A Unified Framework for Document Restoration using Inpainting and Shape-from-Shading

Li Zhang[a,1], Andy M. Yip[b,1], Michael S. Brown[c,1], Chew Lim Tan[d,1]

[a] *School of Computing, National University of Singapore*
[b] *Department of Mathematics, National University of Singapore*
[c] *School of Computing, National University of Singapore*
[d] *School of Computing, National University of Singapore*

Abstract

We present a restoration framework to reduce undesirable distortions in imaged documents. Our framework is based on two components: 1) an image inpainting procedure that can separate non-uniform illumination (and other) artifacts from the printed content; and 2) a Shape-from-Shading (SfS) formulation that can reconstruct the 3D shape of the document's surface. Used either piecewise or in its entirety, this framework can correct a variety of distortions including shading, shadow, ink-bleed, show-through, perspective and geometric distortions, for both camera-imaged and flatbed-imaged documents. Our overall framework is described in detail. In addition, our SfS formulation can be easily modified to target various illumination conditions to suit different real-world applications. Results on images of synthetic and real documents demonstrate the effectiveness of our approach. OCR results are also used to gauge the performance of our approach.

Key words: document image restoration, shading distortion, perspective distortion, geometric warping, digital inpainting, RBF-based smoothing, shape-from-shading, physically-based flattening.

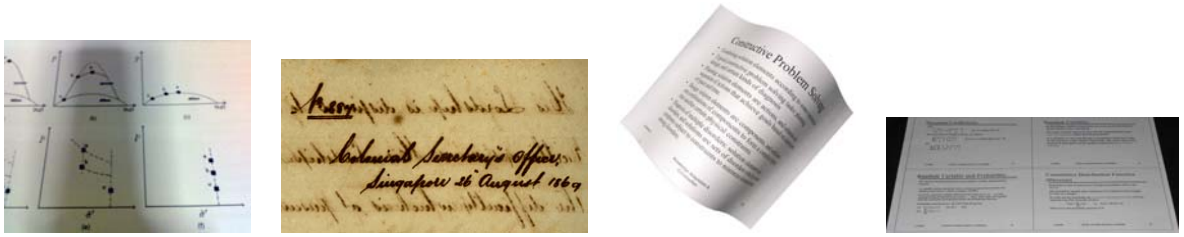
1. Introduction

Document imaging is a fundamental application of computer vision and image processing. The ability to image printed documents has contributed greatly to the creation of vast digital

Email addresses: zhangli@comp.nus.edu.sg (Li Zhang), andyyip@nus.edu.sg (Andy M. Yip), brown@comp.nus.edu.sg (Michael S. Brown), tancl@comp.nus.edu.sg (Chew Lim Tan)

collections now available from libraries and publishers. While traditional document imaging has been performed using flatbed scanning devices, a trend towards more flexible camera-based imaging is also emerging. The goal of document imaging, either via flatbed scanners or digital cameras, is to capture an image that is a reasonable substitute for the original printed content. However, with both approaches unavoidable distortions can be present in the resulting image due to the printed materials' construction (e.g. non-planar), the imaging setup, or environmental conditions, such as non-uniform illumination. Such distortions can make the document difficult to read as well as adversely affect the performance of subsequent processing, namely Optical Character Recognition (OCR) and Document Layout Analysis (DLA).

In this paper, we show how an image inpainting procedure can be combined together with Shape-from-Shading (SfS) to solve a variety of common distortions found in imaged documents. Our framework is reasonably generic and suitable for use with traditional flatbed scanner imaging as well as less-restrictive camera-based imaging. We demonstrate the effectiveness of our framework on four types of distortions: 1) shading/shadow, 2) ink-bleed and show-through, 3) perspective distortion (from camera-imaged materials), and 4) geometric distortion arising in non-planar documents. Examples of these distortions are shown in Figure 1. The inpainting routine can be used by itself to address shading and ink-bleed. However, combined with the SfS approach, we can also correct perspective and geometric distortions.



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

Figure 1: Document images with (a) shadows; (b) background noise; (c) geometric and shading distortions; (d) perspective distortions.

Conference versions of portions of this work have appeared in [1, 2, 3]. In this paper,

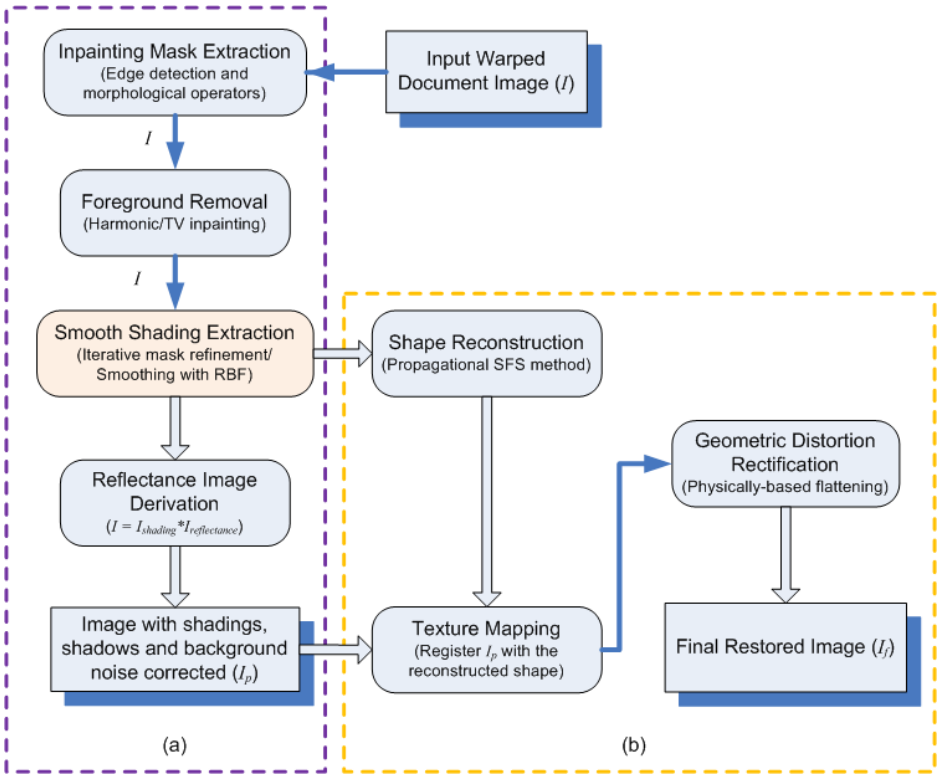


Figure 2: An overview of the restoration framework for (a) shadings, shadows and background noise; (b) perspective and geometric distortions.

we collate these shorter versions and provide more details to the individual components as well as additional experimental results to demonstrate the effectiveness of our framework. In addition, we have extended the application domain to address various background noise as well as perspective distortions.

Figure 2 shows an overview of the various components of our system. To correct shading distortions and background noise from ink-bleed and show-through, we first extract a background layer image based on digital inpainting, which is then used to derive the foreground reflectance image based on the notion of intrinsic images. When dealing with documents with ink-bleed, we show how we can tune our edge detector so that the bleed-through pixels are treated as background and thus separated from the foreground stroke pixels.

If the image contains only smooth shading without other background noise or shadows, we can use a RBF-based smoothing technique to extract a smooth shading image. This

shading image can then be used to reconstruct the surface shape of the document based on the SfS methodology. Here we propose a generic SfS method considering the perspective projection model and various lighting conditions. Our SfS is based on a viscosity framework by solving the image irradiance PDE using Lax-Friedrich Hamiltonian and fast sweeping strategy. The geometric distortions are then removed by mapping the 3D surface back to a plane. Since geometric and photometric distortion are often found together, we can use the photometrically corrected image in the geometric correction step and obtain the final corrected image. Compared to our previous framework proposed in [2, 3], the current framework extends the application domain to a wider range of distortions including background noise and perspective distortions. Compared to the SfS method proposed in [1], the current method is more generic and more accurate attributing to the additional incorporation of the illumination direction. In addition, more comprehensive experiments and comparisons have been conducted to evaluate the proposed methods with a large set of synthetic and real document images displaying various types of distortions.

1.1. Previous Work

Shading distortions in scanned document images can be corrected using binarization techniques such as local thresholding [4], global thresholding [5] or based on the surface shape of the document, typically cylindrical book spines, with the knowledge of the scanner's structure [6]. Shading artifacts in camera-based imaged document are generally more complex especially when they are combined with geometric distortions in arbitrarily warped document images [7]. Boundary interpolation has been used to correct both shadings and geometric distortions on images of warped art materials with iso-parametric folding lines [8]. Sun et al [9] present a system to restore both geometric and photometric artifacts of arbitrarily folded documents by classifying the intensity changes to either illumination changes or reflectance changes based on the notion of intrinsic images [10]. More studies about intrinsic image extraction are done on real-scene images, which treat both shadings and shadows as the illumination intrinsic image and try to separate it from the reflectance image based on color and gradient information [11, 12, 13, 14]. Despite all these efforts on deriving the

intrinsic images, there is no single exact solution because the decomposition of the intensity image into its two intrinsic components is theoretically not unique.

Background noise as part of the image degradations in historical documents have been tackled in various directions among the document image analysis community. One direction is based on binarization techniques such as the multi-stage binarization method proposed by Bar-Yosef et al to restore and recognize ancient Hebrew calligraphy documents [15]. Leedham [16] et al have investigated three global thresholding algorithms and a multi-stage thresholding algorithm to separate text from background in degraded document images and concluded that the given global algorithms do not work well with difficult documents due to over-thresholding while the multi-stage algorithm can do a better job by incrementally remove the noise. Another direction is to separate different layers using classification techniques especially in addressing the bleed-through problem. For example, Drira et al [17] proposed to solve the bleed-through problem by classifying the pixels of an image to background, original text and interfering text, and then replacing the last class with an average background value. Boussellaa [18] et al use a hybrid method to separate foreground from background in Arabic historical documents . There are also other techniques based on wavelet reconstruction [19], color decorrelation [20], to address the bleed-through problem. These latter methods required two registered images from front and back in order to achieve good results.

Perspective distortions are more pervasive in daily imaging applications. Various methods have been proposed in the document image domain based on the presence of orthogonal lines such as parallel text lines, page or paragraph boundary lines, etc. For example, Clark and Mirmehdi [21] used an extension of the 2D projection profile to locate horizontal vanishing point followed by the vertical vanishing point based on the change of line spacings. The vanishing points are then used to recover a frontal-parallel view of the document suitable for OCR. Pilu also proposes a method to extract horizontal and vertical features from the image based on text blocks [22]. Myers et al [23] presented a novel system that extracts text from real-world scenery images and performs perspective rectification. This approach assumes a weak perspective projection in the vertical direction and rectified foreshortening

and shearing using the top line, base line and vertical edge directions.

Often perspective and shading distortions are accompanied with geometric distortions due to a non-planar surface shape of the document. The geometric correction problem has been studied extensively for both scanned and camera-based document images. The proposed methods range from purely 2D-based image processing techniques to 3D-based shape manipulation techniques. Most 2D-based methods try to find spatial transformations between the input and the output images by analyzing the 2D textual content such as known reference points [24], text lines [4], document boundaries [8], and so on. 3D-based methods utilize the shape information to get a more accurate representation of the physical warping. To obtain the surface shape, methods have been proposed to capture it using special setups such as structured lighting [25, 26, 27], laser scans [28] or even CT scans for opaque objects [29]. The captured 3D model is then mapped to a plane through various numerical methods to obtain the restored image. Moreover, approaches are also proposed to reconstruct the shape based on special surface properties [30] or features of the imaging devices such as flatbed scanners [6, 31], stereo vision systems [32] and video cameras [33].

In this paper, we present a generic restoration framework that corrects a wide range of distortions in different types of document images. The photometric distortions such as shading and background noise are addressed using an inpainting based method. Inpainting algorithms have long been used for image restoration applications. Here, we apply them to document images for background noise removal and furthermore combine them with an RBF-based smoothing technique to extract qualitative shading images for both photometric correction and geometric correction. We also propose a SfS model specially formulated for the light-controlled document imaging environment and solve it using a propagational approach followed by minimization-based enhancement. The reconstructed shape is then flattened to a plane using a physically-based flattening method that we proposed earlier. This restoration framework has applied several conventional methodologies such as inpainting and RBF-based surface fitting for photometric correction, and also proposed a hybrid SfS method to reconstruct the surface shape for geometric correction. Encouraging results have been obtained as shown in the subsequent sections.

2. Shading and Background Noise Correction

Document shading is caused by changes of the surface normal with respect to the illumination direction. However, due to the intricate light sources in the normal imaging environment and the inconsistent properties of the surface materials, shadings in real-world images, especially camera acquired images, are often too complicated to be defined by a uniform model. Shadows and background noise are even more irregular, which could be in any shape, color and intensity. This makes it difficult to build a universal model that fits in all sorts of situations. Nevertheless, common properties still exist in a sense that all these distortions can be treated as a layer different from the foreground strokes. The task is therefore to separate this layer from the foreground layer effectively. To do this, we have carefully designed the following procedures which first reconstruct a background layer using edge detection and inpainting techniques and then separate it from the foreground layer based on the notion of intrinsic images. The reconstructed background layer can be further refined to produce a smooth shading image that resembles the actual shading as closely as possible when shadows and background noise are non-existent. This shading image can then be used to reconstruct the document's surface shape for geometric correction purposes if needed.

