Origami: Multiview Rectification of Folded Documents

Shaodi You The University of Tokyo Yasuyuki Matsushita Sudipta Sinha Microsoft Research Yusuke Bou Microsoft

Katsushi Ikeuchi The University of Tokyo

Abstract

Digitally unwrapping paper sheets is a crucial step for document scanning and accurate text recognition. This paper presents a method for automatically rectifying curved or folded paper sheets from a small number of images captured from different viewpoints. Unlike previous techniques that require either an expensive 3D scanner or over-simplified parametric representation of the deformations, our method only uses a few images and is based on general developable surface model that can represent diverse sets of deformation of paper sheets. By exploiting the geometric property of developable surfaces, we develop a robust rectification method based on ridge-aware 3D reconstruction of the paper sheet and ℓ_1 conformal mapping. We evaluate the proposed technique quantitatively and qualitatively using a wide variety of input documents, such as receipts, book pages and letters.

1. Introduction

Digitally recording paper documents for editing and sharing is a common task in our daily life. In practice, paper documents are often curved or folded, and proper rectification is needed for subsequent use of recorded documents, such as text recognition and digital editing. One may physically rectify the paper documents using a flatbed scanner for a class of paper sheets; however, there are a wide variety of documents that are not easy to do so, *e.g.*, pages of an opened book. Therefore, a method of digitally rectifying paper documents is desired, and there have been several studies for achieving this goal.

There are two major challenges in rectifying paper documents. One is accurately inferring the 3D shape of a curved and folded paper sheet, and the other is flattening the inferred shape without introducing distortions. To infer the 3D shape of curved paper sheets, previous approaches either use a specialized hardware setup [18, 17, 1] or assume simplified parametric shape models [28, 30, 26, 10, 21, 32, 17], such as a generalized cylindrical surface (Fig. 2.a). While these methods are shown effective, they are applicable in a rather limited context due to the requirement of bulky setups or restricted classes of paper sheet deformations.

This paper presents a method for digitally rectifying arbitrarily curved and folded paper sheets from a few images recorded from uncalibrated viewpoints. Unlike previous techniques, our method work only with images taken from hand-held cameras; thus, it is more applicable to a wide variety of scenarios. Our method relies on structure from motion (SfM) to obtain the initial sparse 3D point cloud from uncalibrated images. To infer an accurate and dense shape of the paper sheet without losing the high-frequency structures such as folds and creases, we develop a *ridge-aware* surface reconstruction method. To effectively neglect outliers that may present in the initial sparse 3D point cloud due to repetitive textures in document images, we formulate the dense surface reconstruction problem as robust Poisson surface reconstruction using ℓ_1 optimization regularized by the ridge-awareness. For unwrapping, we develop a robust conformal mapping technique by incorporating the ridgeawareness and ℓ_1 optimization in order to avoid global distortion and effect of outliers. The overview of the proposed method is illustrated in Fig. 1.

The primary contributions of our work are threefold. First, we use a ridge-aware regularization in both 3D surface reconstruction and flattening (conformal mapping) to ensure accuracy of each step. The ridge-aware regularization in surface reconstruction enables preserving the sharp structure of folds and creases, and that in flattening avoids global distortion by serving as a non-local regularizer. Second, we extend the conventional Poisson surface reconstruction [9] and least-squares conformal mapping (LSCM) [11] to explicitly deal with outliers by introducing their variants based on a ℓ_1 solution technique. Third, the unique combination of these techniques result in a practical system that allows to rectify curved and folded paper sheets in a convenient manner.









Figure 2. Developable surfaces, specifically a paper sheet, has underlying rulers (straight Gaussian curvatures). Dotted lines indicate the rulers and solid lines indicate the fold-lines/ridges. (a) Generalized cylindrical surface of which the rulers are parallel to each other. (b) Cylinder-like developable surfaces of which the rulers are almost parallel. (c) General developable surface of which the rulers and ridge are in arbitrary direction.

2. Related Work

Digital rectification of curved and folded paper sheets has been actively explored in the past two decades in both the computer vision and document processing areas.

As illustrated in Fig. 2.a, many existing methods assume the paper is curved only in one direction (generalized cylindrical surface) so that it can be parameterized using a 1D smooth function. With this assumption, a variety of techniques can be used to obtain the geometry. Shape from shading has first been used by Wada *et al.* [28], Tan et al. [33, 22], Courteille et al. [5] and Zhang et al. [30]. Shape from boundary method is explored by Tsoi et al. [25, 26]. Multi-view stereo with well calibrated binocular cameras is used by Yamashita et al. [29], Koo et al. [10] and Tsoi et al. [26]. Shape from text line is well explored in both computer vision area and document processing area [4, 33, 6, 27, 15, 7, 21, 16, 32], which assume the content is well formatted print out characters. Liang et al. [12] and Tian et al. [23] also use shape from text line and relaxed the geometry model that surface rulers are not necessary to be strictly parallel (cylinder-like developable surface, Fig. 2.b). With parametric expression, document rectification is through find the inverse function. Although the above method works with a single image input, the strong assumptions on surface geometry, contents and illumination limit the applicability.

