provide potential means to correct the non-uniform illumination and color artifacts. In addition, there are other methods proposed to detect shadows [29–32], and remove shadows from document images or natural images [33–38], which is expected to benefit many text detection and recognition approaches reviewed in [39].

Physically, shadows can be divided into two parts: umbra and penumbra [40]. For weak or medium shadows, the umbra and penumbra have fuzzy boundaries and can both be handled by the methods mentioned above. However, for strong shadows, these methods face challenges. There are two possible reasons for this. On one hand, shadow strength is difficult to estimate for shadow regions. On the other hand, many shadow points belong to shadow boundaries and they are very similar with surrounding texts.To remove these shadows, some works have been completed. Some datasets have been created for research on document shadow removal, for example, the Adobe [14] and HS datasets [21]. However, only a few images in these datasets have strong shadows. Therefore, it is necessary to build up a dataset that includes more images with strong shadows.

Our motivation is to explore a means to remove shadows from document images. In this paper, we solve the problem by mapping an image into a topographic surface, i.e., unshadowed region can be regarded as plateau, umbra as catchment basin, and the penumbra as ridge between plateau and basin, which is shown in Figure 1. This paper devises a design to obtain a shading map using local water-filling (LWF), which helps to estimate shadow strength. To remove shadow boundaries, this paper proposes a local binarized water-filling (LBWF) algorithm to correct illumination distortions of document images. Moreover, we create a dataset that includes many images with strong shadows.

The contributions of this paper are as follows:

(1) This paper designs a local water-filling approach to estimate a shading map using a stimulation of flooding and effusing processes (Section 2.1). This strategy is able to produce an effective map that indicates the shading distribution in a document image.

(2) This paper develops a local binarized water-filling algorithm for penumbra removal (Section 2.4). This provides an effective means to remove strong shadow boundaries, which is a difficult problem for many methods due to the high similarity between penumbra and text.

(3) We create a dataset called OSR for shadow removal in document images, including the controlled illumination environment and natural scenes. Specially, the dataset contains some typical scenes with strong shadows (Section 3.1).

(4) The proposed method's efficiency is superior to some state-of-the-art approaches as the experiments are conducted on an image with a size of 960 × 544 pixels.

The paper is organized as follows. Section 2 presents the proposed method. Section 3 gives the experimental results and analysis. Section 4 concludes this work.

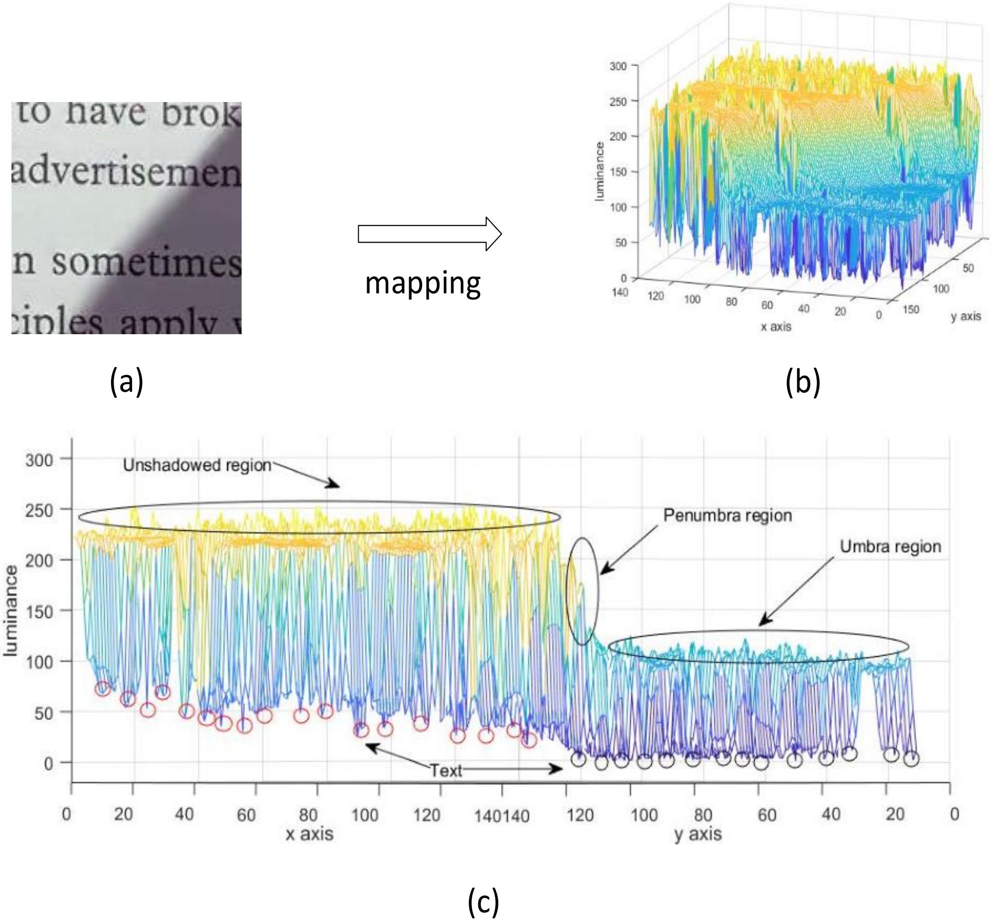

Figure 1. The mapping illustration from one document image with shadows to its topographic structure. (**a**) one document image with shadows, (**b**) the visual topographic structure of image (**a**), (**c**) the clarification of mapping process from image (**a**) to image (**b**): the unshadowed region can be regarded as plateau, umbra as catchment basin, the penumbra as ridge between plateau and basin, and text as the lowest points.

2. The Proposed Method

The flowchart of the proposed method is presented in Figure 2. Firstly, the proposed local water-filling (LWF) algorithm receives an input image with shadows and generates a shading map (see Figure 2b) which represents local background colors of the input image. The shading map can be used to detect umbra (the red) and penumbra (the purple) (as shown in Figure 2c). Then, the umbra can be relighted according to Retinex theory (Figure 2d). Finally, a local binarized water-filling-based (LBWF-based) algorithm was designed to remove the shadow boundaries and produce an unshadowed image (Figure 2e). Notably, Figure 2 shows the topographic structures of the image (a), (b), (d) and (e), indicating how the topographic surface changes.