2.1. Inpainting Mask Generation

To reconstruct a background layer, we first identify pixels that do not belong to this layer. Here we assume that in most images foreground text (or handwritten stroke) pixels are relatively darker than shadings, dim shadows and other noise such as bleed-through and water stains. The task is similar to text localization except that we are interested in all the pixels that may indicate a reflectance change including text and graphics. Text localization has been a widely researched area either on document images or digital videos. The techniques can be broadly classified as component-based [34, 35] or texture-based [36, 37]. The component-based methods usually try to analyze the geometrical arrangements of edges or uniform colored components of the characters. The texture-based methods utilize the texture characteristics of text lines to extract the text. Here we use a component-based

method as the first step to identify pixels that are of high contrast to the background. Next, morphological operations are applied to the edge-detected image to generate a mask of the foreground pixels. The detailed procedure is as follows:

1) Convert color images into gray-scale using the luminance channel from the YUV color model.

2) Detect edges using canny edge detector. Different thresholds can be chosen to better distinguish foreground strokes and unwanted noise such as bleed-through;

3) Perform morphological dilation followed by closing. The purpose of this step is to mask all the foreground pixels as completely as possible to avoid noise after the inpainting process. The size of the structuring element can be tuned manually or adjusted automatically based on an estimated average character height when applicable.

2.2. Harmonic/TV Inpainting

Digital inpainting was introduced by Bertalmio et al[38] and has since been applied to a variety of image processing applications. Here we use it to recover the background layer of the input image which can be further refined to produce a smooth shading image in certain situations. The idea is to fill up the masked foreground regions using the neighboring background pixels. In the case of an image with a uniformly colored background, this essentially recovers the shading in the masked regions based on the assumption that the local shading variation is small. On the other hand, if our main focus is on recovering the foreground content, the quality of the background layer is therefore not so crucial. For example, we probably do not need to recover a fine detail of the bleed-through and water stain pixels. To this purpose, we look into the two non-texture inpainting models - harmonic inpainting and Total Variation (TV) inpainting [39].

Mathematically, inpainting can be considered as a local interpolation problem: Given an image I_0 with a hole Ω_H inside, we want to find an image I that matches I_0 outside the hole and has consistent information inside the hole. To do this, we try to find I that minimizes the following energy in a continuous domain Ω:

$$E(I) = \int_{\Omega} \chi \cdot (I - I_0)^2 d\mathbf{x} + \lambda \int_{\Omega} |\nabla I|^2 d\mathbf{x} \tag{1}$$

where $\lambda > 0$ is a smoothness parameter and χ denotes the characteristic function:

$$\chi(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in \Omega \backslash \Omega_H \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

To minimize the energy in Eq. 1, we solve the Euler-Lagrange equation for the functional $F(I, I_u, I_v) = \chi(I - I_0)^2 + \lambda |\nabla I|^2$ corresponding to Eq. 1:

$$\frac{\partial E}{\partial I} = \frac{\partial F}{\partial I} - \frac{\partial}{\partial u}\frac{\partial F}{\partial I_u} - \frac{\partial}{\partial v}\frac{\partial F}{\partial I_v} = 2[\chi(I - I_0) - \lambda(I_{uu} + I_{vv})] = 0 \tag{3}$$

By applying a gradient-descent method and a discretization using finite difference, we obtain the iterative update formula:

$$I_{i,j}^{n+1} = I_{i,j}^n + \Delta t \left(\frac{\lambda}{h^2}(I_{i+1,j}^n + I_{i-1,j}^n + I_{i,j+1}^n + I_{i,j-1}^n - 4I_{i,j}^n) - \chi_{i,j}(I_{i,j}^n - I_{0i,j}) \right) \tag{4}$$

where h is the grid size and λ is the smoothness parameter which is chosen through trial and error. The time step Δt can be any small constant that makes the iteration stable, which is chosen as 0.2 in our experiments.

We notice that the harmonic inpainting constructs a smooth solution which may cause problems when the foreground pixels at the image boundaries are completely masked or when the interior edges such as folds are occluded due to the overlaid stroke pixels. The edges at these places are often missing after being inpainted with a smoothly filled interior. This can be mitigated with the use of TV inpainting. Instead of using a penalty term $\int |\nabla I|^2 d\mathbf{x}$ in Eq. 1, which is infinite for discontinuous functions, we use $\int |\nabla I| d\mathbf{x}$ which allows discontinuous functions as minimizers. The energy function now becomes:

$$E(I) = \int_\Omega \chi \cdot (I - I_0)^2 d\mathbf{x} + \lambda \int_\Omega |\nabla I| d\mathbf{x} \tag{5}$$

where $\lambda = 2\sigma^2/\nu$. A minimizer for this energy function can be computed using a similar scheme as for harmonic inpainting.

Note that both harmonic and TV inpainting are local models, in which the inpainting is mainly determined by the existing information I_0 in the vicinity of the inpainted domain Ω_H. We can also choose to retain the original background pixels and only inpaint the masked

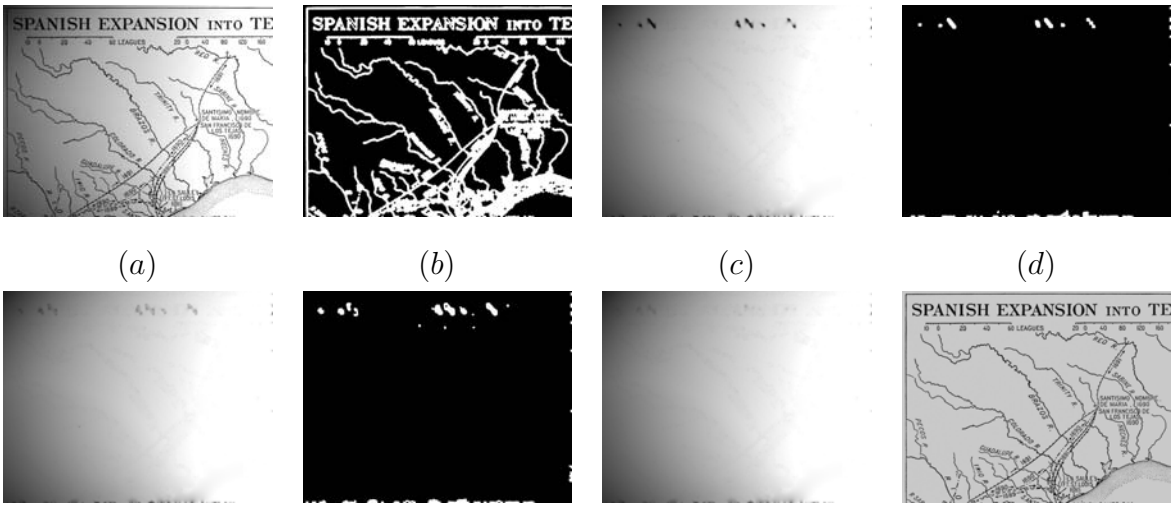
regions to minimize computational cost. Moreover, Eq. 1 has a built-in denoising capacity so that it is robust to noise. The main difference is that harmonic inpainting builds very smooth solutions and thus does not cope well with edges, while TV inpainting is able to restore narrow broken smooth edges which often exist in document images due to overlaid stroke pixels.

2.3. Smoothing with RBF

Although the inpainting process is able to remove all the masked foreground pixels and return an estimated background layer image, the result is not perfect due to the errors in the extracted mask. This is acceptable if our purpose is mainly to remove the background noise and extract the foreground strokes such as bleed-through and water stain removal. However, if we need a smooth shading image for the sake of a subsequent surface reconstruction process, additional refinement step is then necessary.

One way to remove the pepper noise in the inpainted image is to iteratively improve the mask and compute the inpainted image until all the foreground pixels are completely covered. More specifically, the first iteration extracts a mask from the original input image while the subsequent iterations extract masks from the inpainted images constructed in the previous step. Figure 3 shows an example of improving the inpainted image using an iterative approach. In this example, the line strokes are of variable size and thickness. In other scenarios where creases or folds exist, to distinguish reflectance edges from illumination edges caused by creases or folds, the edge detector may need to be tuned specifically to avoid confusions. Nevertheless, even if some creases or folds are mis-identified as reflectance edges, the inpainting algorithm is still able to provide a close estimation of the original edge based on the neighboring pixel values to a certain extent.

Alternatively, we can smooth out the errors by using a smoothing algorithm with radial basis functions (RBF) [40]. This is especially useful when the background layer mainly comprises of shadings caused by smoothly warped surfaces. Consider the pixels in the inpainted image as a set of noisy 3D points $\{(\mathbf{x}_i, I(\mathbf{x}_i)),\ i = 1, 2, \cdots, m\}$ where m is usually equal to width$\times$height, our goal is to find an approximate fitting to these 3D points by using



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

(e) (f) (g) (h)

Figure 3: Background layer improvement using an iterative approach: (a) A map image with synthetic shadings; (b) Inpainting mask (Iteration 1); (c) Inpainted image (Iteration 1); (d) Inpainting mask (Iteration 2); (e) Inpainted image (Iteration 2); (f) Inpainting mask (Iteration 3); (g) Inpainted image (Iteration 3); (h) Final restored image.

a selected subset of points as the collocation points that uniformly span the whole image horizontally and vertically denoted by $\{\mathbf{y}_j,\ j = 1, 2, \cdots, n\}$. The interval Δd between these collocation points can be adjusted to achieve different smoothness. In our experiments, we set it to 40 pixels. The task is therefore to find the coefficients α_j that minimize the least squares error defined as:

$$e = \min_{\alpha_1,\cdots,\alpha_n} \left\{ \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \alpha_j h(\mathbf{x}_i - \mathbf{y}_j) - f(\mathbf{x}_i) \right)^2 \right\} \tag{6}$$

with optional boundary conditions. The kernel function used here is Multiquadrics defined as:

$$h(\mathbf{x}) = \sqrt{||\mathbf{x}||^2 + c^2} \tag{7}$$

where c is a constant with empirical values of 10 for all our experiments. Other kernel functions can also be used such as Gaussian, thin-plate spline, etc.

Figure 4 shows a typical example of how the RBF-based smoothing technique helps improve the shading image using a smoothly warped document page. In particular, the

one-step inpainted image as shown in Figure 4 (c) contains obvious pepper noise due to unmasked ink pixels. To apply the smoothing technique, we first select 12×12 collocation points that uniformly sample the image plane as shown in Figure 4 (d). Next, we compute a smooth fitting to the original noisy data points based on the selected collocation points and the multiquadrics kernel function. A cross section view of the fitted surface at $v = 50$ is shown in Figure 4 (e). Finally, the smoothed shading image is shown in Figure 4 (f).

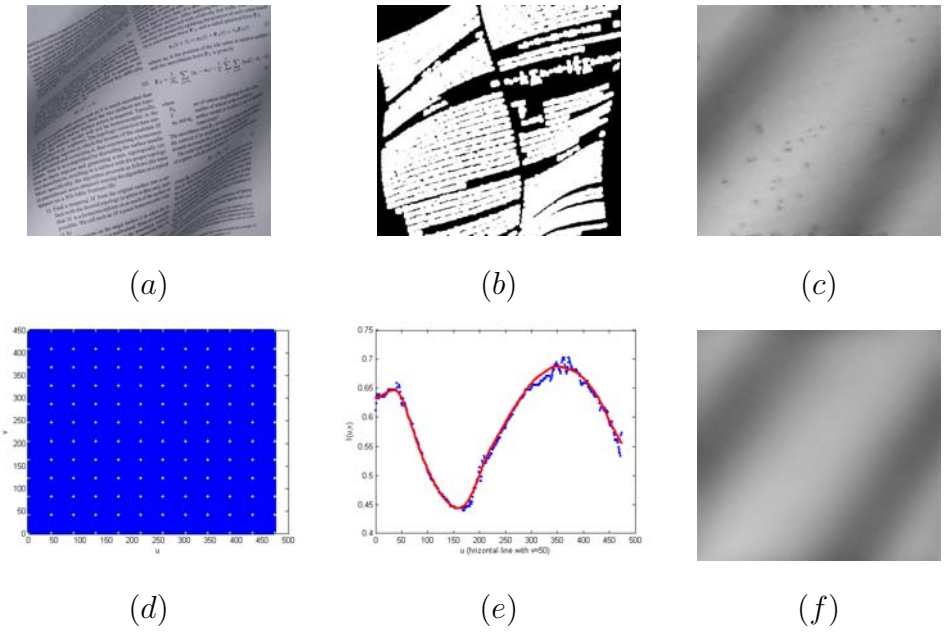


Figure 4: RBF-based smoothing: (a) A warped document page with smooth shadings; (b) Inpainting mask; (c) Inpainted image; (d) Selected collocation points; (e) Cross section view of the fitted surface at $v = 50$; (f) Smoothed shading image.