Alternatively, in order to rectify documents with arbitrary distortion and contents, existing methods employ special devices. Brown *et al.* [1] use a calibrated mirror system to obtain 3D geometry using multi-view stereo. and unwrap the surface with constraints on elastic energy, gravity and collision. The model is inaccurate because developable surfaces are not elastic. Later [2], they use range sensor to directly obtain dense 3D points and flatten the surface using least square conformal mapping [11]. Zhang et al. [31] also use range finder to obtain dense point cloud, and the flattening is done by modifying the elastic constraints to rigid constraints based on Brown et al.'s method [1]. Pilu assumes the dense 3D mesh is already obtained and minimizes the global bending potential energy to flatten the surface [18]. Recently, Meng et al. designed a calibrated active structural light device to retrieve the two parallel 1D curvatures [17]. Their method works with gray-scale contents. None of these existing methods are as practical and convenient as our method that only requires a hand-held camera.

3. Proposed Method

Our method consists of two steps: 3D document surface reconstruction and unwrapping of the reconstructed surface. For now, let us assume that sparse 3D points on the target surface is obtained from the input images via SfM. We will describe the detail of the input and SfM later as implementation details. In the following, we describe the two key steps: ridge-aware surface reconstruction, and robust surface unwrapping.

3.1. Ridge-aware surface reconstruction

One of the major challenges in accurate 3D reconstruction of folded papers from sparse 3D points is to retain high-frequency ridges. Due to the lack of density of the given 3D points, such ridges are typically smoothed out if a conventional interpolation method is used. In addition, for a document-like scene where repetitive textures are commonly observed, outliers in the sparse 3D point estimates



Figure 3. The robust Poisson reconstruction and ridge-aware reconstruction from sparse points.

need to be taken care. We address these problems by developing a robust ridge-aware surface reconstruction method. The proposed method is built upon Poisson surface reconstruction [9], and we make two important modifications to the original method by adding (1) robustness against outliers and (2) ridge-awareness.

Robust Poisson surface reconstruction We denote a set of sparse points that are obtained from SfM as $\{\hat{x}_n, \hat{y}_n, \hat{z}_n\}, n = 1, 2, \cdots, N$, where N is the number of the points. We use the sparse points that are seen from at least three images. For our document input, the typical value of N ranges from 700 to 2000. By properly choosing a reference view, we use a depth map parameterization z(x, y). Now we wish to recover depth values of mesh grid vertices $z_i(x_i, y_i)$, where *i* is the mesh grid index, $1 \le i \le I$. Our method determines depth values $\mathbf{z} = [z_1, \ldots, z_I]^{\top}$ at grid points *i*. The optimal depth values function:

$$\mathbf{z}^* = \operatorname*{argmin}_{\mathbf{z}} E_d(\mathbf{z}) + \lambda E_s(\mathbf{z}), \tag{1}$$

where E_d and E_s are data and smoothness terms, respectively, and λ is a weighting factor for controlling smoothness. While the original Poisson surface reconstruction uses the squared ℓ_2 -norm for both terms, our method uses ℓ_1 norm for the data term E_d for improving robustness against outliers as

$$E_d(\mathbf{z}) = \sum_n \|\hat{z}_n - z_i\|_1.$$
 (2)

In Eq. (2), consistency among the original data points \hat{z}_n and the corresponding depth values z_i is ensured without fitting outliers contained in $\{\hat{z}_n\}$. In a vector form, it is re-written as

$$E_d(\mathbf{z}) = \|\hat{\mathbf{z}} - \mathcal{P}_{\Omega}\mathbf{z}\|_1, \qquad (3)$$

where \mathcal{P}_{Ω} is a permutation matrix that selects and aligns observed entries Ω by ensuring the correspondence between \hat{z}_n and z_i . The smoothness term E_s is defined using the squared Frobenius norm of the gradient of depth vector \mathbf{z} along x and y directions

$$E_s(\mathbf{z}) = \|\nabla^2 \mathbf{z}\|_F^2 = \left\| \left[\frac{\partial^2 \mathbf{z}}{\partial x^2}, \ \frac{\partial^2 \mathbf{z}}{\partial y^2} \right] \right\|_F^2.$$
(4)