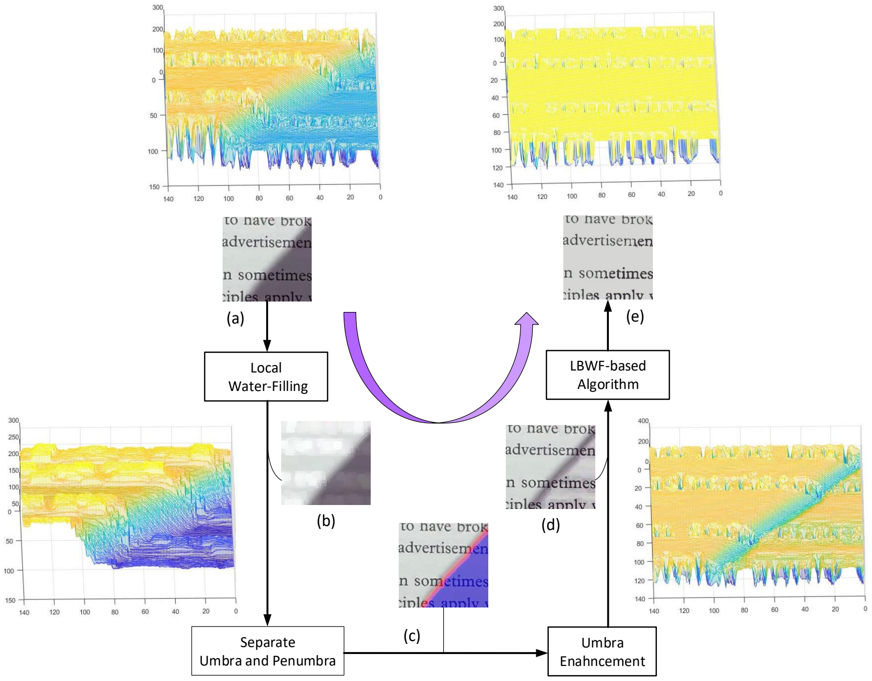

Figure 2. The flowchart of optical shadows removal. (**a**) input image, (**b**) shading map, (**c**) the red represents umbra and the purple represents penumbra, (**d**) the image after umbra enhancement, (**e**) the output image without shadows.

2.1. Local Water-Filling Algorithm

In this section, we report a design to estimate a shading map of the input image using a local water-filling algorithm. It mainly includes two parts: a flooding and effusing part. This paper stimulates this process by solving three core problems: where does the "water" come from; where does the "water" flow out; how is the "water" stored. The proposed algorithm is modeled by figurative flowing of "water". Therefore, some variables need to be defined first before modeling our method.

We set $h(x)$ as the altitude of the topographic surface and $w(x,t)$ as the water level at a point of time t. For a point x_0, its overall altitude $K(x_0,t)$ is the sum of $h(x_0)$ and $w(x_0,t)$, i.e., $K(x_0,t) = h(x_0) + w(x_0,t)$. Figure 3 illustrates a one-dimensional model of plateau and basin. Specially, an essential constraint about $w(x,t)$ is given as follows

$$w(x,t) \geq 0, \forall t, \forall x \in I \tag{1}$$

where I is denoted as the domain of an image. To evaluate $w(x,t)$, the inflow and outflow of water are modeled by three parts as below.

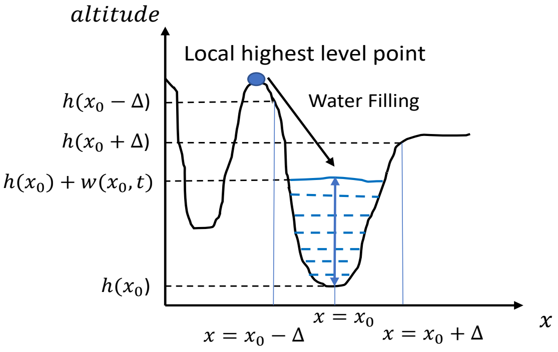

Figure 3. One-dimensional topographic model of a basin and its neighborhood, coupled with the water-filling direction.

Where does the "water" come from? The water is simulated at the pixel-wise in the input image, which is similar with the techniques developed by [24,41,42]. In our study, locality means that the water comes from the neighboring pixels, in other words, the pixel with the highest intensity (or altitude) is selected as water source. It is denoted by

$$h_m(x_0) = max\{h_x\}, h_x \in NeighboringPixels \tag{2}$$

$NeighboringPixels$ represents a number of neighboring pixels of point x_0. It can be concluded that $h_m(x_0) \geq h_x$. Thus, to meet Equation (1), the flooding process can be modeled by

$$w_f(x_0, t) = h_m(x_0, t) - K(x_0, t) \tag{3}$$

Where does the "water" flow out? We consider the effusion process through the pixel's surroundings in a dynamic changing manner. The effusing process for 1D case can be modeled by

$$\begin{aligned} w_e(x_0, t) \propto min\{K(x_0 + \Delta, t) - K(x_0, t), 0\} \\ +min\{K(x_0 - \Delta, t) - K(x_0, t), 0\} \end{aligned} \tag{4}$$

It can be seen that the $w_e(x_0, t)$ is non-positive, which represents the amount of effusion water for point x_0. The water only flows into the lower places.

How is the "water" stored? The change in water level depends on flood and effusion results, and it is the sum of the two components. Meanwhile, considering the previous water level, the final altitude of x_0 is formulated by an iterative form

$$K(x_0, t + \Delta) = K(x_0, t) + w_f(x_0, t) + \alpha \cdot w_e(x_0, t) \tag{5}$$

For a 2D image, the iterative update process of the overall altitude can be written as

$$\begin{aligned} K(x_0, y_0, t + \Delta_t) = K(x_0, y_0, t) + (h_m(x_0, y_0, t) - K(x_0, y_0, t)) \\ +\alpha \cdot \{min\{K(x_0 + \Delta_x, y_0, t) - K(x_0, y_0, t), 0\} \\ +min\{K(x_0 - \Delta_x, y_0, t) - K(x_0, y_0, t), 0\} \\ +min\{K(x_0, y_0 + \Delta_y, t) - K(x_0, y_0, t), 0\} \\ +min\{K(x_0, y_0 - \Delta_y, t) - K(x_0, y_0, t), 0\}\} \end{aligned} \tag{6}$$

where Δ_t represents the changing time, Δ_x and Δ_y are defined as distances from (x_0, y_0) to its neighboring pixels. The α is an important parameter that controls the speed of the effusion process. α should be set carefully and it is expected to be limited in a suitable ratio in order to store the water. For LWF, the parameter α should be no greater than 0.25 due to the use of four neighboring points. In practice, $\alpha = 0.22$ may provide a satisfactory result. The iteration process will come to an end if the difference between two continuous altitudes is small enough or it reaches the maximum iteration number. Three iterations is enough to generate a proper shading map that represents the local background color. The shading map can be used to separate umbra and penumbra.