2.4. Background Layer Removal

Once the background layer is extracted, the foreground layer image can be derived based on the notion of intrinsic images [10]. For Lambertian surfaces, the intensity image is equal to the product of the shading image and the reflectance image. Therefore, the idea is to treat the foreground layer as the reflectance image and the background layer as the illumination image. Consider the luminance component of the HSV model, we have $I = I_s \cdot I_r$. Now given the background layer image I_s, the foreground layer image I_r can be computed as: $I_r = e^{log\, I - log\, I_s}$. The photometrically restored image can be computed as: $I_p = k \cdot I_r$, where $k \in [0, 1]$.

3. Experimental Results for Photometric Correction

To demonstrate the performance of the aforementioned restoration procedures, we have conducted experiments on a data set including 18 synthetic images with smooth shading distortions and 35 real images with different types of distortions ranging from smooth shadings, non-smooth shadings to shadows and background degradations. The images we are dealing with are mainly text documents without large photos or graphical regions. The synthetic images are generated by adding controlled illumination such as an off-centered spot light source over an original clean document image. The real images are captured using either normal digital cameras or mobile phone cameras under unconstrained imaging conditions. In addition, the images include both modern documents and historical manuscripts which are either in color or gray-scale. The inpainted images are produced using the harmonic inpainting algorithm with at most two iterations.

3.1. Results on Synthetic Document Images

Figure 5 demonstrates the performance of the proposed method using a synthetic example picked up from the synthetic image set. The original clean image is a digital image of a textual document containing mainly text information. The distorted image is obtained by simulating bad illumination conditions using an off-centered spot light source with large distance attenuations. The inpainted image shown in Figure 5 (d) is obtained after two iterations of the inpainting process. The final restored reflectance image in Figure 5 (e) is shown to be a close resemblance of the original clean document image.

3.2. Results on Real Document Images

Figure 6 demonstrates the restoration results of two real images with smooth shading distortions due to a flash light source applied on the warped surface shape. Figure 6 $(b_1)(b_2)$ show the inpainting masks and $(c_1)(c_2)$ show the inpainted background layer images produced using the harmonic inpainting algorithm after one iteration. Figure 6 $(d_1)(d_2)$ are the smooth shading images obtained after the refinement step using a RBF-based smoothing technique. Clearly, the refined images contain less noise than the inpainted image after

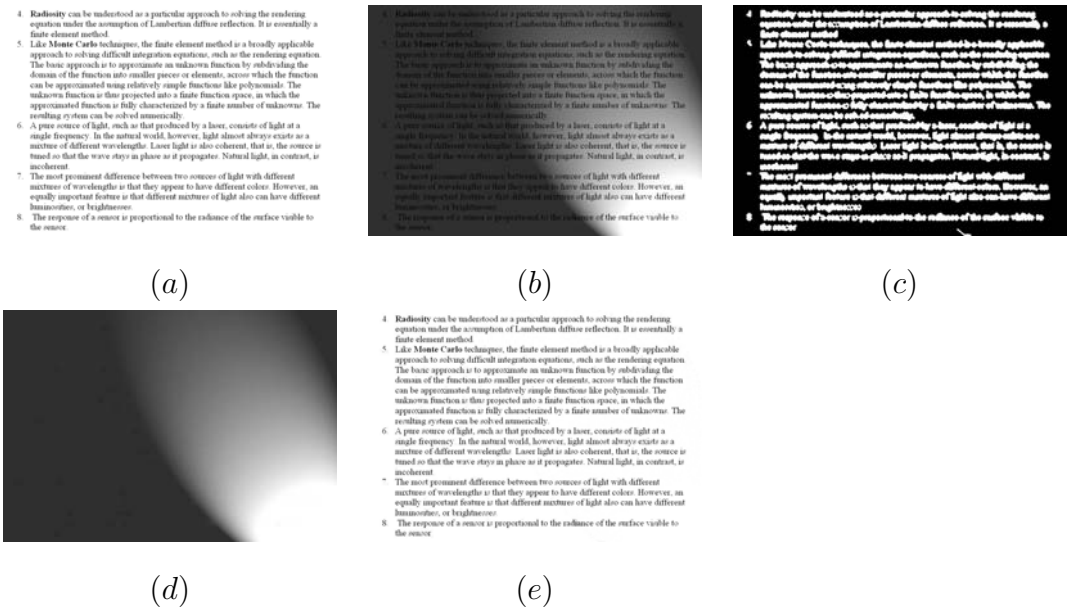
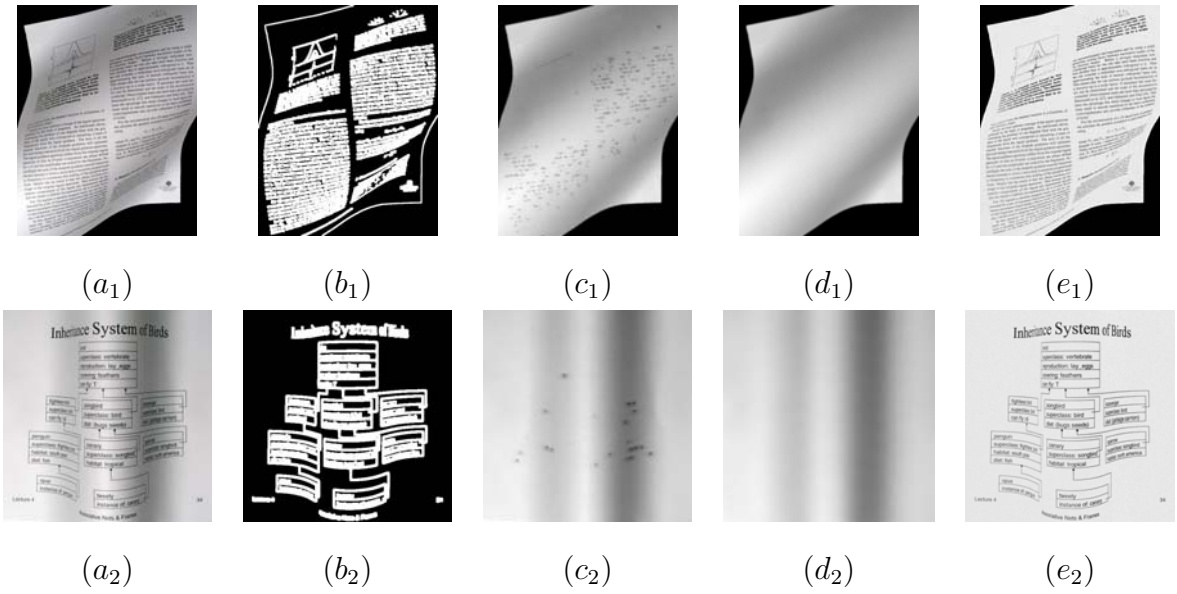


Figure 5: Restoration results of document images with synthetic shadings: (a) Original clean document image; (b) Distorted image with synthetic shadings; (c) Inpainting mask; (d) Inpainted image; (e) Photometrically restored image.

one iteration. Finally, Figure 6 $(e_1)(e_2)$ show the restored images after photometric correction, which are visually more legible notwithstanding the geometric distortions. We will demonstrate how these refined shading images can be used to reconstruct the surface shape effectively for further geometric correction in the next Section.

Figure 7 $(a_1)(a_2)$ show two images taken using mobile phone cameras with normal indoor lightings. Shadows of the mobile phone or neighboring articles are cast on the document and result in large shading variations across the image. Our method is able to separate the shadows effectively from the original image and produce a set of much more machine-processable restored images as shown in Figure 7 $(d_1)(d_2)$.

Figure 8 shows the results of two sample document images provided by the National Archive of Singapore with distortions including ink-bleed, water stains or smudges. The original images are seriously distorted and hard to read due to the ink bleed-through problems and material degradations. Figure 9 shows an enlarged portion of the original image (a_1) in figure 8. It can be seen from the enlarged image in figure 9 that after restoration, the written text is much more readable which facilitates various scholarly studies and inves-



(a_1) (b_1) (c_1) (d_1) (e_1)

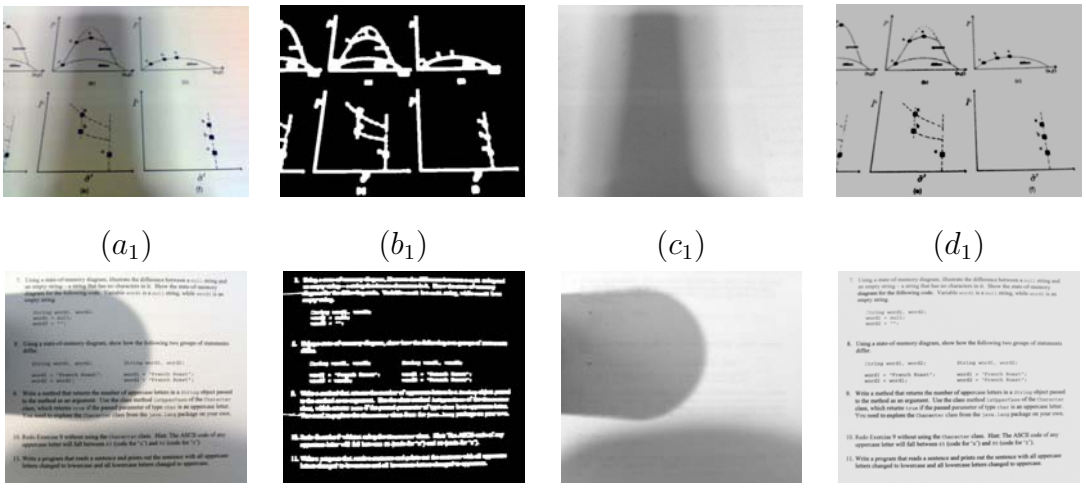
(a_2) (b_2) (c_2) (d_2) (e_2)

Figure 6: Restoration results of real document images with smooth shadings: $(a_1)(a_2)$ A real document image with shading distortions; $(b_1)(b_2)$ Inpainting mask; $(c_1)(c_2)$ One-pass inpainted image; $(d_1)(d_2)$ Smooth shading image extracted after RBF smoothing; $(e_1)(e_2)$ Photometrically restored image with $k = 0.9$.

tigations. In particular, the reconstructed background layer contains all the bleed-through pixels, water stain noise and smudges, which is then extracted from the original image to obtain the foreground reflectance image.

Figure 10 demonstrates the restoration of different degrees of show-through effects in duplex printed document pages using two typical examples. As we can see from the original images shown in Figure 10 (a) that some pixels from the back-side of the page are so strong that they interfere with the recognition of the foreground pixels. On the other hand, they are still relatively of low intensity which makes them distinguishable from the foreground strokes. Figure 10 (b) and (c) show the extracted inpainting mask and the inpainted background layer image, respectively. It is clear that the show-through effects are mostly captured in the background layer image which leaves a clean foreground image as shown in Figure 10 (d).

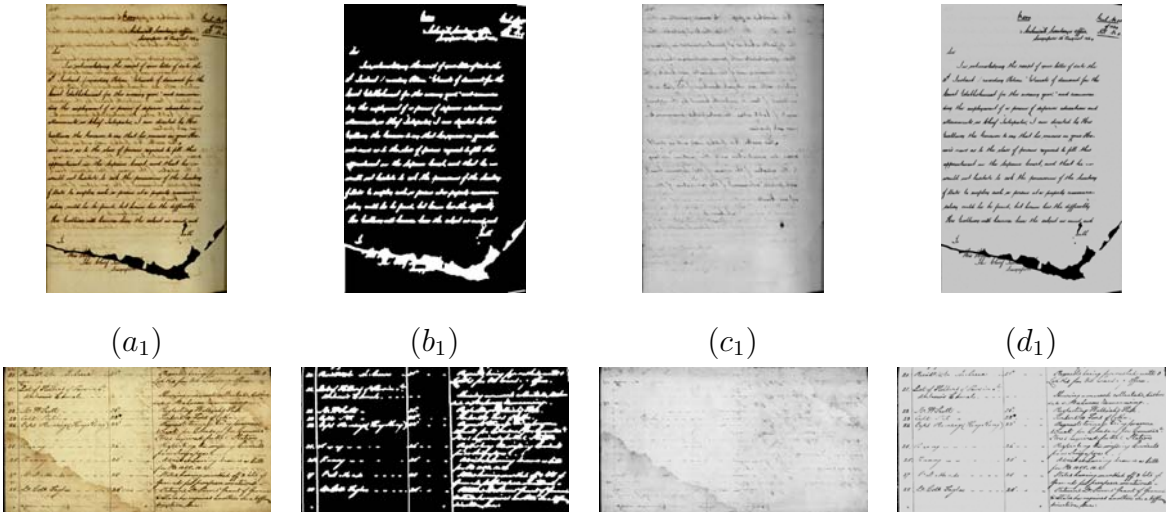
To demonstrate the computational complexity of the proposed restoration procedure, Table 1 tabulates the time taken for each of the restoration steps on the eleven images shown in Figure 6, 7, 8 & 10. All the experiments are run on an Intel Pentium III 996MHz PC (512 MB RAM) with implementations using Matlab. The surface smoothing step is only



(a_1) (b_1) (c_1) (d_1)

(a_2) (b_2) (c_2) (d_2)

Figure 7: Restoration results of real document images with non-smooth shadings or shadows: (a) Original distorted image; (b) Extracted inpainting mask; (c) Reconstructed background layer image; (d) Restored image.