By preparing a sparse derivative matrix D that replaces the Laplace operator ∇^2 in a linear form:

$$\mathbf{D} = \begin{bmatrix} d_{i,j} = \begin{cases} 2 & \text{if } i = j \\ -1 & \text{if } z_j \text{ is left/right to } z_i \\ 0 & \text{otherwise} \end{cases}, \quad (5)$$
$$d_{i+I,j} = \begin{cases} 2 & \text{if } i = j \\ -1 & z_j \text{ is above/below } z_i \\ 0 & \text{otherwise} \end{cases}$$

the problem is then viewed as a specialized form of the Lasso problem [24] as

$$\mathbf{z}^* = \underset{\mathbf{z}}{\operatorname{argmin}} \|\hat{\mathbf{z}} - \mathcal{P}_{\Omega}\mathbf{z}\|_1 + \lambda \|\mathbf{D}\mathbf{z}\|_2^2.$$
(6)

While the problem of Eq. (6) does not have a closed form solution, we employ a type of iteratively reweighted least squares (IRLS) method [3] for efficiently deriving the solution. By rewriting the data terms in Eq. (6) as a weighted ℓ_2 norm using a diagonal weight matrix **W** whose elements are all positive, we have

$$\mathbf{z}^{*} = \underset{\mathbf{z}}{\operatorname{argmin}} \left(\hat{\mathbf{z}} - \mathcal{P}_{\Omega} \mathbf{z} \right)^{\top} \mathbf{W}^{\top} \mathbf{W} \left(\hat{\mathbf{z}} - \mathcal{P}_{\Omega} \mathbf{z} \right) + \lambda \mathbf{z}^{\top} \mathbf{D}^{\top} \mathbf{D} \mathbf{z}.$$
 (7)

Note that the smoothness term is unchanged. Only the data term is rewritten from ℓ_2 to ℓ_1 . Our method iteratively updates the estimate of z and weight matrix W by alternating between two steps.

Step 1: Estimating z

We denote $\mathbf{A} = \begin{bmatrix} \mathbf{W} \mathcal{P}_{\Omega} \\ \sqrt{\lambda} \mathbf{D} \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} \mathbf{W} \hat{\mathbf{z}} \\ \mathbf{0}_{2I \times 1} \end{bmatrix}$, where $\mathbf{0}_{2I \times 1}$ is a zero array with 2I elements. Then, Eq. (7) is rewritten as $\mathbf{z}^* = \operatorname{argmin}_z || \mathbf{A}\mathbf{z} - \mathbf{b} ||_2^2$, which is a squared ℓ_2 sparse linear system that has a closed form solution. We add a small regularizer $\alpha \mathbf{I}$ for stabilizing the solution as

$$\mathbf{z}^* = [\mathbf{A}^\top \mathbf{A} + \alpha \mathbf{I}]^{-1} \mathbf{A}^\top \mathbf{b}, \tag{8}$$

where I is an identity matrix, α is a small positive scalar set to $\alpha = 1.0e - 8$.

Step 2: Updating W

The matrix \mathbf{W} is initialized to identity. For each iteration, \mathbf{W} is updated based on the residual $\mathbf{r} = \mathbf{W} \mathcal{P}_{\Omega} \mathbf{z}^* - \mathbf{W} \mathbf{b}$. The *i*-th diagonal element of \mathbf{W} is updated using the *i*-th element of the residual \mathbf{r} as

$$w_i = \frac{1}{|r_i| + \epsilon},\tag{9}$$

where $\epsilon = 1.0e-8$ is a small positive scalar used to avoid zero division. These steps are repeated until convergence; namely, until the estimate at *t*-th iteration $\mathbf{z}^{*(t)}$ does not vary much from the previous estimate $\mathbf{z}^{*(t-1)}$, *i.e.*, $\|\mathbf{z}^{*(t)} - \mathbf{z}^{*(t-1)}\|_2 < 1.0e - 8$. Figure 3.c is an example of the reconstructed mesh.

Ridge-aware reconstruction Developable surfaces are ruled [19], *i.e.*, containing straight lines on the surface as illustrated in Fig. 2. Our method exploits this geometric property to identify and add ridge constraint for more accurate surface reconstruction. Since extraction of ridges from images is difficult and so is from the sparse 3D points, we take a sequential approach by using the reconstructed mesh defined for z^* via the robust Poisson reconstruction described earlier, to determining the fold lines.