2.2. Separate Umbra and Penumbra

The shading map in Figure 2b is an image with three channels. To obtain the umbra and penumbra mask, a series of steps are designed to reach the goal.

Firstly, for each channel, a medium filtering and a binary threshold operation are adopted to generate a binary image, indicating shadow regions and unshadowed regions. Then, three channels are merged together. For a point, at least one of the three channels must be classified as shadow. It will be regarded as an umbra point. The umbra mask can be obtained by the pixel classification one by one.

Next, umbra masking is performed on a succession of dilation operations, generating an expanded shadow mask. In practice, two times of dilation are expected to be enough. Finally, the expanded shadow mask is subtracted by the umbra mask, producing the penumbra mask. In Figure 2c, the blue and red represent umbra and penumbra, respectively.

2.3. Umbra Enhancement

For umbra enhancement, an effective strategy to correct illumination is to relight umbra based on Retinex theory [43]. It requires the calculation of an enhancement scale that can be expressed as a ratio between a global reference background color and a local background color. Let G be the global reference background intensity, it can be expressed by

$$G_i = \frac{1}{n} \sum L_i(x, y), (x, y) \in Unshadowed\ Region \tag{7}$$

where $i \in \{r, g, b\}$, n represents the number of pixels in an unshadowed region, $L(x, y)$ is the local background in Figure 2b. G is the global background color with three channels.

Then, the enhancement scale can be easily obtained through the equation $\eta(x, y) = \frac{G}{L(x,y)}$. Hence, umbra can be enhanced by a multiplication of a pixel's intensity and the enhancement scale $\eta(x, y)$.

Penumbra are located between umbra and lighted regions, and are generally regarded as the shadow boundaries. The penumbra varies widely and makes it difficult to estimate the enhancement scale. In this paper, we put forward a solution to solve the problem in the next section.

2.4. Local Binarized Water-Filling

To solve the issue associated with the penumbra, we propose an algorithm to correct the illumination distortions, called the local binarized water-filling algorithm (LBWF-based algorithm). The overall structure of LBWF is similar to that of LWF, but there are some differences. Two main differences between LBWF and LWF are the following: the iteration number of LBWF is one; the parameter α of the effusion process is set to one. This setting of parameters not only speeds up the effusion process, but also reduces background noise. It is able to produce different and significant results compared with LWF. Experiments indicate that LBWF is more likely to suppress the effects of penumbra and keep the integrity of text, which can be found in Figure 4c.

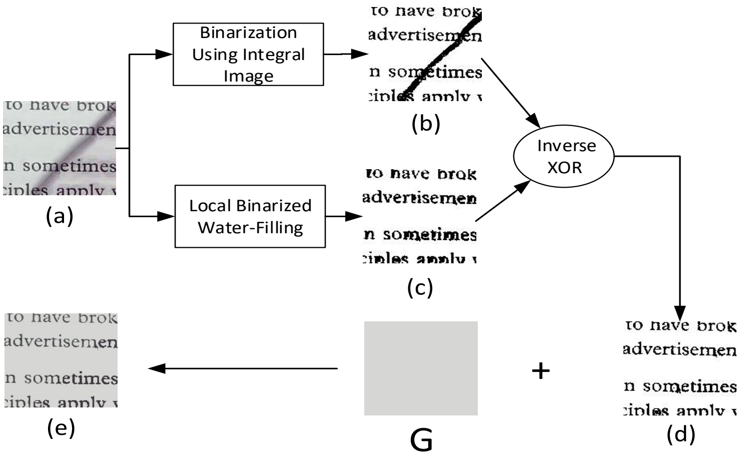

Figure 4. Shadow boundaries removal using the LBWF-based algorithm. (**a**) is the image from Figure 2b,d the binarization result of an adaptive thresholding method [20], (**c**) the binarization result of the proposed LBWF, (**d**) the inverse result of the XOR operation between the images (**b**) and the image (**c**), (**e**) the final result of shadow removal by combing the image (**d**) and the global background color G, which includes three channels and comes from Equation (7).

LBWF is able to produce a gray-level image with only text and background, which is indicated in Figure 4c. The penumbra between text lines can be suppressed well, which verifies the effectiveness of LBWF. To obtain a better result, a binary image (Figure 4b) is generated by the integral image-based method [20]. Then, an inverse XOR operation is carried out to produce a clearer image. Finally, the global background color G is combined with Figure 4d to generate an unshadowed result (Figure 4e). Overall, the algorithmic description is presented in the form of pseudocode, as shown in (Algorithm 1).

Algorithm 1 Algorithm of removing shadows from a document image.

```
Input: A document image with shadows: $\mathbf{I}$.
Output: An unshadowed image: $\mathbf{UI}$.
Obtain the shading map $\mathbf{S}$ using local water-filling algorithm.
    Split image $\mathbf{I}$ into three channels in RGB color space. For each channel, Equation (6) is carried out in an iterative manner three times. Threshold parameter $\alpha = 0.22$. Merge the results of three channels into a shading map $\mathbf{S}$.
Separate umbra and penumbra. With $\mathbf{S}$ obtained, median filtering and OTSU binarization are operated for three channels, generating binary masks $\mathbf{BI_B}$, $\mathbf{BI_G}$, $\mathbf{BI_R}$.
    A voting strategy of $\mathbf{BI_B}$, $\mathbf{BI_G}$, and $\mathbf{BI_R}$ is used to determine the shadow region mask $\mathbf{SR}$.
    $\mathbf{SR} :\leftarrow \phi(\mathbf{SR})$, $\phi$ represents the removal of the border noise with a size of 2 pixels.
    $UmbraMask = \psi(\mathbf{SR})$, Umbra mask is generated by an dilation operation $\psi$ on $\mathbf{SR}$.
    $DilatedUmbra \leftarrow \psi(UmbraMask)$ is carried out at least two times.
    $PenumbraMask = DilateUmbra - UmbraMask$ is to produce the penumbra mask.
Umbra enhancement. Calculate the average background intensity in the non-shadow region by Equation (7): $G_i = \frac{1}{n}\sum L_i(x,y), (x,y) \in Unshadowed\ Region$.
    Compute umbra enhancement: $\eta(x,y) = \frac{G}{L(x,y)}$.
    Remove umbra: $\mathbf{UI_{umbra}}(x,y) = \mathbf{I}(x,y) \times \eta(x,y)$
Penumbra removal using LBWF-based algorithm. Obtain a binarization image $\mathbf{B_1}$ from $\mathbf{UI_{umbra}}$ using integral image technique, which corresponds to Figure 4b.
    Compute a new binarization result $\mathbf{B_2}$ from $\mathbf{UI_{umbra}}$ using the Local Binarized Water-Filling technique. Split image $\mathbf{UI_{umbra}}$ into three channels in the RGB color space. For each channel, Equation (6) are carried out once. Threshold parameter $\alpha = 1$.
    $\mathbf{B_3} = InverseXOR(\mathbf{B_1}, \mathbf{B_2})$, this corresponds to Figure 4d.
    $\mathbf{UI} \leftarrow \mathbf{B_3} + G$
return UI
```