(a_1) (b_1) (c_1) (d_1)

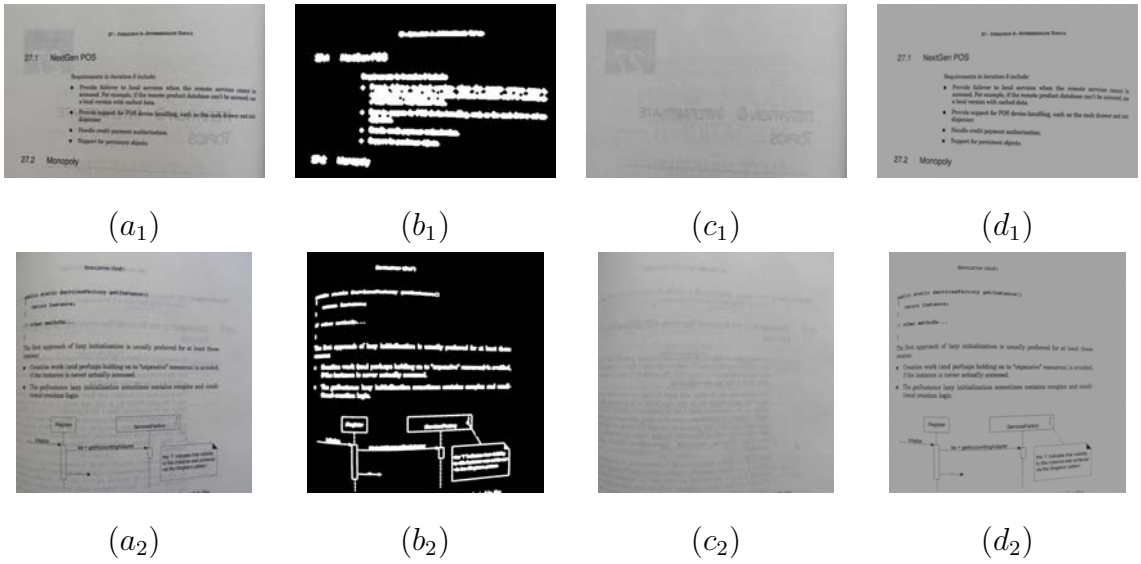
(a_2) (b_2) (c_2) (d_2)

Figure 8: Restoration results of real degraded historical document images with background noise: (a) Original distorted image; (b) Extracted inpainting mask; (c) Reconstructed background layer image; (d) Restored image.

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

Figure 9: Restoration results of an enlarged portion of the image shown in Figure 8 (a_1).

applicable to the two images with smooth shadings shown in Figure 6. Therefore, the time of this step is only measured for the first two images. In particular, the time of the mask extraction step and the RBF smoothing step mainly depends on the size of the image. The time of the inpainting step depends on both the size of the image and the density of the masked regions.



(a_1) (b_1) (c_1) (d_1)

(a_2) (b_2) (c_2) (d_2)

Figure 10: Restoration results of duplex printed document images with show-through effects: (a) Original distorted image; (b) Extracted inpainting mask; (c) Reconstructed background layer image; (d) Restored image.

3.3. Comparisons with Existing Methods

In this experiment, we use a synthetic image to compare our result with those from two existing binarization methods. In particular, Niblack's method [41] is a typical representative of the local adaptive thresholding methods and Lu and Tan's [5] global thresholding method

Table 1: Running time of each restoration step on the images shown in Figure 6, 7, 8 & 10.

Images	Size	Mask Gen. (s)	Inpainting (s)	RBF (s)	Total (s)
Figure 6 (a_1)	634×816	3.72	141.13	39.74	184.59
Figure 6 (a_2)	570×578	1.47	42.12	21.81	65.40
Figure 7 (a_1)	597×467	1.33	19.60	-	20.93
Figure 7 (a_2)	1056×928	3.64	102.59	-	106.23
Figure 8 (a_1)	640×1026	2.34	125.56	-	127.90
Figure 8 (a_2)	1104×597	2.35	97.54	-	99.89
Figure 10 (a_1)	480×320	1.04	11.91	-	12.95
Figure 10 (a_2)	1112×1134	4.22	107.01	-	111.23

is specially designed to deal with badly illuminated document images which is similar to our focus here. For comparable reasons, we only choose images with smooth shadings since Lu and Tan's method only works on shadings that fit in a polynomial distribution. In particular, the variance gains and the window size for Niblack's method are set to be -0.2 and 20, respectively. The parameters used in Lu and Tan's method follow the suggested values as given in their paper [5]. Following this setup, Figure 11 $(b)(c)(d)$ show the restoration results of the distorted textual image obtained from our method, Niblack's method and Lu and Tan's method, respectively. It is observed that Niblack's method produces a large amount of background noise at regions with uniform pixel intensities such as the white margins. This is because when such white margins fully occupy a local window, the local threshold will be determined by the shading variations instead of the uniform reflectance and thus results in an erroneous classification. Lu and Tan's method produces a better restored image but some text pixels in the upper right portion are eliminated due to over-thresholding. In contrast, our method achieves better result both visually and in term of OCR performance. Typically, the average word precision rates achieved for ten similar textual document images on their restored images using the three methods (namely, our method, Niblack's, and Lu and Tan's) are 96.8%, 77.5% and 90.4%, respectively.

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

Figure 11: Restoration result of a badly illuminated textual document image and its comparison with the results from existing methods: (a) Input image with shading distortions; (b) Restored foreground reflectance image using our method; (c) Result from Niblack's method; (d) Result from Lu and Tan's global thresholding method.

4. Geometric and Perspective Distortion Correction

Shape-based geometric correction methods have achieved rather promising results in the past decades. All these attribute to the critical shape acquisition step which provides an accurate 3D model to represent the actual physical warpings. Besides capturing shapes using special setups such as structured lighting or laser scanners, efforts have been made to reconstruct shapes using various shape recovery techniques. Here, we look at how SfS techniques can be used with a physically-based flattening method to rectify both geometric and perspective distortions in real document images.

4.1. Shape Recovery

Shape recovery is a classic and fundamental problem in computer vision. Its goal is to derive a 3D scene description from one or more 2D images. Over the years, researchers have developed a variety of techniques to tackle this problem known as Shape-from-X where X can be shading, stereo, motion, texture, etc. In particular, Shape-from-Shading tries to make use of the shading variations in a single 2D image to reconstruct the surface shape of the object. The research in this field was pioneered by Horn who first formulated the SfS problem as to find the solution of a nonlinear first-order PDE called the brightness equation [42]. Following this, a series of variational methods [43, 44, 45] are developed, which try to minimize an energy function that often comprises of an integral of the brightness error to

find the solution. Later Oliensis and Dupuis [46] propose to cast the SfS problem as an optimal control problem and directly find the depth map of the surfaces. This brought out a new set of propagation approaches based on the theory of viscosity solutions to Hamilton-Jacobi equations [47, 48, 49]. According to the numerical schemes used to estimate the viscosity solutions, these methods can be further divided into two categories. The first class of methods are based on the monotonicity of the solution along the characteristic direction including level set and fast marching methods [50, 51, 52, 53]. Various adaptations of the fast marching method have been developed to handle oblique light source [54] and perspective projection [55, 56]. On the other hand, the second class of methods make use of iteration strategies. Rouy and Tourin [47] exploit an upwind and monotone scheme to solve the discretized Eikonal equation iteratively and the convergence property is shown. Tsai et al combine the upwind monotone Godunov Hamiltonian with a Gauss-Seidel iteration method to reconstruct surfaces with good efficiency [57]. More comprehensive surveys can be found in [58, 59].

As discussed earlier, perspective distortions are caused by the perspective projection property of the pinhole camera model, which makes the faraway points appear smaller than those closer to the camera. When the document being imaged has a non-planar surface shape such as a thick bound book, the resultant image will often appear warped because the distances from the surface points to the camera vary. The surface shape can thus be represented as a depth map that records the distance from each surface point to the image plane. We can treat pure perspective distortions on planar surfaces as geometric distortions too. The goal of the SfS method here is to recover the depth map of the document surface based on the shading variations in the image so that this shape can be used for subsequent restoration procedures.

4.1.1. SfS Formulations in Document Imaging Domain

To formulate the SfS problem in the document imaging domain, we make the assumption that the document surfaces follow Lambertian reflection, no interreflections exist and sensor noise is not considered. In general, we use $z(x, y)$ to represent the depth value of a surface

point $(x, y, z(x, y))$ whose projection is denoted by (u, v) in the image plane Ω. By definition, $z(x, y) \equiv z(u, v)$. Depending on the illumination conditions, the image irradiance equation is formulated in different forms. In particular, we discuss the following two illumination situations: distant point light source and close point light source.

Distant point light source

Given a Lambertian surface, the reflectance map under a distant point light source, also known as a directional light source, is defined as:

$$I(u, v) = N \cdot L = \frac{(-p, -q, 1)}{\sqrt{p^2 + q^2 + 1}} \cdot \frac{(\alpha, \beta, \gamma)}{\sqrt{\alpha^2 + \beta^2 + \gamma^2}} \tag{8}$$

where $I(u, v)$ is the image irradiance at the image point (u, v) corresponding to the surface point (x, y), (α, β, γ) is the illumination direction and $(-p, -q, 1)$ is the surface normal at point (x, y) with $p = \frac{\partial z}{\partial u}$ and $q = \frac{\partial z}{\partial v}$. This is the general image irradiance equation under a distant oblique light source. In particular, if the light source is right on top of the surface with $L = (0, 0, 1)$, the image irradiance equation becomes the Eikonal equation:

$$\sqrt{p^2 + q^2} = \sqrt{\frac{1}{I(u, v)^2} - 1} \tag{9}$$

Close point light source

Distant point light source is often difficult to obtain in real-life situations. In fact, it is easier and more practical to capture images using close point light sources such as the on-camera flash than under specially-built lighting environment. The on-camera flash can be modeled as a close point light source with known position in the camera's coordinate system. More generally, this can be any point light source with a known location in reference to the optical center. Now, we can represent the document's surface $S(u, v)$ with respect to the image domain Ω as:

$$\left\{ S(u, v) = \frac{z(u, v)}{f} \cdot \ (u', v', f) \quad (u, v) \in \Omega \right\} \tag{10}$$

where $z(u, v)$ is the distance from $x - y$ plane to the surface point (x, y), f is the camera's focal length, (u_0, v_0) is the principle component and $u' = u - u_0$, $v' = v - v_0$. The intrinsic

parameters of the camera can be obtained using some simple camera calibration procedures [60].

To find the surface normal at each point (u, v), we calculate the tangent vectors in both u and v directions, respectively and compute their cross product. The downward normal can thus be derived as:

$$N(u,v) = (\frac{pz}{f}, \frac{qz}{f}, -\frac{z(u'p + v'q + z)}{f^2}) \tag{11}$$

where $p = \frac{\partial z}{\partial u}$ and $q = \frac{\partial z}{\partial v}$.

Suppose a close point light source is located at (α, β, γ), the illumination direction for each point (u, v) can be written as:

$$\begin{aligned} L(u,v) &= (l_1, l_2, l_3) = (\alpha - x, \beta - y, \gamma - z) \\ &= (\alpha - \frac{u'z}{f}, \beta - \frac{v'z}{f}, \gamma - z) \end{aligned} \tag{12}$$

Using Eq. 11 and 12 with the assumption of Lambertian reflection, the image irradiance equation can thus be derived from Lambert's cosine law as:

$$I(u,v) = \frac{N(u,v) \cdot L(u,v)}{\|N(u,v)\| \, \|L(u,v)\|} = \frac{n_1 l_1 + n_2 l_2 + n_3 l_3}{\sqrt{n_1^2 + n_2^2 + n_3^2}\sqrt{l_1^2 + l_2^2 + l_3^2}} \tag{13}$$

where $(n_1, n_2, n_3) = (pf, qf, -u'p - v'q - z)$.

4.1.2. Lax-Friedrichs Based Viscosity Solution

To solve the image irradiance equation in Eq. 8 & 13, we first write it in the form of a static Hamilton-Jacobi equation:

$$\begin{cases} H(u, v, z, \nabla z) = R(u, v), & (u, v) \in \Omega \\ z(u, v) = b(u, v), & (u, v) \in \Gamma \subset \Omega \end{cases} \tag{14}$$

where Ω denotes the image plane, Γ denotes a set of points whose value $z(u, v)$ is known to be $b(u, v)$, although they may be located in the interior of Ω.