For each point z(x, y) on the mesh defined for z^* , we compute the Hessian K as

$$\mathbf{K}(z) = \begin{bmatrix} \frac{\partial^2 z}{\partial x^2} & \frac{\partial^2 z}{\partial x \partial y} \\ \frac{\partial^2 z}{\partial x \partial y} & \frac{\partial^2 z}{\partial y^2} \end{bmatrix}.$$
 (10)

By Eigen decomposition of the Hessian,

$$\mathbf{K}(z) = [\mathbf{p}_1, \mathbf{p}_2] \begin{bmatrix} \kappa_1 & 0\\ 0 & \kappa_2 \end{bmatrix} [\mathbf{p}_1, \mathbf{p}_2]^\top, \quad (11)$$

we obtain two principal curvatures $|\kappa_1|$ and $|\kappa_2|$ $(|\kappa_1| \le |\kappa_2|)$ and corresponding eigenvectors \mathbf{p}_1 and \mathbf{p}_2 .

For a developable surface, the smaller curvature $\kappa_1 = 0$ holds at any location. In other words, the surface contains a straight line along direction \mathbf{p}_1 at any point z_i . As we can see in Fig. 2, this property is most significant at ridges, where the curvature is zero along the ridge while it becomes significant in its tangent direction. Based on this observation, we determine ridge candidates using the greater principle curvature κ_2 . Specifically, for a mesh point $z_i(x_i, y_i)$, if $|\kappa_2(i)|$ is greater than the threshold κ_{th} , it is then regarded as a ridge candidate. Figure 3.d depicts an example of the ridge candidates.

Using the ridge candidates, we re-weight its smoothness constraints in Eq. (4) as:

$$\tilde{d}_{i,j} = \varphi(\langle \mathbf{p}_1, \mathbf{e}_1 \rangle) d_{i,j}
\tilde{d}_{i+I,j} = \varphi(\langle \mathbf{p}_1, \mathbf{e}_2 \rangle) d_{i+1,j},$$
(12)



Figure 4. Obtain coordinates of vertices on local basis determined by triangle.

where \langle , \rangle is the inner product, $\mathbf{e}_1 = [1, 0]^{\top}, \mathbf{e}_2 = [0, 1]^{\top}$ are orthonormal bases. $\varphi(\cdot)$ is a convex monotonic function defined as $\varphi(x) = \frac{\beta^{x^2} - 1}{\beta - 1}$, which gives a greater weight $\beta \gg 1$ along the ridge and smaller weight in the orthogonal direction. Similar as Eq. (12), for ridge candidates we also add directional smoothness constraints in slant direction $\mathbf{e}_3 = [\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}]^{\top}$ and $\mathbf{e}_4 = [\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}]^{\top}$. By updating the smoothness of Eq. (4) to incorporate the ridge-awareness, we again solve the robust Poisson reconstruction for obtaining the final reconstruction. Figure 3.e shows the improvement of surface reconstruction with the ridge-awareness.

3.2. Surface Unwrapping

Once we obtain the 3D surface reconstruction, our next step is to unwrap the surface. We take a conformal mapping approach to this problem. While the least-squares conformal mapping (LSCM) [11, 2] is a viable choice for us because of its efficiency, we make two extensions to address the issues of outliers and global distortion. Namely, we use a robust estimation scheme for solving the problem and incorporate the ridge-aware constraint again to this problem.

Conformal Mapping We first triangulate the mesh grid defined for \mathbf{z} using the following rule: For each point $z_i(x_i, y_i)$, two triangles are generated; one is with its upper neighbor and left neighbor, the other is with its lower neighbor and right neighbor. We denote the triangular mesh as $\{\mathcal{T}, \mathbf{z}\}$. The conformal mapping aims to find a corresponding mesh in 2D space, denoted as $\{\mathcal{T}, \mathbf{u}\}$, where $\mathbf{u} = (u_i, v_i), i = 1, \ldots, I$, with the same topology by best preserving the conformality of all the triangles. As illustrated in Fig. 4, for a given triangle with three vertices with the global coordinates $(x_1, y_1, z_1), (x_2, y_2, z_2),$ and (x_3, y_3, z_3) , we convert their coordinates to the local 2D coordinates $(X_1, Y_1), (X_2, Y_2),$ and (X_3, Y_3) . Then, the conformal tis formulated for its mapping

$$\mathbf{u}_{\mathcal{T}} = \begin{bmatrix} u_1, u_2, u_3, v_1, v_2, v_3 \end{bmatrix}^{\top} \text{ as}$$
$$\frac{1}{S} \begin{bmatrix} \Delta X_1 & \Delta X_2 & \Delta X_3 & -\Delta Y_1 & -\Delta Y_2 & -\Delta Y_3 \\ \Delta Y_1 & \Delta Y_2 & \Delta Y_3 & \Delta X_1 & \Delta X_2 & \Delta X_3 \end{bmatrix} \mathbf{u}_{\mathcal{T}} = \mathbf{0}, (13)$$

where $\Delta X_1 = (X_3 - X_2)$, $\Delta X_2 = (X_1 - X_3)$, and $\Delta X_3 = (X_2 - X_1)$, and ΔY are defined in a similar manner, and S is the area of triangle \mathcal{T} .