3. Experimental Analysis

Visual and quantitative results are provided in this section. Our method runs on a PC with 3.5 GHz Xeon machine, and it is implemented by C++ and open source in computer vision (OpenCV) under the Visual Studio 2015 development environment. We compared our approach with two approaches whose codes are available online [11,24]. All the methods are performed on the same PC with a Windows 10 Operating System and 64GB RAM installed, and each method utilizes a suite of parameters. Each method runs five times to obtain the average running time.

3.1. Dataset

Previous researchers have proposed related datasets for shadow removal in document images, for example, the Adobe [14] and HS datasets [21]. To verify the proposed method's effectiveness, these datasets are selected for the evaluation. Since there are a few strong shadow datasets available for optical shadow removal, we create one for evaluation, which is called the OSR dataset. It consists of two parts: the first part contains 237 images (controlled group, OSR_CG) with ground truth which are created under a control environment, and the other has 24 images (natural group,OSR_NG) without ground truth which are obtained from the Internet or captured under natural scenes.

The OSR_CG was created in a room. The documents were taken from books, newspapers, booklets, etc. They are typical documents. In the process of creating the dataset, two persons worked together. Firstly, the document was fixed on a desk, and a smart phone holder was adjusted to ensure our iPhone XR was well positioned to take photos. Then, one person created the source light using a lamp and remained still at all times. The other person created occlusions using objects such as hands and pens. Each time, the moving magnitude of occlusion was as small as possible. The clear images were captured first and then the images with shadows were captured. To align shadow images and clear images, the iPhone XR was not touched, and images were captured and controlled using an earphone wire. The documents, desk, and the smartphone were not touched and their positions were not changed throughout the process. These measures can guarantee the ground truth captured under uniform white illumination.

The size of the controlled group is 960 × 544 (96 dpi), some examples are shown in Figure 5. We also built up the ground truth for shadow regions manually using photoshop, which can be employed for visual comparison and quantitative analysis. The images in the natural group are of different sizes and they are captured with various illuminations and shadow strengths. The OSR dataset is available to the public: "https://github.com/BingshuCV/DocumentShadowRemoval".

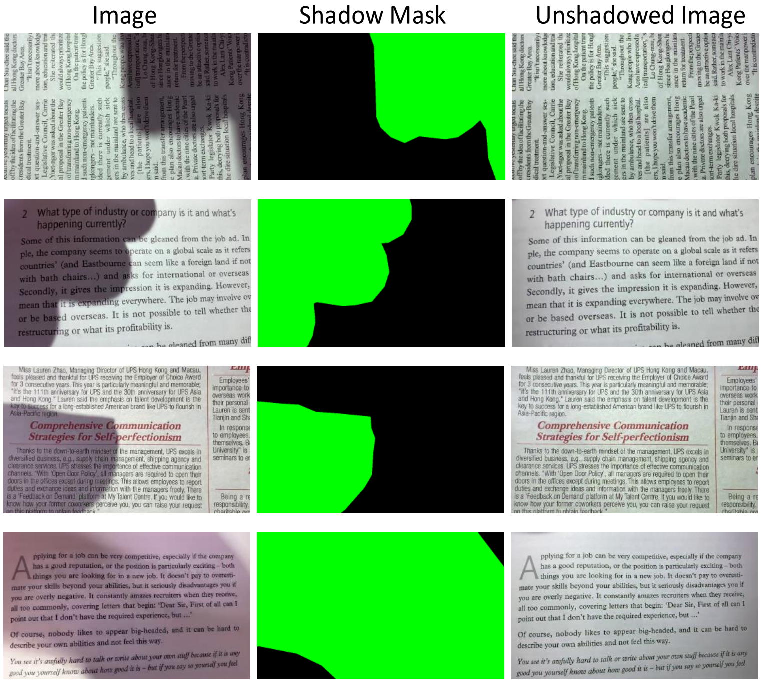

Figure 5. Some examples of the proposed dataset. Specifically, the middle column gives the shadow masks and the green areas indicate the shadow regions. The right column represents the ground truth.

3.2. Evaluation Metrics

To measure the effect of shadow removal, one of the most commonly used evaluation metrics is Mean Squared Error (MSE). It is defined by

$$MSE(R, GT) = \frac{1}{n} \sum (R(x, y) - GT(x, y))^2 \quad (8)$$

where R, GT, and I represents the result image after shadow removal, ground truth, and input image, respectively. n is denoted as the number of pixels. This metric is widely used to evaluate the quality of algorithms. Further, we also employed an evaluation metric $ErrorRatio$ [44] for the assessment of methods, which is shown as follows:

$$ErrorRatio = \frac{RMSE(R, GT)}{RMSE(I, GT)} \tag{9}$$

where $RMSE$ is the root MSE (i.e., $\sqrt{MSE}$). For an image, the area of shadow regions is usually uncertain. When the ratio of the shadow regions (i.e., the green parts labeled in the ground truth in Figure 5) to the whole image is small, the evaluation result may be influenced by the lighted regions (i.e., the black parts labeled in the ground truth in Figure 5). For fairness, only the shadow regions are considered in the evaluation.

In addition, the Structural SIMilarity (SSIM) index [45] is also considered for evaluating the structural similarity between the prediction and ground truth.