In the case of a distant oblique light source as given by Eq. 8 [1], we have:

$$\begin{cases} H(u, v, \nabla z) = I\sqrt{p^2 + q^2 + 1} + p\bar{\alpha} + q\bar{\beta} - \bar{\gamma} \\ R(u, v) = 0 \end{cases} \tag{15}$$

where $(\bar{\alpha}, \bar{\beta}, \bar{\gamma})$ is the normalized illumination direction and $\nabla z = (p, q)$.

Similarly, for a close point light source as described by Eq. 13 [3], we have:

$$\begin{cases} H(u, v, z, \nabla z) & = I\sqrt{n_1^2 + n_2^2 + n_3^2}\sqrt{l_1^2 + l_2^2 + l_3^2} + Ap + Bq + C \\ R(u, v) = 0 \end{cases} \tag{16}$$

where $A = f(\alpha f - \gamma u)$, $B = f(\beta f - \gamma v)$, $C = -fz(\gamma - z)$.

To solve Eq. 15 & 16, we use an iterative sweeping strategy [57] to solve for $z(u, v)$ with an update formula based on Lax-Friedrichs Hamiltonian [61] given as:

$$\begin{aligned} z_{u,v}^{n+1} &= \frac{1}{\frac{\sigma_u}{\Delta u} + \frac{\sigma_v}{\Delta v}}(R(u, v) - H(u, v, z, p, q) + \sigma_u u_m + \sigma_v v_m) \\ p = \frac{z_{u+1,v} - z_{u-1,v}}{2\Delta u} \quad & q = \frac{z_{u,v+1} - z_{u,v-1}}{2\Delta v} \\ u_m = \frac{z_{u+1,v} + z_{u-1,v}}{2\Delta u} \quad & v_m = \frac{z_{u,v+1} + z_{u,v-1}}{2\Delta v} \end{aligned} \tag{17}$$

where $(\Delta u, \Delta v)$ is the grid size, σ_u and σ_v are artificial viscosities satisfying $\sigma_u \geq \max|\frac{\partial H}{\partial p}|$ and $\sigma_v \geq \max|\frac{\partial H}{\partial q}|$.

In particular, for Eq. 15, we let

$$\begin{aligned} \sigma_u &= \max_{u,v,p,q}\left|\frac{\partial H}{\partial p}\right| = \max_{u,v}\{\max\{|I + \bar{\alpha}|, |I - \bar{\alpha}|\}\} \\ \sigma_v &= \max_{u,v,p,q}\left|\frac{\partial H}{\partial q}\right| = \max_{u,v}\{\max\{|I + \bar{\beta}|, |I - \bar{\beta}|\}\} \end{aligned} \tag{18}$$

Similarly, for Eq. 16, we let

$$\begin{aligned} \sigma_u &= \max_{u,v,z,p,q}\left|\frac{\partial H}{\partial p}\right| = \max_{u,v}\{\max\{|p_m I_p + A|, |p_m I_p - A|\}\} \\ \sigma_v &= \max_{u,v,z,p,q}\left|\frac{\partial H}{\partial q}\right| = \max_{u,v}\{\max\{|q_m I_p + B|, |q_m I_p - B|\}\} \\ p_m &= \max_{p,q}\left|\frac{\partial H}{\partial p}\right| = \sqrt{f^2 + u'^2} \\ q_m &= \max_{p,q}\left|\frac{\partial H}{\partial q}\right| = \sqrt{f^2 + v'^2} \end{aligned} \tag{19}$$

where $I_p = I\sqrt{l_1^2 + l_2^2 + l_3^2}$.

An iterative sweeping strategy is applied to sweep through the image grid in four alternating directions to update the values of $z(u, v)$ which are initialized with the boundary

values $b(u, v)$. Height values are re-calculated at the four boundaries where the update formula fails to compute. The complexity of the fast sweeping algorithm is $O(N)$ where N is the number of grid points.

4.2. Physically-based Shape Flattening

Once the surface shape $z(u, v)$ is obtained, we will construct a uniform sampled mesh and map the image that has gone through shading correction to the reconstructed shape based on the perspective projection principle: $x/u' = y/v' = z/f$. Next, the shape will be flattened to a plane through a numerical simulation process [26, 28, 62] which subsequently corrects the geometric distortions to produce the final restored image.

5. Experimental Results for Geometric Correction

To evaluate our SfS formulations and the performance of the viscosity framework, we first conduct several experiments on synthetic surfaces which are either parametric surfaces generated using mathematical functions or geometric surfaces captured from range scans. Synthetic shadings computed using our formula are compared with real shadings and show very similar patterns. The reconstructed shapes based on synthetic shadings are also shown to be close resemblances of the ground truth shapes. Moreover, we use several real document images to demonstrate how effectively our SfS method and digital flattening procedures can handle both geometric and perspective distortions in real situations.

5.1. Results on Synthetic Surfaces

First, to evaluate the accuracy of our shading extraction routine and the correctness of our image irradiance equation, we use an analytical cylindrical shape to generate a synthetic shading image using Eq. 13 and compare it with the real shading image extracted using our inpainting and RBF-based smoothing algorithm. Figure 12 (a) is a camera image captured from a real cylinder textured with a document page. The extracted shading image is shown in Figure 12 (b). On the other hand, Figure 12 (d) is an analytical cylindrical shape used to simulate the real cylinder. Figure 12 (e) shows the synthetic shading image generated

using the same camera's focal length and the same light source location as the real imaging condition. The shading meshes shown in (c) and (f) demonstrate a high similarity. This shows that our shading extraction procedure is fairly accurate and the general image irradiance equation in Eq. 13 can be used to model real imaging situations effectively. Finally, Figure 12 (g) and (h) show the reconstructed shape based on the real shading image in (b) and the corresponding restored image, respectively.

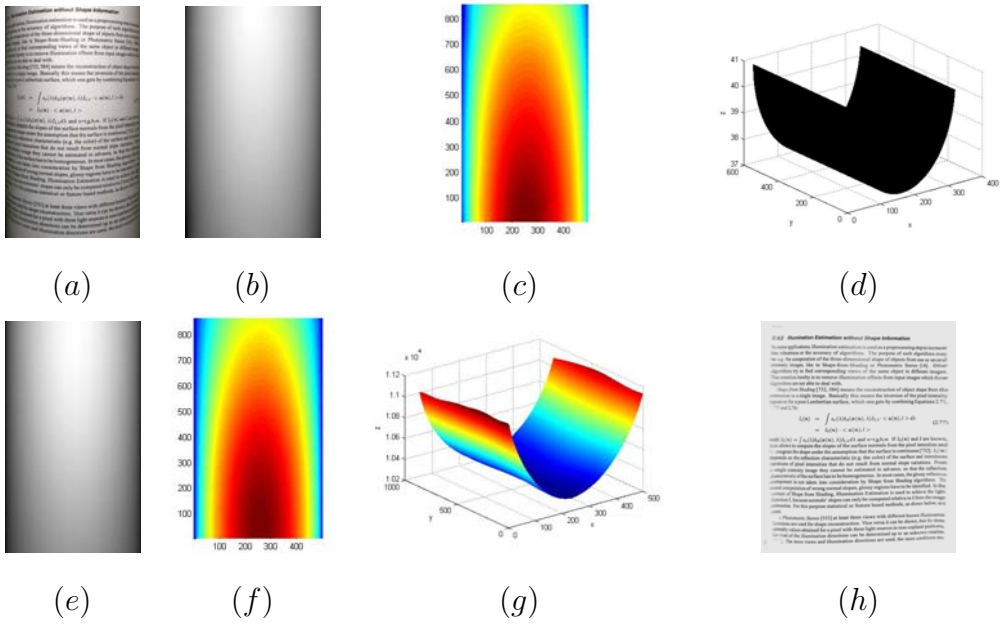


Figure 12: (a) Image of a real cylinder with a document texture; (b) Extracted shading image; (c) Color mapped shading mesh of (b); (d) Analytical cylinder; (e) Synthetic shading image produced under the same imaging condition; (f) Color mapped shading mesh of (e); (g) Reconstructed surface shape from (b); (h) Final restored image. (Color in print)

Next, we evaluate the viscosity framework by using three synthetic shading images generated from known parametric surfaces. First, the synthetic vase is generated using the formula provided in [58] as shown in Figure 13 (a_1). The grid size is set to be $\Delta x = \Delta y = 0.00625$ with an image of size 161×161. The second shape is given by Tankus [63]: $z(x,y) = 2cos(\sqrt{x^2+(y-2)^2}+100$ as shown in Figure 13 (b_1). The third shape is obtained from [64]: $f(x,y) = 2\pi\sqrt{[cos(2\pi x)sin(2\pi y)]^2 + [sin(2\pi x)cos(2\pi y)]^2}$ as shown in Figure 13 (c_1). The shading images are generated based on Eq. 9 under a distant frontal light source

$L = (0, 0, 1)$ and 8 under an oblique light source $L = (1, 0, 1)$ as shown in the second and fourth column of Figure 13, respectively. In addition, p and q are discretized using the forward difference of the surface height z.

In the case of a distant frontal light source, we use the formulation described in Eq. 15 with $(\alpha, \beta, \gamma) = (0, 0, 1)$. By applying the iterative sweeping scheme based on Lax-Friedrichs Hamiltonian, we obtain the reconstructed surface as shown in the third column of Figure 13. In particular, the vase surface is initialized with $z = 0$ along the two vertical boundaries. The second shape is initialized with its four boundary values. The last shape is initialized with the five singular points at $(0.25, 0.25), (0.75, 0.75), (0.25, 0.75), (0.75, 0.25)$ and $(0.5, 0.5)$. As we can see that the results are close resemblances of their original surfaces. In addition, we also tried to apply a high order WENO scheme [64] to solve our SfS formulation in the case of an oblique light source. This is in comparison with the low order Lax-Friedrich Hamiltonian. The results are visually closer to the original images as shown in the fifth column of Figure 13. The number of iterations and the total time taken to converge to the solution are given in Table. 2. The convergence criterion used in our experiments is $\max_{u,v} |z_{u,v}^{n+1} - z_{u,v}^{n}| \leq 0.01$.

Table 2: Efficiency evaluation on the three synthetic surfaces shown in Figure 13.

Surfaces	First order Lax-Friedrichs scheme				High order scheme	
	Frontal light source		Oblique light source		Oblique light source	
	Iterations	Time (s)	Iterations	Time (s)	Iterations	Time (s)
Figure 13(a_1)	17	3.6007	77	19.7733	81	27.6636
Figure 13(b_1)	45	5.0929	93	16.3444	192	39.3084
Figure 13(c_1)	26	5.1727	177	50.9447	195	73.7268

5.2. Comparisons Using Mozart Bust

To compare our results with those of existing approaches, we use the classic example of Mozart Bust provided by Tsai [58]. The true depth map is captured using a range scanner as shown in Figure 14 (a). Using the same shading generation method described in Section 4.1.1, we obtain the shading image under an oblique light source $L = (1, 0, 1)$ as

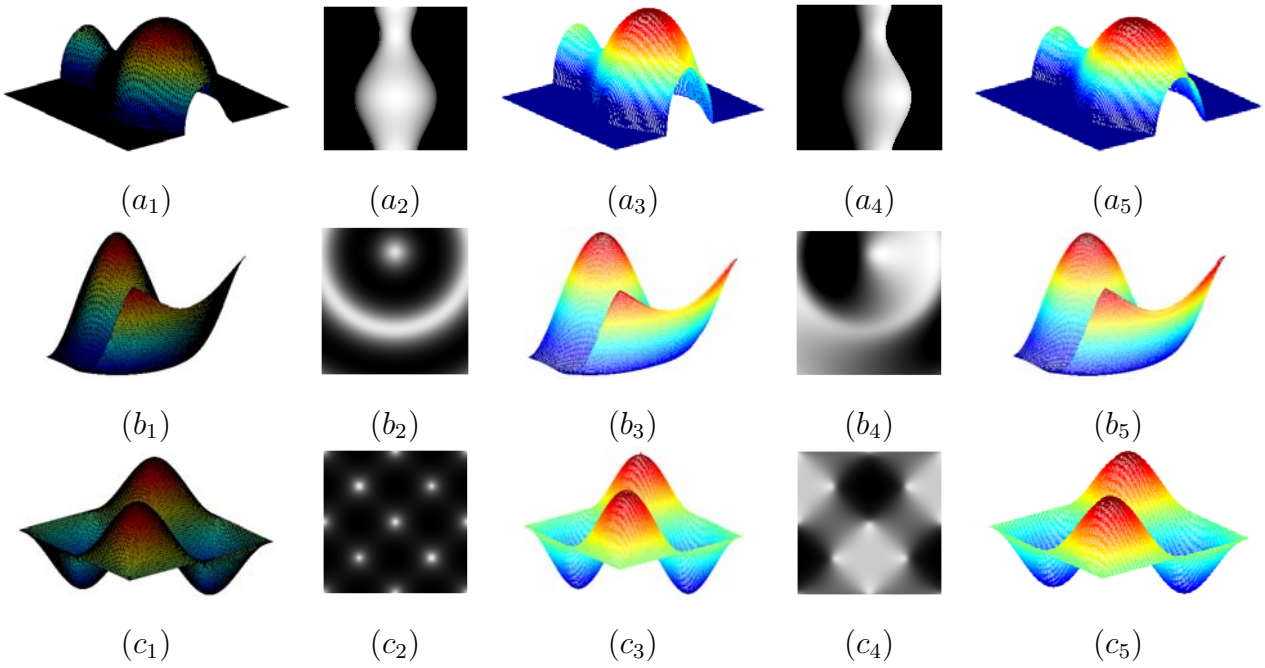


Figure 13: $(a_1)(b_1)(c_1)$ Original surface (ground truth); $(a_2)(b_2)(c_2)$ Shading image with a distant frontal light source $L = (0, 0, 1)$; $(a_3)(b_3)(c_3)$ Reconstructed surface based on Lax-Friedrichs Hamiltonian; $(a_4)(b_4)(c_4)$ Shading image with an oblique light source $L = (1, 0, 1)$; $(a_5)(b_5)(c_5)$ Reconstructed surface based on a high order WENO scheme. (Color in print)

shown in Figure 14 (b). Figure 14 (c) gives the reconstructed shape based on the Hamilton-Jacobi equation solver discussed in Section 4.1.2 with an initialization of the singular point on the nose tip. To evaluate the accuracy of the reconstructed shape, we measure its absolute distance from the original true depth map. The result shows that most regions are well aligned with an average distance of 1.18 mm. In addition, we compare our method with all the algorithms reported in [58]. Figure 14 $(d)(e)(f)$ show the results on three selected approaches. We can see that the current method gives a relatively better reconstruction.