For all the triangles, the conformality constraints is formulated as:

$$\mathbf{C}\mathbf{u}=\mathbf{0},\tag{14}$$

where **0** is a zero array and **C** is a $2J \times 2I$ sparse matrix with the following non-zero elements

$$\mathbf{C} = \begin{bmatrix} c_{j,i} = \frac{\Delta X}{S_j}, & c_{j,i+I} = -\frac{\Delta Y}{S_j} \\ c_{j+J,i} = \frac{\Delta Y}{S_j}, & c_{j+J,i+I} = \frac{\Delta X}{S_j} \end{bmatrix},$$
(15)

where $1 \le j \le J$ is the index of triangles and $1 \le i \le I$ is the index of points.

Ridge constraints To avoid the global distortion in unwrapping, we use ridge and boundary constraints to regularize the solution. Since the conformal mapping preserves straight lines on ridges on a developable surface, these constraints serve as a non-local regularizer in the reconstruction. We formulate this collinearity in the same form as conformality constraints, so that it can be solved in the same framework. As illustrated in Fig. 4, when three points are collinear, (x_2, y_2, z_2) is also lying on the X axis; therefore, $Y_2 = 0$. In addition, the area of the triangle S is zero. Hence, the ridge constraints can be formed in a similar manner to Eq. (13) as

$$\begin{bmatrix} \Delta X_1 & \Delta X_2 & \Delta X_3 & 0 & 0 & 0\\ 0 & 0 & 0 & \Delta X_1 & \Delta X_2 & \Delta X_3 \end{bmatrix} \mathbf{u}_{\mathcal{R}} = \mathbf{0}.$$
(16)

To form the constraints, for each ridge, we select its two end points and one in the middle. Similarly, the boundary constraints are obtained in the form of Eq. (16), if we assume the flattened document has straight boundaries (not straight in the original 3D space).

All the ridge and boundary constraints can be represented as a linear system:

$$\mathbf{R}\mathbf{u} = \mathbf{0}.\tag{17}$$

Robust conformal mapping To robustify the conformal mapping, we employ an ℓ_1 objective function instead of the squared ℓ_2 norm. Putting together Eqs. (14) and (17) in the ℓ_1 sense, we have:

$$\mathbf{u}^* = \underset{\mathbf{u}}{\operatorname{argmin}} \parallel \mathbf{C}\mathbf{u} \parallel_1 + \gamma \parallel \mathbf{R}\mathbf{u} \parallel_1, \tag{18}$$



Figure 5. ℓ_1 conformal mapping with non-local constraints has better robustness than the original LSCM.

where γ controls the importance of ridge and boundary constraints. To avoid the trivial solution $\mathbf{u} = 0$, we fix two points of \mathbf{u} to $(u_i, v_i) = (0, 0)$ and $(u_j, v_j) = (0, 1)$. Equation (18) is then rewritten as

$$\mathbf{u}^* = \underset{\mathbf{u}}{\operatorname{argmin}} \parallel \mathbf{C}\mathbf{u} \parallel_1 + \gamma \parallel \mathbf{R}\mathbf{u} \parallel_1 + \theta \parallel E_{\operatorname{fix}} \parallel_2^2, \quad (19)$$

where $E_{\rm fix}$ is the energy function for the two fixed points. We solve the objective function using the iterative reweighted method [3]. Figure 5 shows one of the results with the conventional LSCM (ℓ_2 method) and our ℓ_1 solution method.

3.3. Implementation details

Structure from motion For the first step of 3D reconstruction, we obtain sparse 3D points using SfM. Since SfM does not need extra-equipment nor calibration, input data can be conveniently collected by a hand-held camera. Specifically, for each document, we record five to ten images from different viewpoints. When using a smart phone camera that supports burst shot or HD video recording, data can be acquired within a few seconds. While our method is not restricted to a particular SfM algorithm, in this work, we use the method developed by Snavely *et al.* [20, 8]. Figure 1.a is an example of the input data and recovered sparse 3D points.