3.3. Comparisons with the State-of-the-Art Methods

In comparison to the state-of-the-art methods, we choose a water-filling method [24] and a 3D point cloud-based method [11]. Both represent state-of-the-art techniques for shadow removal in document images. Specifically, we compared these with a CNN model [38]. Quantitative comparisons are presented in Tables 1–3. Visual comparisons are shown in Figures 6–8.

Table 1. Quantitative comparisons of our method and some state-of-the-art approaches for the Adobe dataset with evaluation metrics MSE, $Error\ Ratio$, $SSIM$.

Evaluation Metric	MSE	Error Ratio	SSIM
Kligler et al. [11]	2062.2	2.9489	0.802
Jung et al. [24]	9167.0	6.2104	0.683
Ours	105.8	0.6385	0.927

Table 2. Quantitative comparisons of our method and some state-of-the-art approaches for the HS dataset with evaluation metrics MSE, $Error\ Ratio$, $SSIM$.

Evaluation Metric	MSE	Error Ratio	SSIM
Kligler et al. [11]	517.6	0.5641	0.878
Jung et al. [24]	1287.3	0.8980	0.861
Ours	158.2	0.3059	0.885

Table 3. Quantitative comparisons of our method and some state-of-the-art approaches for the OSR_CG dataset with three metrics: MSE, $Error\ Ratio$, and the average running time on an image with a size of 960 × 544 pixels.

Evaluation Metric	MSE	Error Ratio	SSIM	Running Time (Seconds/Frame)
Kligler et al. [11]	1555.2	0.7160	0.892	8.84
Jung et al. [24]	2313.8	0.9216	0.885	1.396
Ours	1282.4	0.685	0.875	0.265

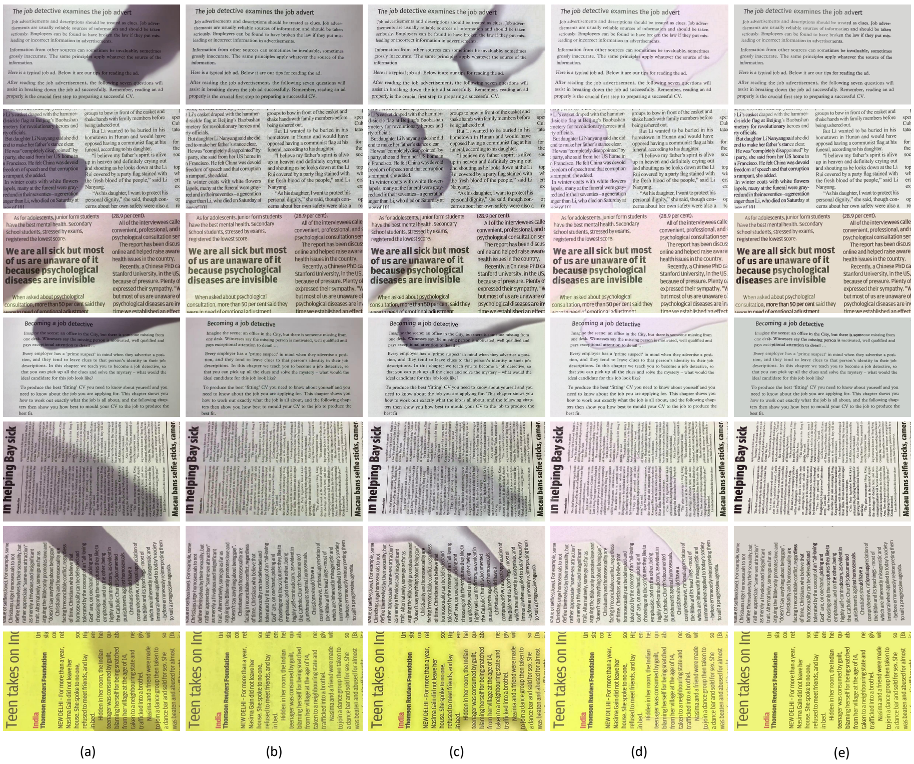

Figure 6. The visual comparisons of some state-of-the-art methods for the proposed OSR_CG dataset. (**a**) the input images, (**b**) the ground truth, (**c**) the results of [11], (**d**) the results of [24], (**e**) our results.

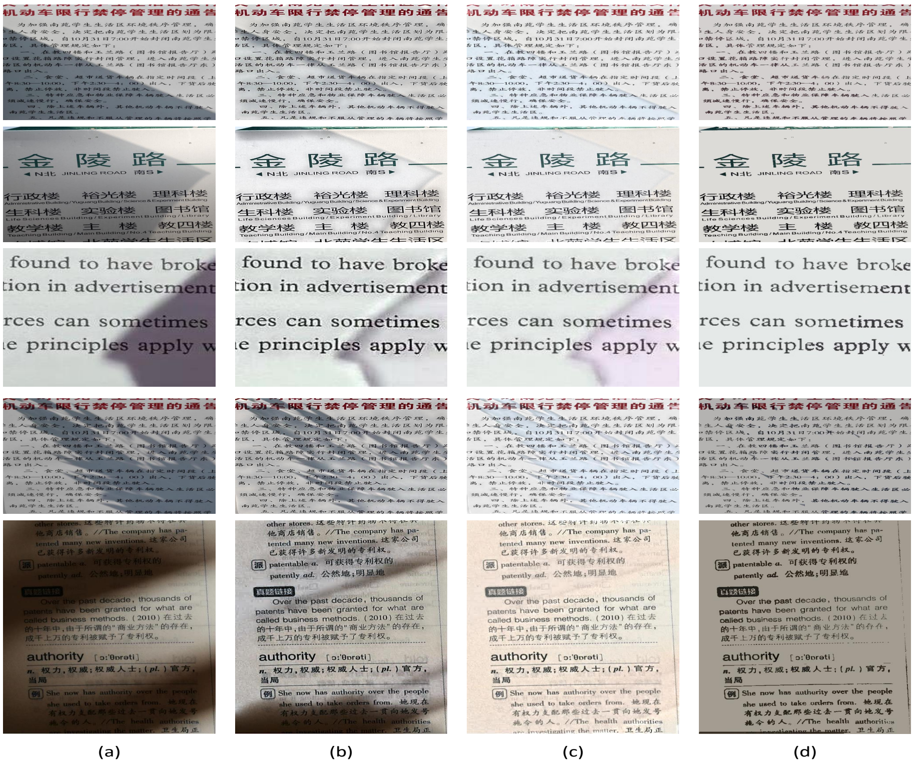

Figure 7. The visual comparisons of some state-of-the-art methods for the proposed OSR_NG dataset. (**a**) the input images, (**b**) the results of [11], (**c**) the results of [24], (**d**) our results.