5.3. Results on Real Document Images

Moreover, we have conducted experiments on both scanned and camera-captured real document images with various perspective or geometric distortions. One typical geometric distortion in scanned images is the spine warping that often appears when scanning thick bound documents. In our experiments, an Espon color scanner is used to get scanned images from thick bound conference proceedings. In particular, the scanner's uniform light

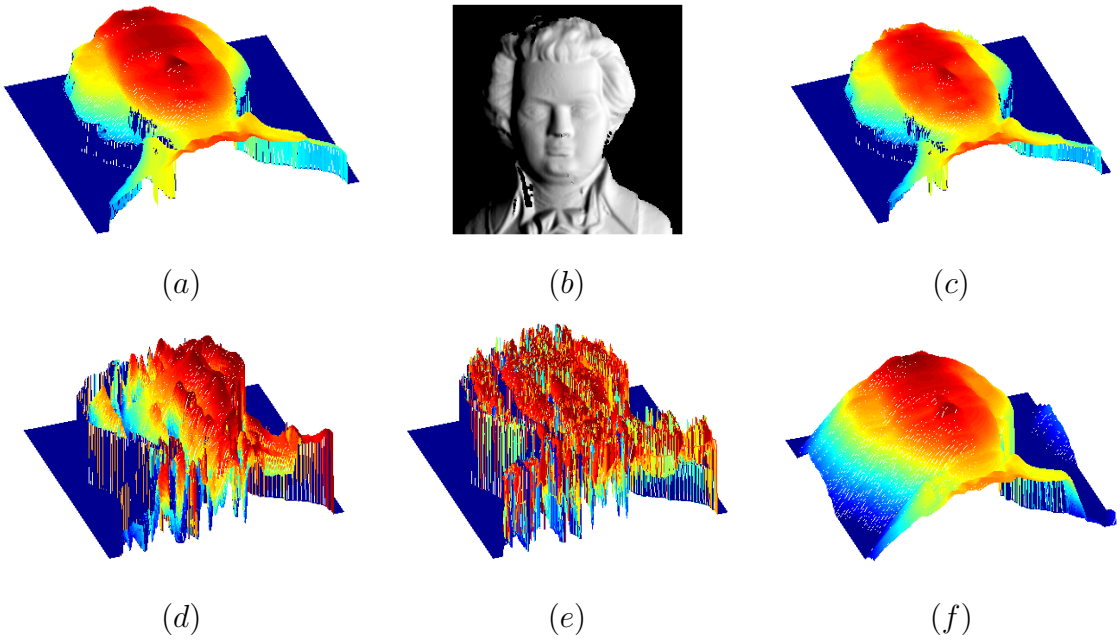


Figure 14: (a) Original depth map of Mozart Bust; (b) Shading image generated with $L = (1, 0, 1)$; (c) Shape reconstructed using the viscosity framework; (d) Shape reconstructed using Tsai's linear approximation method; (e) Shape reconstructed using Lee and Rosenfeld's method; (f) Shape reconstructed using Bichsel and Pentland's method. (Color in print)

ray is modeled as an oblique directional light source with $L = (1, 0, 5.67)$. The lens follows perspective projection along the spine direction and the focal length is obtained through a calibration process. On the other hand, all the camera images are taken in a relatively dark environment. The light source used is either the on-camera flash or a small halogen light both simulating a close point light source. The location of the light source is measured with respect to the camera's optical center. The camera's focal length f and the principle component (u_0, v_0) are obtained through a camera calibration process. Typically, for an image of size 1600×1200, we get $f = 1286$ and $(u_0, v_0) = (820, 585)$ in pixel size. The images are cropped to avoid lens distortions near the corners. The shading image is often normalized to a certain dynamic range for effective SfS processing. Harmonic inpainting algorithm is used to generate the inpainted image.

Figure 15 shows an example of scanned thick bound document with geometric distortions at the spine region. Because of the constraint imaging environment of the scanner, we have a fixed light source and a frontal imaging view. Images of thick bound books obtained from

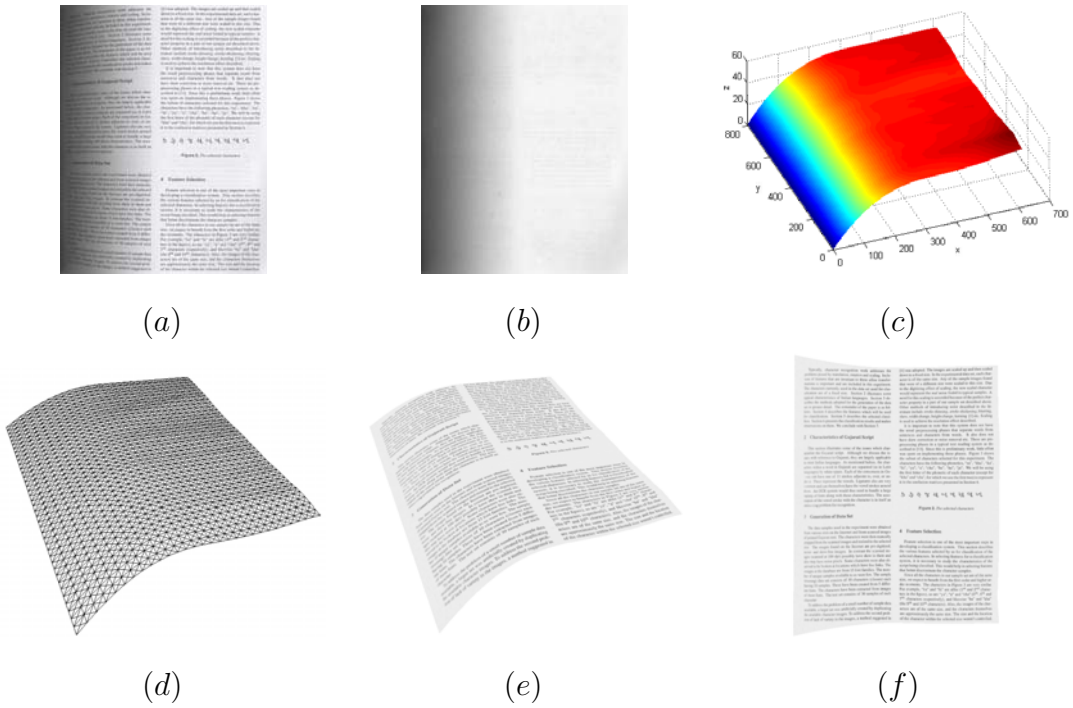
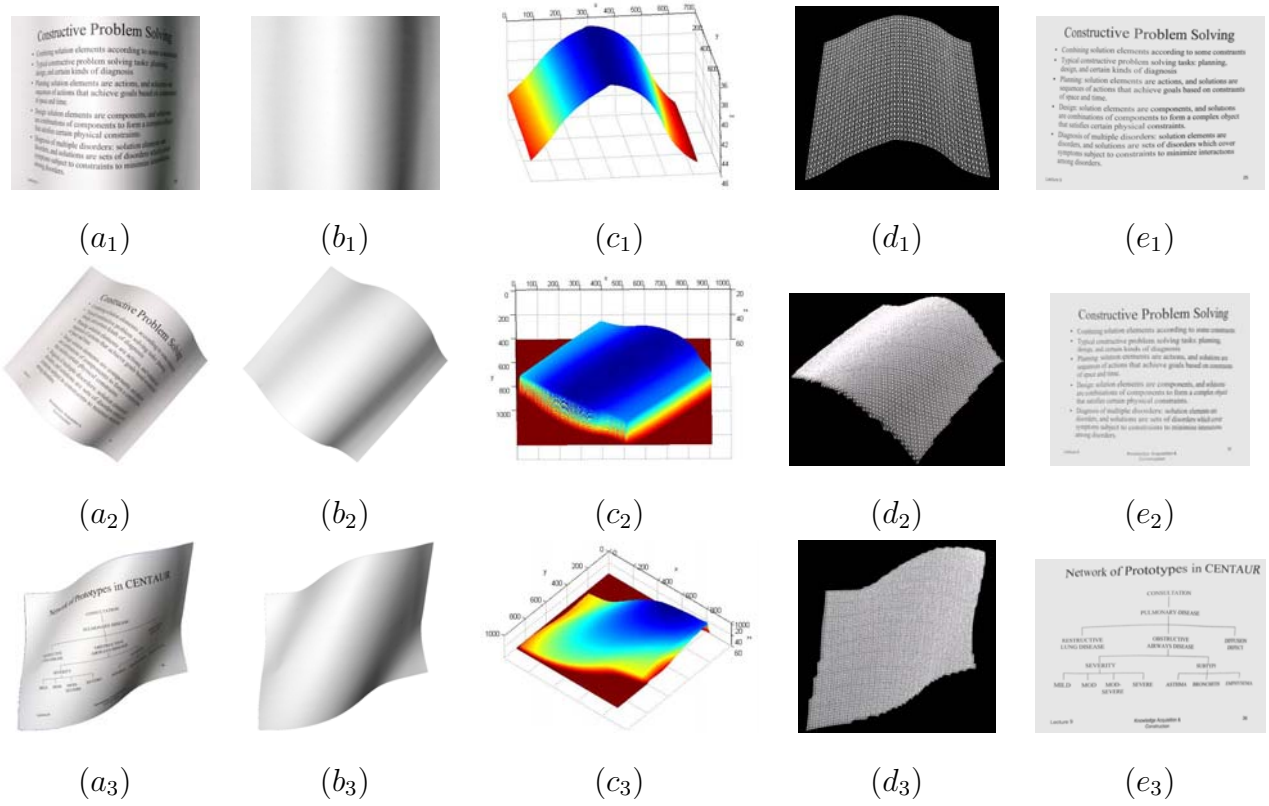


Figure 15: (a) Scanned thick bound book image; (b) Shading image obtained after one-step inpainting; (c) Shape reconstructed using the viscosity framework with an oblique light source; (d) 3D mesh of the reconstructed surface; (e) 3D mesh textured with the shading corrected image; (f) Final restored image.

this typical setup therefore have similar properties. Figure 15 shows a typical example of the restoration process and some of the intermediate results. As we can see from Figure 15 (d), the reconstructed surface shape is a close estimation of the cylindrical warping of the book spine. Some minor uneven surface patches due to water creases are also detected near the page boundary in accordance to a slight change of shading. Finally, the restored image as shown in Figure 15 (f) demonstrates a great improvement over the original distorted image and also gives a representative result when applying our framework on common scanned thick bound document images.

Figure 16 shows three examples of geometrically distorted document images with different types of contents captured under the halogen light source at arbitrary locations. In particular, Figure 16 (a_1) is an arbitrarily warped document page captured under a light source at $L = (-25, 0, -34)$. Figure 16 (c_1) is the reconstructed surface shape by initializing the singular points with an estimated distance from the surface to the camera set as $b_0 = 35.2$. Next, Figure 16 (e_1) shows the uniformly sampled mesh with the original image



(a_1) (b_1) (c_1) (d_1) (e_1)

(a_2) (b_2) (c_2) (d_2) (e_2)

(a_3) (b_3) (c_3) (d_3) (e_3)

Figure 16: $(a_1)(a_2)(a_3)$ Distorted real document images; $(b_1)(b_2)(b_3)$ Extracted shading image; $(c_1)(c_2)(c_3)$ Reconstructed surface shape; $(d_1)(d_2)(d_3)$ 3D mesh of the surface shape; $(e_1)(e_2)(e_3)$ Final restored image with all the distortions removed.

as the texture and Figure 16 (f_1) gives the final restored image.