Image warping After obtaining the flattened mesh grid $\mathbf{u} = \{u_i, v_i\}$ after conformal mapping, we warp the image. We begin with choosing one image with the largest content area for warping out from the input images. To establish the correspondence between the input image and $\{u_i, v_i\}$, we first project back the 3D mesh points $\{z_i(x_i, y_i)\}$ to the image coordinates $\{\tilde{x}_i, \tilde{y}_i\}$. This is done using the camera information obtained from SfM. Lastly we warp the image content according to the correspondence between $\{\tilde{x}_i, \tilde{y}_i\}$ and $\{u_i, v_i\}$. We use bilinear interpolation for points located between the reference mesh grid. Figure 1.e shows an example of the final rectification results.



Figure 6. Rectification results of the proposed method on various origami cases and contents.

4. Experiments

We evaluate the proposed technique qualitatively and quantitatively using a wide variety of input documents from different sources. The first set of experiments demonstrate that our method can handle different paper types, document contents and various types of foldings. Next, we report a quantitative evaluation of our rectification method using a global and a local distortion metrics. Finally, we demonstrate the superior performance and advantages of our method over existing approaches [2, 31].

4.1. Test Sequences

We test our method on typical foldings of a paper sheet. The input images as well as the results from our method are shown in Fig. 6. Specifically, the first six sequences (I – VI) contain documents with no fold lines, one fold line, two to three parallel fold lines, and two to three crossing fold lines respectively. The other six sequences (VII – XII) contain documents with an increasing number of fold lines and there irregular fold lines were intentionally added to make the rectification more challenging.

The documents photographed in sequence I - XII were either placed on a planar or curved background surface or held in hand (I). Sequence VII contains a shopping receipt



Figure 7. Comparison on global and local distortion metric using data with the ground truth in Fig. 6. RA stands for the proposed ridge-aware reconstruction while Po stands for Poisson reconstruction. L1 stands for the proposed ℓ_1 conformal mapping with non-local constraints; L2 stands for LSCM (Brown *et al.*) and Geo stands for geodesic unwrapping (Zhang *et al.*). Po + L2 generates results with obvious failure and is thus not compared.

on a paper roll whereas II and VIII contain pages from a book. Sequences III, IV, IX and X contain folded letters placed within envelopes. Sequence V, VI, XI and XII contain examples of documents kept in purse/notebooks or other small spaces.

Our method does not rely on the contents, formatting, layout, color; thus is generally applicable as long as the paper is textured which allows us to extract keypoints for SfM.

4.2. Quantitative Evaluation Metrics

We quantitatively evaluate the global and local distortion between the ground truth digital image and our rectified result using a global and local metric. The digital version of 6 out of the 12 test documents are available to us and we treat them as the ground truth. We normalize the height of ground truth images to 1000 pixels. Here we ignore the photometric distortion introduced by the printer or the shading caused by scene illumination.

Global distortion metric To evaluate the global distortion, we register the rectified image to the ground truth using a global affine transform. We find about 2K SIFT matches [14] between the rectified image (feature positions are de-



Figure 8. Visual comparison of rectification results obtained with six different methods. The abbreviated method names are consistent with Fig. 7 and the text. More results can be found in the supplementary material.

noted as $\mathbf{p} = (p_i, q_i, 1)$) and the ground truth (feature positions are denoted as $\hat{\mathbf{p}} = (\hat{p}_i, \hat{q}_i, 1)$). Then the global transform

$$\mathbf{T} = \begin{bmatrix} a_1 & a_2 & t_1 \\ a_3 & a_4 & t_2 \\ 0 & 0 & s \end{bmatrix},$$
 (20)

is estimated by minimizing the squared error:

$$\mathbf{T}^* = \underset{\mathbf{T}}{\operatorname{argmin}} \parallel \mathbf{T}\mathbf{p} - \hat{\mathbf{p}} \parallel_2^2.$$
 (21)

We define the global distortion metric \mathcal{G} as the normalized determinant of the affine part of T:

$$G = |a_1 a_4 - a_2 a_3| / s^2$$

$$\mathcal{G} = \max(G, 1/G).$$
(22)

The identity transformation will have $\mathcal{G} = 1$; and a larger number of $\mathcal{G} = 1$ indicates the higher degree of global distortion, *i.e.*, lower accuracy. The result on global distortion is summarized in Fig. 7.