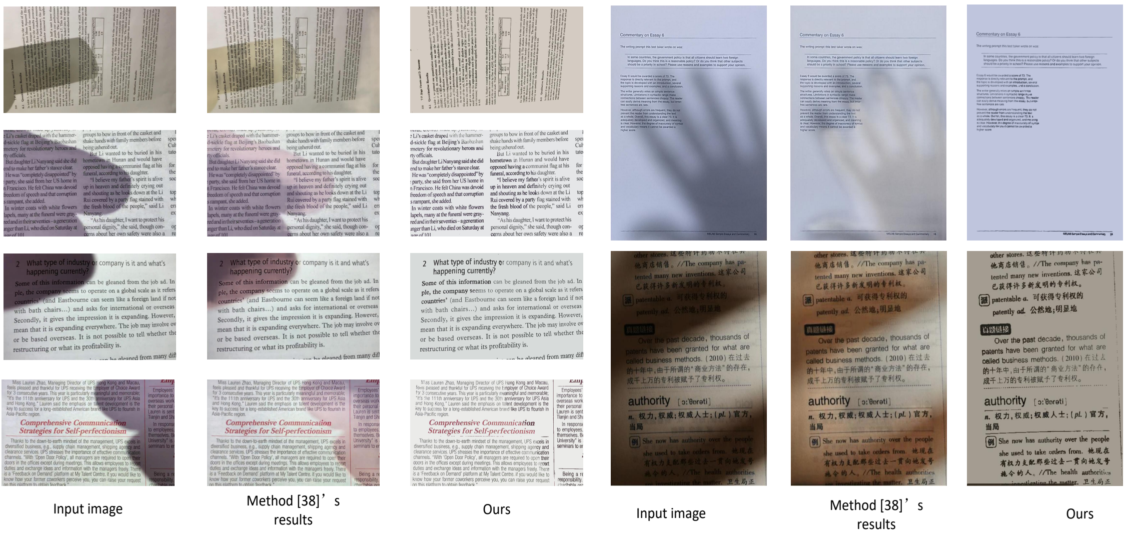

Figure 8. Visual comparison between our method and a CNN model [38].

3.3.1. Quantitative Comparison

In terms of quantitative comparisons, we utilize three evaluation metrics: the MSE, $Error\ Ratio$, $SSIM$. For MSE and $Error\ Ratio$ evaluation metrics, lower values indicate that the method can remove shadows effectively and the produced images are closer to the ground truth. For $SSIM$, the higher the better.

It can be seen from Tables 1–3 that our results are much lower than those the methods in [11,24]. For example, in Table 1, our $Error\ Ratio$ is only 21.65% of method [11], 10.28% of method [24]; our $MSE = 105.8$ is much lower than those of method [11] with $MSE = 2062.2$, method [24] with $MSE = 9167.0$. Meanwhile, Tables 2 and 3 also demonstrate that our method is superior to the methods in [11,24].

The metric $SSIM$ values of the methods are relatively close to each other, but there are differences. Table 1 shows that our method (0.927) is higher than the approach in [11] (0.802) and the approach in [24] (0.683). In Table 2, our method achieves 0.885, better than 0.878 of [11] and 0.861 of [24]. Although our method is inferior to the compared methods in Table 3, the differences are relatively small.

Therefore, our method performs better than the state-of-the-art methods [11,24] in the evaluation metrics. The performance differences are statistically significant. The advantages of the proposed method are demonstrated.

Moreover, we also provide the running time comparison by conducting methods on an image size of 960 × 544 pixels. Our method takes 0.265 s to process one frame, only accounting for one-sixth of the computational cost of the method [24]. A large number of water-filling processes designed in [24] can lead to an large increase of computational cost. As can be seen from the Table 3, the approach proposed by [11] requires 8.84 s to remove shadows for one frame, which is almost over 6 times the computational cost of [24] and 34 times the computational cost of our method. The reason why the method [11] runs slowly is because it spends a long time on visibility detection at the 3D point cloud transformation stage.

The computational complexity of [11] is $O(nlogn)$ for n points. For [24], its computational complexity is $O(pn)$ for n points, p is the number of iterations. Our method's computational complexity is $O(mn)$ for n points, m is the running time of local water-filling. m is less than p. Meanwhile, the number of iterations p or runs m is always set as a constant value and is far lower than the number of points n. Thus, the computational complexities of the method in [24] and ours are of a similar level and far fewer than that of the method in [11].

3.3.2. Visual Results

It can be seen from Figure 6 that visual comparisons of seven images with shadows are presented. Our method achieves better visual results than the compared approaches.

The approach in [24] employed a global incremental filling of catchment basins and corrected illumination distortions on the luminance channel of the $YCbCr$ color space. It is based on the assumption that the color information of shadow regions remains unchanged while the intensity decreases. Figure 6 shows that the method in [24] produces unnatural colors, for example, the shadow regions become pink. This is because for strong shadows, the assumption in [24] is hard to meet. The approach [11] produces many artifacts on shadow boundaries, making the image difficult to perceive visually. The reason may derive from the fact that the 3D point cloud transformation is not able to distinguish shadow points from texts due to the high similarity between some shadow points and text.

The proposed method is inspired by the techniques in [24,41,42] and implemented based on RGB color space, which is defined by the three chromaticities of the red, green, and blue. The method presents a new way to process umbra and penumbra, respectively. By integration with the LBWF-based module, shadow boundaries can be addressed appropriately. The color information belonging to shadow regions appears more natural.

To further demonstrate the effectiveness of our method, we conducted experiments on natural images shown in Figure 7. It can be seen from the figure that the approach in [11] has issues when dealing with nonuniform, strong shadows and the approach in [24] tends to change the color of output images. The proposed method may generate clean unshadowed images.

Figures 6 and 7 show that the methods in [11,24] produce more artifacts than ours, which is in accordance with the quantitative comparison in Table 3. Visual comparisons and quantitative results demonstrate the effectiveness and efficiency of our proposed method.