Figure 17 shows three examples of real document images with pure perspective distortions or mixed perspective and geometric distortions. The light source used here is the on-camera flash at position $(-1.8, -2.4, 0)$ with respect to the camera's coordinate system. Figure 17 (a_1) is a planar page image with pure perspective distortions. The reconstruction shape in Figure 17 (c_1) is obtained by initializing the bottom of the page to be at a distance of $b_0 = 10$ to the camera. The black background is computed as faraway from the camera. Next, Figure 17 $(d_1)(e_1)$ show the uniformly sampled mesh and the one with the shading-corrected image as the texture. Finally, Figure 17 (f_1) gives the restored image. In addition, Figure 17 $(a_2)(a_3)$ show that the proposed method is also able to handle images with both perspective and geometric distortions.

Both Figure 16 and 17 show that the restored images are greatly improved compared

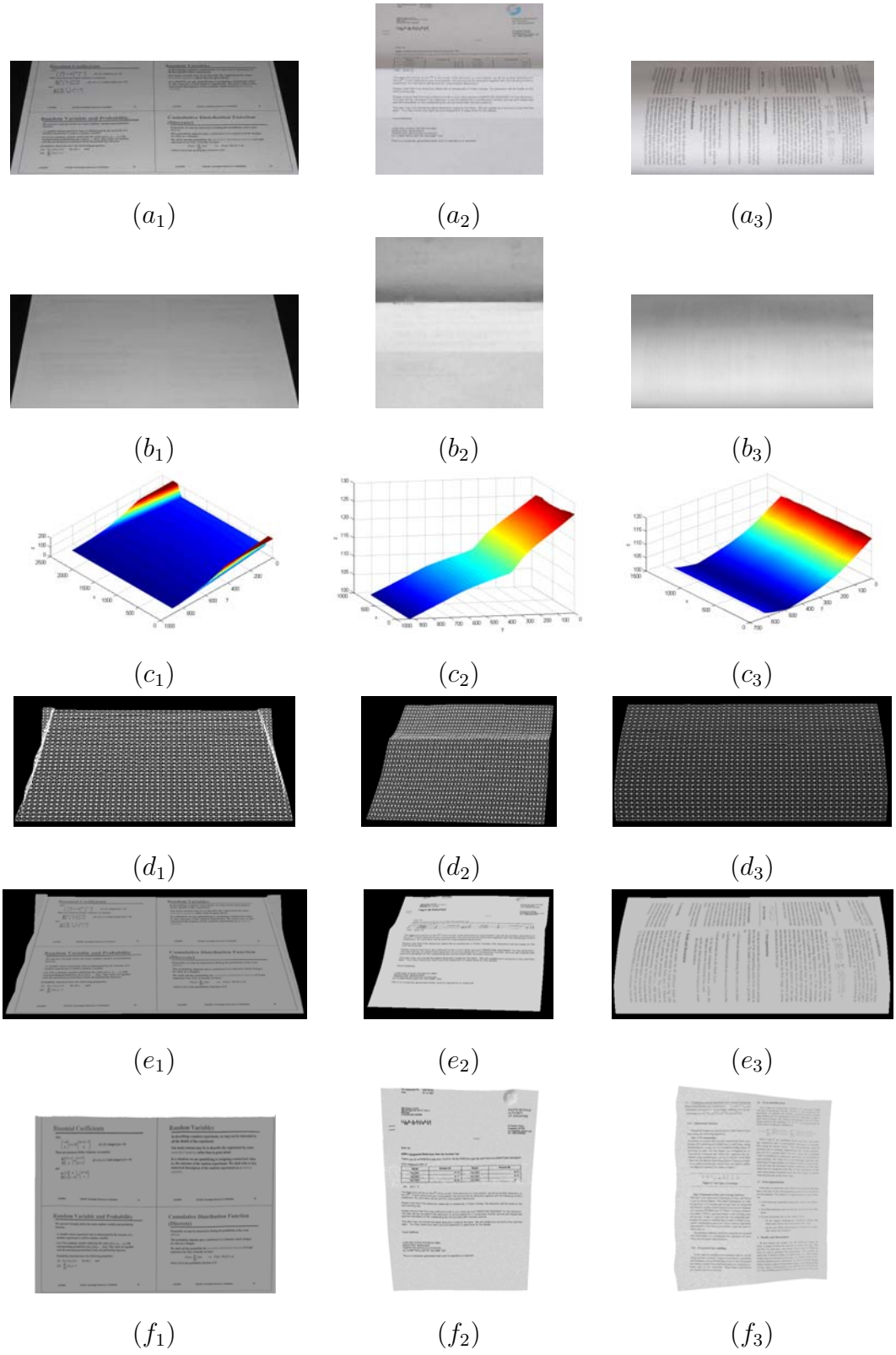


Figure 17: $(a_1)(a_2)(a_3)$ Camera-captured document images with distortions; $(b_1)(b_2)(b_3)$ Extracted shading image; $(c_1)(c_2)(c_3)$ Reconstructed surface shape; $(d_1)(d_2)(d_3)$ 3D mesh of the surface shape; $(e_1)(e_2)(e_3)$ 3D mesh with the shading-corrected image as texture; $(f_1)(f_2)(f_3)$ Final restored image with all the distortions removed.

to the original images although some distortions still remain due to the inaccurate surface reconstruction based on irregular real-world shadings. The OCR performance is greatly improved on the restored images with an average increase of 48.6% in terms of word precision over a set of 30 document images with approximately 3,500 words. Table 3 gives the word precision P_w and character precision P_c on the six examples shown in Figure 16 and 17. One thing worth mentioning is that Figure 17 (a_3) does not achieve a good OCR performance even after restoration because a great loss of pixel values results in a blurring effect after the rectification. This can be improved by using a super resolution technique to further enhance the quality of the image.

6. Discussion

Despite the good performance the proposed photometric correction procedure on a general set of document images, it may also produce less satisfactory results in some extreme cases or on certain types of images. For example, in the case of the bleed-through removal, if the degradations are so severe that the bleed-through pixels are even more intense than the foreground pixels, confusions will arise. Figure 18 (a_1) gives an example of a degraded image with serious bleed-through distortions. After applying the inpainting method, the restored image as shown in Figure 18 (d_1) is still difficult to read because many bleed-through pixels are detected incorrectly as foreground pixels while many original foreground pixels are eliminated. Another situation is when the original image contains large color figures, some parts of the figures will not be correctly masked due to their large sizes and thus the shadings around the figures will not be correctly approximated. This results in errors in the final restored image at the figure region. One way to remedy this problem is to detect large figure components beforehand and mask them as a whole. The inpainting method will then fill in the whole region accordingly. Although there are still some errors in the approximated shadings inside the filled region, the restored image is overall satisfactory if the primary concern is not on the fine-grained details of the figure itself as shown in Figure 18 (a_2).

The SfS method we propose here is able to effectively recover smoothly warped surfaces when the extracted shading image is of good quality. Due to the unconstrained imaging

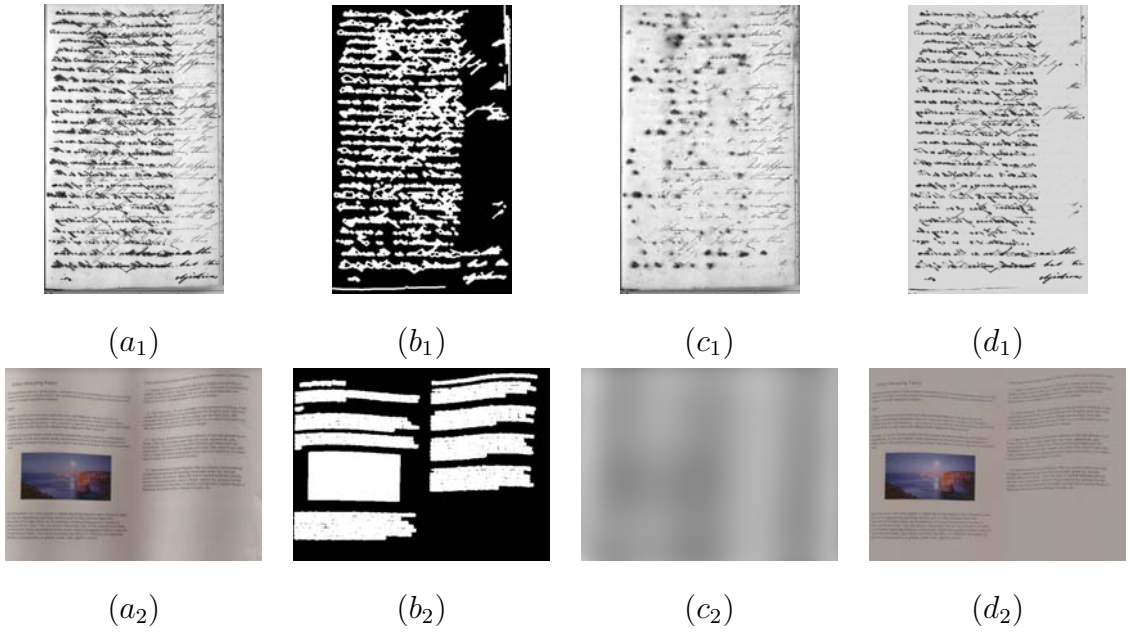


Figure 18: Restoration results of real document images with severe background noise or large embedded figures: $(a_1)(a_2)$ Original distorted image; $(b_1)(b_2)$ Extracted inpainting mask; $(c_1)(c_2)$ Reconstructed background layer image; $(d_1)(d_2)$ Restored image.

environment when using digital cameras and the difficulty in extracting a perfect shading image, it is hard to obtain a very accurate reconstruction from real images as in the synthetic case. However, our objective is to be able to use a single input image to produce a rough estimation of the physical warping so that it can be used to restore the geometric and perspective distortions. As our experiments demonstrate, the restored images based on this approach are greatly improved compared to the original images in terms of the OCR performance. In addition, the SfS approach integrates well into our full restoration framework, which performs all the restorations at one go including shadings, shadows, perspective and geometric distortions. Despite the simplicity and effectiveness of the SfS method, there are still limitations. In particular, the approach targets smooth surface variations better than irregular folds or crumples, although it is still able to recover the ridges to a certain smoothness level as shown in Figure 17 (c_2). Furthermore, due to the importance of the shadings in the reconstruction process, certain requirements on the imaging environment need to be followed such as the use of a single point light source in camera imaging. Some boundary values need to be predefined in order to compute the viscosity solution. Inter-reflections

Table 3: OCR results on both original images and restored images in Figure 16 & 17.

OCR Results		Figure 16			Figure 17		
		(a_1)	(a_2)	(a_3)	(a_1)	(a_2)	(a_3)
Original image	P_w	65.9%	58.3%	27.3%	69.8%	73.0%	37.6%
	P_c	68.6%	62.4%	21.6%	70.3%	76.8%	38.2%
Restored image	P_w	95.7%	91.7%	84.8%	96.5%	94.6%	70.2%
	P_c	96.3%	93.1%	83.4%	97.3%	95.5%	74.8%

also need to be avoided because they may affect shadings significantly. Last but not least, if large text blocks or figures are masked, the inpainting method may not be able to estimate an accurate shading image and therefore cause errors in the reconstructed surface shape. Nevertheless, we can always substitute the SfS method with any other shape recovery methods such as methods based on Shape-from-Stereo, Structure-from-Motion or even use special setups to capture the surface shape if they are available. With all the above substitutes, the other components of the whole framework are not affected, which demonstrates great flexibility and modularity.

7. Conclusion

We have presented a restoration framework based on inpainting and Shape-from-Shading that can be used to correct a variety of distortions in imaged documents. The inpainting component can be used to separate out the illumination from the printed content. This is useful for illumination rectification as well as aiding in the reduction of noise such as show-through and ink-bleed. SfS is used to acquire a 3D approximation of the document surface which can be used to correct geometric and perspective distortions. The whole framework is tested on various real document images for each of the component and shows good results. Although the restored images are not perfect due to various restrictions in the document contents or the imaging environment, this framework does demonstrate a functional solution to restore various distortions of a single input image without the help of any special setups. Further improvements of the inpainting method and the SfS method are being explored,

which will allow an even better performed system.

Acknowledgment

This research is supported in part by National University of Singapore URC grant R252-000-202-112 and Agency for Science, Technology and Research (A*STAR) grant 0421010085.

References

[1] L. Zhang, A. Yip, C. Tan, Shape-from-shading based on lax-friedrichs fast sweeping and regularization techniques with applications to document image restoration, IEEE Conf. on Computer Vision and Pattern Recognition, 2007.