Local distortion metric We also evaluate the local distortions in our results by computing a dense correspondence field using SIFT-flow [13] between the rectified image and the ground truth image. The frequency distribution of local displacements are shown in Fig. 7 and compared with existing methods. We found the SIFT-flow based registration



Figure 9. Comparison of the global distortion metric between our method (top) and Zhang *et al.* [31] and Brown *et al.* [2] with varying point density and noise. Note that lower values indicate higher accuracy. (b) Frequency distribution of local distortion metrics for the corresponding experiments. The local metric also demonstrates the superiority of our method with higher noise and sparser point sets.

to be more useful for our assessment compared to that obtained using sparse SIFT features since the sparse method tends to neglect matches with large deformation which can bias the evaluation.

4.3. Comparison with existing methods

Brown *et al.*'s [2] and Zhang *et al.*'s [31] methods, which are applicable to general curves and foldings are compared. Methods with an assumption of cylindrical surface do not work with general data and are thus not compared.

Real data Data with the ground truth (Fig. 6) are used for the evaluation. Both methods [2, 31] rely on the use of 3D range finder to obtain the dense geometry which are not available from our data. Thus we use our reconstructed surface as input for these methods and compare the performance the surface flattening quality. We also compare the behavior of our ridge-aware reconstruction to standard Poisson reconstruction. As illustrated in Fig. 7, the global and local distortion are used to evaluate the reconstruction quality. And a visual comparison is summarized in Fig. 8. The proposed method has better performance on both global and local distortion evaluations.

Synthetic data We generated synthetic dense 3D points to compare with [2, 31] since both those methods require dense points. To evaluate our method on such data, we generate point sets whose size varies from 2K (typical of SfM) to 300K (typical of range finder data) and we also inject varying degree of Gaussian noise. Figure 9 shows a comparison of our method with [2, 31] based on the local and global distortion metrics. These experiments demonstrate that with low noise and high point density, all three methods are comparable in accuracy. However, when the point

set is sparser or when the noise level is higher, the proposed method is more accurate than the existing methods [2, 31].

5. Conclusion and Future Work

In this paper, we propose a method for automatically rectifying curved or folded paper sheets from a small number of images captured from different viewpoints. We use SfM to obtain sparse 3D points from images and propose ridge-aware surface reconstruction method which utilizes the geometric property of developable surface for accurate and dense 3D reconstruction of paper sheets. We also robustify the reconstruction by using ℓ_1 optimization. After obtaining the surface geometry, we unwrap the surface by adopting conformal mapping with both local and non-local constraints in a robust estimation scheme. For the future work, we consider correct the shading of the document and further improve the geometric rectification.

References

- M. S. Brown and W. B. Seales. Image restoration of arbitrarily warped documents. *Pattern Analysis and Machine Intelligence* (*TPAMI*), *IEEE Transactions on*, 26(10):1295–1306, 2004. 1, 2
- [2] M. S. Brown, M. Sun, R. Yang, L. Yun, and W. B. Seales. Restoring 2d content from distorted documents. *Pattern Analysis and Machine Intelligence (TPAMI), IEEE Transactions on*, 29(11):1904– 1916, 2007. 2, 4, 6, 8
- [3] E. J. Candes, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted l₁ minimization. *Journal of Fourier analysis and applications*, 14(5-6):877–905, 2008. 3, 5
- [4] H. Cao, X. Ding, and C. Liu. A cylindrical surface model to rectify the bound document image. In *Computer Vision (ICCV). Ninth IEEE International Conference on*, pages 228–233. IEEE, 2003. 2