3.3.3. In Comparison with a Deep Learning Method

Convolutional Neural Network (CNN) models, as a representative of deep learning techniques, have achieved impressive results in various fields. Recently, some CNN models of shadow removal have been proposed to process natural images and these have performed well. To compare with other existing deep learning methods, herein, we compare with a CNN model proposed by [38]. The CNN model can only process an image size of 640 × 480. Therefore, the test images need to be adjusted to this size and then processed. The comparison results are presented in Figure 8. It is pretty clear that many artifacts are left using the approach in [38], resulting in an image that is difficult to percieve visually. The possible reason for this is that the approach in [38] was originally designed to remove shadows from natural images. Thus, it is not suitable for use on images of documents. One potential solution to this problem is to fine-tune a model on a document shadow dataset and redesign the CNN structure. In this regard, the training data should be prepared appropriately in future. In contrast, our method can remove shadows effectively.

It should be noted that the results of some scenarios need to be improved, which is shown in Figure 9. When the colored text is covered with strong shadows, e.g., the red text in the first row and the blue text in the second row, the output text of our method tends to be black. The color degradation might lead to visual inconsistencies. Color constancy methods [26,27] could be considered to address this issue. In this regard, more research needs to be invested in the future.

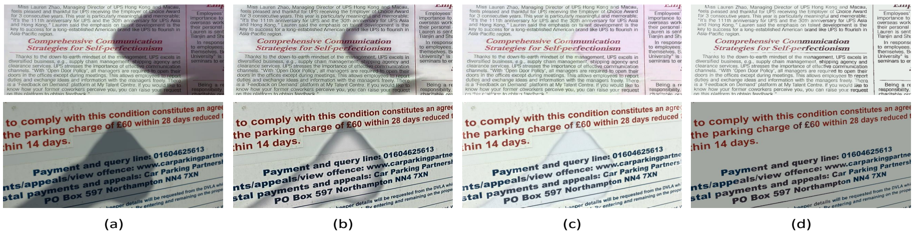

Figure 9. The visual comparisons of some state-of-the-art methods for text shadow removal. (**a**) the input images, (**b**) the results of [11], (**c**) the results of [24], (**d**) our results.

4. Conclusions

In this paper, we proposed a local water-filling-based method for shadow removal. The main objective was to build up a topographic structure using pixels of a document image. An LWF algorithm was developed to estimate the shading map, which was used to divide shadows into umbra and penumbra. We adopted a divide-and-conquer strategy to process umbra and penumbra. Umbra was enhanced by Retinex theory, and penumbra was handled by the proposed LBWF-based algorithm. The strategy offers a powerful way to eliminate shadows, particularly strong shadow boundaries, and produce a clear and easy-to-read document. Moreover, a dataset was created that includes images with strong shadows and is available to the public. Experimental results performed on three datasets indicate that the proposed method outperforms some state-of-the-art methods in terms of effectiveness and efficiency.

Although our method is expected to be a promising technique for document binarization and recognition, we must to point out that the proposed method might produce unsatisfactory results when the shadow regions contain colored text. The output text tends to be dark and lack color information. It may bring discordant visual perception and this limitation will be addressed in the future work.

Author Contributions: Conceptualization, methodology, validation, investigation, data curation, writing—original draft preparation, B.W.; writing—review and editing, visualization, supervision, funding acquisition, C.L.P.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded in part by the National Key Research and Development Program of China under number 2019YFA0706200 and 2019YFB1703600, in part by the National Natural Science Foundation of China grant under number 61702195, 61751202, U1813203, U1801262, 61751205, in part by the Science and Technology Major Project of Guangzhou under number 202007030006, in part by The Science and Technology Development Fund, Macau SAR (File no. 079/2017/A2, and 0119/2018/A3), in part by the Multiyear Research Grants of University of Macau.

Acknowledgments: The authors thank Seungjun Jung (KAIST) and Netanel Kligler (Technion) for their code sharing. We thank Yong Zhao (Peking University) for the interesting discussion. Thanks for Xiaodong Cun's (University of Macau) help in conducting some experiments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lynch, D.K. Shadows. *Appl. Opt.* **2015**, *54*, B154–B164. [CrossRef] [PubMed]
2. Ibarra-Arenado, M.J.; Tjahjadi, T.; Pérez-Oria, J. Shadow detection in still road images using chrominance properties of shadows and spectral power distribution of the illumination. *Sensors* **2020**, *20*, 1012. [CrossRef] [PubMed]
3. Dong, K.; Muhammad, A.; Kang, P. Convolutional Neural Network-Based Shadow Detection in Images Using Visible Light Camera Sensor. *Sensors* **2018**, *18*, 960.
4. Amin, B.; Riaz, M.M.; Ghafoor, A. Automatic shadow detection and removal using image matting. *Signal Process.* **2020**, *170*, 107415. [CrossRef]
5. Hu, X.; Zhu, L.; Fu, C.; Qin, J.; Heng, P. Direction-aware spatial context fea-tures for shadow detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7454–7462.