[2] L. Zhang, A. Yip, C. Tan, Photometric and geometric restoration of document images using inpainting and shape-from-shading, 22nd AAAI Conference on Artificial Intelligence, 2007.

[3] L. Zhang, A. Yip, C. Tan, A restoration framework for correcting photometric and geometric distortions in camera-based document images, Int'l Conf. on Computer Vision, 2007.

[4] Z. Zhang, C. Tan, Correcting document image warping based on regression of curved text lines, 7th Int'l Conf. on Document Analysis and Recognition 1 (2003) 589–593.

[5] S. Lu, C. L. Tan, Thresholding of badly illuminated document images through photometric correction, ACM Symposium on Document Engineering (2007) 3–8.

[6] L. Zhang, Z. Zhang, C. Tan, T. Xia, 3d geometric and optical modeling of warped document images, IEEE Int'l Conf. on Computer Vision and Pattern Recognition 1 (2005) 337–342.

[7] L. Zhang, A. Yip, C. Tan, Removing shading distortions in camera-based document images using inpainting and surface fitting with radial basis functions, International Conference on Document Analysis and Recognition (2007) 984–988.

[8] M. Brown, Y. Tsoi, Geometric and shading correction for images of printed materials using boundary, IEEE Trans. on Image Processing 15 (6) (2006) 1544–1554.

[9] M. Sun, R. Yang, L. Yun, G. Landon, B. Seales, M. Brown, Geometric and photometric restoration of distorted documents, IEEE Int'l Conf. on Computer Vision 2 (2005) 17–21.

[10] H. Barrow, J. Tenenbaum, Recovering intrinsic scene characteristics from images, Computer Vision Systems (1978) 3–26.

[11] B. V. Funt, M. S. Drew, M. Brockington, Recovering shading from color images, 2nd European Conference on Computer Vision (1992) 124–132.

[12] M. Tappen, W. Freeman, E. Adelson, Recovering intrinsic images from a single image, IEEE Trans. on Pattern Recognition and Machine Intelligence 27 (9) (2005) 1459–1472.

[13] Y.Weiss, Deriving intrinsic images from image sequences, IEEE Int'l Conf. on Computer Vision 2 (2001) 68–75.

[14] J. Toro, D. Ziou, M. F. Auclair-Fortier, Recovering the shading image under known illumination, 1st Canadian Conf. on Computer and Robot Vision (2004) 92–96.

[15] I. Bar-Yosef, I. Beckman, K. Kedem, I. Dinstein, Binarization, character extraction, and writer identification of historical hebrew calligraphy documents, International Journal on Document Analysis and Recognition (2007) 89–99.

[16] G. Leedham, S. Varma, A. Patankar, V. Govindaraju, Separating text and background in degraded document images - a comparison of global threshholding techniques for multi-stage threshholding, 8th International Workshop on Frontiers in Handwriting Recognition (2002) 244–249.

[17] F. Drira, Towards restoring historic documents degraded over time, Second International Conference on Document Image Analysis for Libraries (2006) 350–357.

[18] W. Boussellaa, A. Zahour, A. Alimi, A methodology for the separation of foreground/background in arabic historical manuscripts using hybrid methods, ACM symposium on Applied computing (2007) 605–609.

[19] C. L. Tan, R. Cao, P. Shen, Restoration of archival documents using a wavelet technique, IEEE Trans. on Pattern Analysis and Machine Intelligence 24 (2002) 1399–1404.

[20] A. Tonazzini, E. Salerno, M. Mochi, L. Bedini, Bleed-through removal from degraded documents using a color decorrelation method, Document Analysis Systems (2004) 229–240.

[21] P. Clark, M. Mirmehdi, Rectifying perspective views of text in 3d scenes using vanishing points, Pattern Recognition 36 (2003) 2673–2686.

[22] M. Pilu, Extraction of illusory linear clues in perspectively skewed documents, Computer Vision and Pattern Recognition (2001) 363–368.

[23] G. K. Myers, R. C. Bolles, Q. T. Luong, J. A. Herson, H. B. Aradhye, Rectification and recognition of text in 3-d scenes, International Journal on Document Analysis and Recognition (2005) 147–158.

[24] Y. Tang, C. Suen, Image transformation approach to nonlinear shape restoration, IEEE Trans. on Systems, Man, and Cybernetics 23 (1) (1993) 155–171.

[25] M. Pilu, Undoing page curl distortion using applicable surfaces, Int'l Conf. on Computer Vision and Pattern Recognition 1 (2001) 67–72.

[26] M. Brown, W. Seales, Image restoration of arbitrarily warped documents, IEEE Trans. on Pattern Analysis and Machine Intelligence 26 (10) (2004) 1295–1306.

[27] M. Brown, C. Pisula, Conformal deskewing of non-planar documents, IEEE Computer Society Conf. on Computer Vision and Pattern Recognition 1 (2005) 998–1004.

[28] K. Chua, L. Zhang, Y. Zhang, C. Tan, A fast and stable approach for restoration of warped document

images, 8th Int'l Conf. on Document Analysis and Recognition 1 (2005) 384–388.

[29] Y. Lin, W. Seales, Opaque document imaging: Building images of inaccessible texts, IEEE Int'l Conf. on Computer Vision 1 (2005) 662–669.

[30] H. Cao, X. Ding, C. Liu, A cylindrical model to rectify the bound document image, IEEE Int'l Conf. on Computer Vision 2 (2003) 228–233.

[31] F. Courteille, A. Crouzil, J.-D. Durou, P. Gurdjos, Shape from shading for the digitization of curved documents 18 (5) (2007) 301–316.

[32] A. Yamashita, A. Kawarago, T. Kaneko, K. Miura, Shape reconstruction and image restoration for non-flat surface of document with a stereo vision system, 17th Int'l Conf. on Pattern Recognition 1 (2004) 482–485.

[33] A. Iketani, T. Sato, S. Ikeda, Video mosaicing for curved documents based on structure from motion, 18th Int'l Conf. on Pattern Recognition 1 (2006) 391–396.

[34] P. Clark, M. Mirmehdi, Recognizing text in real scenes, Int'l Journal on Document Analysis and Recognition 4 (2002) 243–257.

[35] M. Pietikainen, O. Okun, Edge-based method for text detection from complex document images, 6th Int'l Conf. on Document Analysis and Recognition (2001) 286–291.

[36] Y. Zhong, H. Zhang, A. Jain, Automatic caption localization in compressed video, IEEE Pattern Analysis and Machine Intelligence 22 (4) (2000) 385–392.

[37] H. Li, D. Doermann, O. Kia, Automatic text detection and tracking in digital video, IEEE Trans. on Image Processing 9 (1) (2000) 147–156.

[38] M. Bertalmio, G. Sapiro, C. Ballester, V. Caselles, Image inpainting, SIGGRAPH 1 (2000) 417–424.

[39] T. Chan, J. Shen, Mathematical models for local nontexture inpaintings, SIAM Journal on Applied Mathematics 62 (3) (2002) 1019–1043.

[40] J. Carr, R. Beatson, B. McCallum, W. Fright, T. McLennan, T. Mitchell, Smooth surface reconstruction from noisy range data, Graphite 2003 (2003) 119 –297.

[41] W. Niblack, An Introduction to Digital Image Processing, Prentice-Hall, Englewood Cliffs, 1986.

[42] B. Horn, Obtaining shape from shading information, McGraw-Hill, New York, 1975.

[43] K. Ikeuchi, B. Horn, Numerical shape from shading and occluding boundaries, Journal of Artificial Intelligence 17 (1-3) (1981) 141–184.

[44] B. Horn, M. Brooks, The variational approach to shape from shading, Computer Vision, Graphics, and Image Processing 33 (2) (1986) 174–208.

[45] R. Frankot, R. Chellappa, A method for enforcing integrability in shape from shading algorithms, IEEE Pattern Analysis and Machine Intelligence 10 (1988) 439–451.

[46] J. Oliensis, P. Dupuis, Direct method for reconstructing shape from shading, SPIE 1 (1991) 116–128.

[47] E. Rouy, A. Tourin, A viscosity soluition approach to shape from shading, SIAM Journal on Numerical Analysis 29 (3) (1992) 867–884.

[48] E. Prados, O. Faugeras, E. Rouy, Shape from shading and viscosity solutions, European Conf. on Computer Vision 2 (2002) 790–804.

[49] E. Prados, F. Camilli, O. Faugeras, A unifying and rigorous shape from shading method adapted to realistic data and applications, Journal of Mathematical Imaging and Vision 25 (3) (2006) 307–328.

[50] S. Osher, J. Sethian, Fronts propagating with curvature dependent speed: Algorithms based on the hamilton-jacobi formulation, Journal of Computational Physics 79 (1988) 12–49.

[51] R. Kimmel, A. Bruckstein, Tracking level sets by level sets: A method for solving the shape from shading problem, Computer Vision and Image Understanding 62 (2) (1995) 47–58.

[52] J. Sethian, Level Set Methods and Fast Marching Methods, Cambridge University Press, New York, 1999.

[53] E. Prados, S. Soatto, Fast marching method for generic shape from shading, 3rd Int'l Workshop on Variational, Geometric and Level Set Methods in Computer Vision 1 (2005) 320–331.

[54] R. Kimmel, J. Sethian, Optimal algorithm for shape from shading and path planning, Journal of Mathematical Imaging and Vision 14 (2) (2001) 237–244.

[55] A. Tankus, N. Sochen, Y. Yeshurun, Perspective shape-from-shading by fast marching, IEEE Conf. on Computer Vision and Pattern Recognition 1 (2004) 43–49.

[56] S. Yuen, Y. Tsui, Y. Leung, R. Chen, Fast marching method for shape from shading under perspective projection, Visualization, Imaging, and Image Processing (2002) 584–589.

[57] Y. Tsai, L. Cheng, S. Osher, H. Zhao, Fast sweeping algorithms for a class of hamilton-jacobi equations, SIAM J. on Numerical Analysis 41 (2) (2003) 659–672.

[58] R. Zhang, P. Tsai, J. Cryer, M. Shah, Shape from shading: A survey, IEEE Trans. on Pattern Analysis and Machine Intelligence 21 (8) (1999) 690–706.

[59] J. Durou, M. Falcone, M. Sagona, A survey of numerical methods for shape from shading, research Report from IRIT (2004).

[60] Z. Zhang, Flexible camera calibration by viewing a plane from unknown orientations, In Proceedings of IEEE International Conference on Computer Vision 1 (1999) 666–673.

[61] C. Kao, S. Osher, J. Qian, Lax-friedrichs sweeping scheme for static hamilton-jacobi equations, Journal of Computational Physics 196 (1) (2004) 367–391.

[62] L. Zhang, A. Yip, C. Tan, An improved physically-based method for geometric restoration of distorted document images, IEEE Trans. on Pattern Analysis and Machine Intelligence 30 (4) (2008) 728–734.

[63] A. Tankus, N. Sochen, Y. Yeshurun, A new perspective on shape-from-shading, IEEE International Conference on Computer Vision 2 (2003) 862–869.

[64] Y. Zhang, H. Zhao, J. Qian, High order fast sweeping methods for eikonal equations, Proc. of the 74th SEG Annual International Meeting (2004) 1901–1904.

Vitae

Li Zhang received the B.Sc. (Hons), M.Sc. and Ph.D. degree from school of computing, National University of Singapore in 2003, 2004 and 2008, respectively. She is a member of the Pattern Recognition and Machine Intelligence Association (PREMIA) in Singapore. Her research interests include computer vision, pattern recognition, document image analysis and processing, information extraction and retrieval.

Andy M. Yip received the B.Sc. in mathematics from the Chinese University of Hong Kong in 1998, the M.Phil. degree in mathematics from the University of Hong Kong in 2000, and the Ph.D. in mathematics from University of California, Los Angeles in 2005. He joined the Department of Mathematics at National University of Singapore in July 2005 as an Assistant Professor. His research interests include variational and PDE methods in image processing and data mining algorithms.

Michael S. Brown obtained the BS and PhD in Computer Science from the University of Kentucky in 1995 and 2001 respectively. He was a visiting PhD student at the University of North Carolina at Chapel Hill from 1998-2000. He is currently the Sung Kah Kay Assistant Professor in the School of Computing at the National University of Singapore. His research interests include Computer Vision, Image Processing and Computers Graphics.

Chew Lim Tan is a Professor in the Department of Computer Science, School of Computing, National University of Singapore. He received his B.Sc. (Hons) degree in physics in 1971 from University of Singapore, his M.Sc. degree in radiation studies in 1973 from University of Surrey, UK, and his Ph.D. degree in computer science in 1986 from University of Virginia, U.S.A. His research interests include document image analysis, text and natural language processing, neural networks and genetic programming. He has published more than 300 research publications in these areas. He is an associate editor of Pattern Recognition, associate editor of Pattern Recognition Letters, an editorial member of the International

Journal on Document Analysis and Recognition. He is a member of the Governing Board of the International Association of Pattern Recognition (IAPR). He is also a senior member of IEEE.