- [5] F. Courteille, A. Crouzil, J.-D. Durou, and P. Gurdjos. Shape from shading for the digitization of curved documents. *Machine Vision* and Applications, 18(5):301–316, 2007. 2
- [6] H. Ezaki, S. Uchida, A. Asano, and H. Sakoe. Dewarping of document image by global optimization. In *Document Analysis and Recognition (ICDAR). Eighth International Conference on*, pages 302–306. IEEE, 2005. 2
- [7] B. Fu, M. Wu, R. Li, W. Li, Z. Xu, and C. Yang. A model-based book dewarping method using text line detection. In *Proc. 2nd Int. Workshop on Camera Based Document Analysis and Recognition, Curitiba, Barazil*, pages 63–70, 2007. 2
- [8] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004. 5
- [9] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In Proceedings of the fourth Eurographics symposium on Geometry processing, 2006. 1, 3
- [10] H. I. Koo, J. Kim, and N. I. Cho. Composition of a dewarped and enhanced document image from two view images. *Image Processing* (*TIP*), *IEEE Transactions on*, 18(7):1551–1562, 2009. 1, 2
- [11] B. Lévy, S. Petitjean, N. Ray, and J. Maillot. Least squares conformal maps for automatic texture atlas generation. In ACM Transactions on Graphics (TOG), volume 21, pages 362–371. ACM, 2002. 1, 2, 4
- [12] J. Liang, D. DeMenthon, and D. Doermann. Geometric rectification of camera-captured document images. *Pattern Analysis and Machine Intelligence (TPAMI), IEEE Transactions on*, 30(4):591–605, 2008.
- [13] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. T. Freeman. Sift flow: Dense correspondence across different scenes. In *Tenth European Conference on Computer Vision (ECCV)*, pages 28–42. Springer, 2008. 7
- [14] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 60(2):91– 110, 2004. 7
- [15] S. Lu, B. M. Chen, and C. C. Ko. A partition approach for the restoration of camera images of planar and curled document. *Image and Vision Computing*, 24(8):837–848, 2006. 2
- [16] G. Meng, C. Pan, S. Xiang, J. Duan, and N. Zheng. Metric rectification of curved document images. *Pattern Analysis and Machine Intelligence (TPAMI), IEEE Transactions on*, 34(4):707–722, 2012.
- [17] G. Meng, Y. Wang, S. Qu, S. Xiang, and C. Pan. Active flattening of curved document images via two structured beams. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pages 3890–3897. IEEE, 2014. 1, 2
- [18] M. Pilu. Undoing paper curl distortion using applicable surfaces. In Computer Vision and Pattern Recognition (CVPR), IEEE Conference on, volume 1, pages I–67. IEEE, 2001. 1, 2
- [19] E. Portnoy. Developable surfaces in hyperbolic space. Pacific Journal of Mathematics, 57(1):281–288, 1975. 4
- [20] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: exploring photo collections in 3d. ACM transactions on graphics (TOG), 25(3):835–846, 2006. 5
- [21] N. Stamatopoulos, B. Gatos, I. Pratikakis, and S. J. Perantonis. Goaloriented rectification of camera-based document images. *Image Processing (TIP), IEEE Transactions on*, 20(4):910–920, 2011. 1, 2
- [22] C. L. Tan, L. Zhang, Z. Zhang, and T. Xia. Restoring warped document images through 3d shape modeling. *Pattern Analysis and Machine Intelligence (TPAMI), IEEE Transactions on*, 28(2):195–208, 2006. 2
- [23] Y. Tian and S. G. Narasimhan. Rectification and 3d reconstruction of curved document images. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pages 377–384. IEEE, 2011.
- [24] R. Tibshirani. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), pages 267–288, 1996. 3

- [25] Y.-C. Tsoi and M. S. Brown. Geometric and shading correction for images of printed materials: a unified approach using boundary. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, volume 1, pages 240–246. IEEE Computer Society, 2004. 2
- [26] Y.-C. Tsoi and M. S. Brown. Multi-view document rectification using boundary. In *Computer Vision and Pattern Recognition (CVPR)*, *IEEE Conference on*, pages 1–8. IEEE, 2007. 1, 2
- [27] A. Ulges, C. H. Lampert, and T. M. Breuel. Document image dewarping using robust estimation of curled text lines. In *Document Analysis and Recognition (ICDAR). Proceedings. Eighth International Conference on*, pages 1001–1005. IEEE, 2005. 2
- [28] T. Wada, H. Ukida, and T. Matsuyama. Shape from shading with interreflections under a proximal light source: Distortion-free copying of an unfolded book. *International Journal of Computer Vision* (*IJCV*), 24(2):125–135, 1997. 1, 2
- [29] A. Yamashita, A. Kawarago, T. Kaneko, and K. T. Miura. Shape reconstruction and image restoration for non-flat surfaces of documents with a stereo vision system. In *Pattern Recognition. ICPR* 2004. 17th International Conference on, volume 1, pages 482–485. IEEE, 2004. 2
- [30] L. Zhang, A. M. Yip, M. S. Brown, and C. L. Tan. A unified framework for document restoration using inpainting and shape-fromshading. *Pattern Recognition*, 42(11):2961–2978, 2009. 1, 2
- [31] L. Zhang, Y. Zhang, and C. L. Tan. An improved physically-based method for geometric restoration of distorted document images. *Pattern Analysis and Machine Intelligence (TPAMI), IEEE Transactions* on, 30(4):728–734, 2008. 2, 6, 8
- [32] Z. Zhang, X. Liang, and Y. Ma. Unwrapping low-rank textures on generalized cylindrical surfaces. In *Computer Vision (ICCV), IEEE International Conference on*, pages 1347–1354. IEEE, 2011. 1, 2
- [33] Z. Zhang, C. Lim, and L. Fan. Estimation of 3d shape of warped document surface for image restoration. In *Pattern Recognition* (*ICPR*). Proceedings of the 17th International Conference on, volume 1, pages 486–489. IEEE, 2004. 2