6. Wang, B.; Chen, C.L. Optical reflection invariant-based method for moving shadows removal. *Opt. Eng.* **2018**, *57*, 093102. [CrossRef]
7. Lee, G.B.; Lee, M.J.; Lee, W.K.; Park, J.H.; Kim, T.H. Shadow detection based on regions of light sources for object extraction in nighttime video. *Sensors* **2017**, *17*, 659. [CrossRef]
8. Manuel, I.A.; Tardi, T.; Juan, P.-O.; Sandra, R.-G.; Agustín, J.-A. Shadow-based vehicle detection in urban traffic. *Sensors* **2017**, *17*, 975.
9. Michael, S.B.; Yau-Chat, T. Geometric and shading correction for images of printed materials using boundary. *IEEE Trans. Image Process.* **2006**, *15*, 1544–1554.
10. Zhang, L.; Yip, A.; Brown, M.; Tan, C.L. A unified framework for document restoration using inpainting and shape-from-shading. *Pattern Recognit.* **2009**, *42*, 2961–2978. [CrossRef]
11. Kligler, N.; Katz, S.; Tal, A. Document enhancement using visibility detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2374–2382.
12. Feng, S.; Chen, C.P. Fuzzy broad learning system: A novel neuro-fuzzy model for regression and classification. *IEEE Trans. Cybern.* **2020**, *50*, 414–424. [CrossRef]
13. Alotaibi, F.; Abdullah, M.T.; Abdullah, R.B.H.; Rahmat, R.W.B.O.K.; Hashem, I.A.T.; Sangaiah, A.K. Optical character recognition for quranic image similarity matching. *IEEE Access* **2017**, *6*, 554–562. [CrossRef]
14. Bako, S.; Darabi, S.; Shechtman, E.; Wang, J.; Sen, P. Removing shadows from images of documents. In Proceedings of the IEEE Conference on Asian Conference on Computer Vision, Taipei, Taiwan, 20–24 November 2016; pp. 173–183.
15. Chen, X.; Lin, L.; Gao, Y. Parallel nonparametric binarization for degraded document images. *Neurocomputing* **2016**, *189*, 43–52. [CrossRef]
16. Feng, S. A novel variational model for noise robust document image binarization. *Neurocomputing* **2019**, *325*, 288–302. [CrossRef]
17. Michalak, H.; Okarma, K. Fast Binarization of Unevenly Illuminated Document Images Based on Background Estimation for Optical Character Recognition Purposes. *J. Univers. Comput. Sci.* **2019**, *25*, 627–646.
18. Xu, L.; Wang, Y.; Li, X.; Pan, M. Recognition of Handwritten Chinese Characters Based on Concept Learning. *IEEE Access* **2019**, *7*, 102039–102053. [CrossRef]
19. Ma, L.; Long, C.; Duan, L.; Zhang, X.; Zhao, Q. Segmentation and Recognition for Historical Tibetan Document Images. *IEEE Access* **2020**, *8*, 52641–52651. [CrossRef]
20. Bradley, D.; Roth, G. Adaptive thresholding using the integral image. *J. Graph. Tools* **2007**, *12*, 13–21. [CrossRef]
21. Shah, V.; Gandhi, V. An iterative approach for shadow removal in document images. In Proceedings of the IEEE Conference on International Acoustics, Speech and Signal Processing, Calgary, AB, Canada, 15–20 April 2018; pp. 1892–1896.
22. Howe, N. Document binarization with automatic parameter tuning. *Int. J. Doc. Anal. Recognit.* **2013**, *16*, 247–258. [CrossRef]
23. Su, B.; Lu, S.; Tan, C. Robust document image binarization technique for degraded document images. *IEEE Trans. Image Process.* **2012**, *22*, 1408–1417.
24. Jung, S.; Hasan, M.A.; Kim, C. Water-Filling: An Efficient Algorithm for Digitized Document Shadow Removal. In Proceedings of the IEEE Conference on Asian Conference on Computer Vision, Perth, Australia, 2–6 December 2018; pp. 398–414.
25. Zhang, J.; Cao, Y.; Fang, S.; Kang, Y.; Wen, C. Fast haze removal for nighttime image using maximum reflectance prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 22–25 July 2017; pp. 7418–7426.
26. Zhang, J.; Cao, Y.; Wang, Y.; Wen, C.; Chen, C.W. Fully point-wise convolutional neural network for modeling statistical regularities in natural images. In Proceedings of the ACM International Conference on Multimedia, Seoul, Korea, 22–26 October 2018; pp. 984–992.
27. Barron, J.; Tsai, T.Y. Fast fourier color constancy. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 22–25 July 2017; pp. 886–894.
28. Barron, J. Convolutional color constancy. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 13–16 December 2015; pp. 379–387.

29. Nguyen, V.; Yago Vicente, T.F.; Zhao, M.; Hoai, M.; Samaras, D. Shadow detection with conditional generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4510–4518.
30. Wang, J.; Li, X.; Yang, J. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1788–1797.
31. Maltezos, E.; Doulamis, A.; Ioannidis, C. Improving the visualisation of 3D textured models via shadow detection and removal. In Proceedings of the IEEE 9th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games), Athens, Greece, 6–8 September 2017; pp. 161–164.
32. Chai, D.; Newsam, S.; Zhang, H.K.; Qiu, Y.; Huang, J. Cloud and cloud shadow detection in Landsat imagery based on deep convolutional neural networks. *Remote Sens. Environ.* **2019**, *225*, 307–316. [CrossRef]
33. Oliveira, D.M.; Lins, R.D.; Silva, G.F. Shading removal of illustrated documents. In Proceedings of the International Conference on Image Analysis and Recognition, Aveiro, Portugal, 26–28 June 2013; pp. 308–317.
34. Zhang, L.; Yip, A.; Tan, C.L. Removing shading distortions in camera-based document images using inpainting and surface fitting with radial basis functions. In Proceedings of the International Conference on Document Analysis and Recognition, Curitiba, Brazil, 23–26 September 2007; pp. 984–988.
35. Xu, Y.; Wen, J.; Fei, L.; Zhang, Z. Review of video and image defogging algorithms and related studies on image restoration and enhancement. *IEEE Access* **2015**, *4*, 165–188. [CrossRef]
36. Xie, Z.; Huang, Y.; Jin, L.; Liu, Y.; Zhu, Y.; Gao, L. Weakly supervised precise segmentation for historical document images. *Neurocomputing* **2019**, *350*, 271–281. [CrossRef]
37. Wang, J.; Chuang, Y. Shadow Removal of Text Document Images By Estimating Local and global Background Colors. In Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Barcelona, Spain, 4–8 May 2020; pp. 1534–1538.
38. Cun, X.; Pun, C.M.; Shi, C. Towards Ghost-free Shadow Removal via Dual Hierarchical Aggregation Network and Shadow Matting GAN. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020.
39. Alnefaie, A.; Gupta, D.; Bhuyan, M.; Razzak, I.; Gupta, P.; Prasad, M. End-to-End Analysis for Text Detection and Recognition in Natural Scene Images. In Proceedings of the International Joint Conference on Neural Networks, Glasgow, UK, 19–24 July 2020.
40. Wang, B.; Chen, C.L. Moving cast shadows segmentation using illumination invariant feature. *IEEE Trans. Multimed.* **2019**, *22*, 2221–2233. [CrossRef]
41. Vincent, L.; Soille, P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *6*, 583–589. [CrossRef]
42. Roerdink, J.B.; Meijster, A. Watershed transform: Definitions, algorithms and parallelization strategies. *Fundam. Inform.* **2000**, *41*, 187–228. [CrossRef]
43. Land, E.; McCannJ, J. Lightness and retinex theory. *Josa* **1971**, *61*, 1–11. [CrossRef]
44. Gong, H.; Cosker, D. Interactive Shadow Removal and Ground Truth for Variable Scene Categories. In Proceedings of the British Machine Vision Conference, Nottingham, UK, 1–5 September 2014; pp. 1–11.
45. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.