

Robust Document Warping with Interpolated Vector Fields

David C. Schneider
dc.schneid@googlemail.com

Marco Block
block@inf.fu-berlin.de

Raúl Rojas
rojas@inf.fu-berlin.de

Institute of Computer Science, Freie Universität Berlin
Takustraße 9, 14195 Berlin, Germany

Abstract

This paper describes a new versatile algorithm for correcting nonlinear distortions, such as curvature of book pages, in camera based document processing. We introduce the idea of using local orientation features to interpolate a vector field from which a warping mesh is derived. Ultimately, the image is corrected by approximating the nonlinear distortion with multiple linear projections. Since the algorithm does not derive the mesh directly from text baselines it is robust over arbitrarily complex text layouts. We describe a baseline detector for extracting the required local orientation features. We also sketch a method for correcting nonlinear distortions of a document's vertical axis with our algorithm.

1 Introduction

The “warping” of document images with the aim of obtaining text lines rectilinear in the image coordinate system is a crucial preprocessing step in camera based document processing. A rectified document image is a precondition for most optical character recognition systems; it is desirable for electronic archiving and reprography.

In this paper we present a new method for modelling and correcting distortion. It was developed with a focus on curvature problems (e.g. images of open books) but it is capable of correcting other distortions as well. Our approach makes fewer assumptions about the layout of the document to be corrected than other methods reported in the literature and is thus capable of correcting a wider range of documents.

The paper is structured as follows. First we briefly review related work. Then we introduce the idea of using vector field interpolation for warping mesh construction. Next, we describe the baseline detector we use to obtain source data for a vector field from a document image. Finally, we give examples and report on the algorithm's capability to

improve OCR results.

2 Related work

The approaches to the curvature correction problem described in the literature can be divided into two groups: Those requiring additional hardware and software-only methods.

In the first group, typically a geometric shape with known properties is projected onto the document, the projection is recognized in the digital image and a distortion model is derived from the differences between projection and ideal shape. For example, Pilu [6] projects a bright square on the document, photographs it at several locations and computes the model from the distortion of the square. The method yields precise models and is capable of correcting not only curved but also crumpled documents. Chu et al. [3] use a rotating laser beam to project lines on the document and compute the model from the curvature of the lines.

On the software side, Zhang et al. [10] derive their method from an accurate investigation in the image formation process of a flat bed scanner, limiting, however, the algorithm to scanned images. Several approaches, including ours, recognize that most printed text documents already have intrinsic properties that are known a priori so that a projection is not necessary to compute a deviance: For the vast majority of documents the lines of text can be assumed to be parallel to the document's X axis. The algorithms thus aim at recognizing the text baselines, i.e. the imaginary lines most characters “rest on”. From these a distortion model is derived which is ultimately used to correct the image. The approaches differ in their strategy for baseline recognition, in their choice of the distortion model and the methods used for computing the correction. Cao et al. [2] parametrize a cylindrical model; consequently the range of distortions they can correct is limited. Wu et al. [8] as well as Zhang and Tan [10] describe the baselines with splines and abandon the idea of modelling the

three-dimensional page and flattening the model. Instead, they use the splines to construct two-dimensional meshes and employ image warping techniques for rectification.

3 A new method for mesh construction

Our approach follows [10] and [8] in so far as we work in 2D space and use deformation meshes for rectification. The common strategy can be described as follows. Calling the image coordinate system the XY system and the document coordinate system the UV system (where U is the orientation of the text), two conditions must hold to rectify an image: (1) The mesh must be fit to the distorted document in the image so that the mesh coordinate system approximates the UV system. (2) A mapping $\mathcal{M} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ from the mesh system into the XY system must be known. Then the image can be rectified by resampling: Denoting by \mathbf{I} the original image, by \mathbf{J} the rectified image and by $\mathbf{I}_{x,y}$ the pixel at coordinates (x, y) in \mathbf{I} , we have $\mathbf{J}_{u,v} = \mathbf{I}_{\mathcal{M}(u,v)}$ for all pixels. An interpolation scheme should be used to deal with non-integer coordinates.

We developed a new method for mesh construction for the following reason: The approaches reported in the literature typically use the splines that model the text baselines directly as basis for their meshes (e.g. Gordon surfaces in [10]). This works well for single text blocks but has considerable disadvantages in complex document layouts: The baselines are neither guaranteed to stretch over the entire page (e.g. in multi column, left or center aligned documents) nor do they necessarily follow a common vertical raster. Thus there is no guarantee that a sufficient number of suitable lines can be found. Moreover, if unsuitable lines are discarded for mesh construction, potentially valuable information about the distortion is wasted.

3.1 Using vector fields as curvature models

In our approach we derive the mesh from an intermediate representation of curvature information: We build up a vector field $\mathcal{V} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ mapping each point P in the XY system to a vector $\mathcal{V}(P)$ that points in the tangent direction of a projection of a U-parallel line into the XY-system.¹ An obvious example for such a line is a text baseline. Thus our method can be used with any baseline detection algorithm (see section 3.3 for a description of ours). In the remainder of this section we will assume that we have a set of splines describing detected baselines. Note, however, that it is not necessary to model full baselines to employ our dewarping method; this point will be elaborated below.

¹Note that this projection is a curve in the XY system in the case of a nonlinear distortion.

Let S be a baseline spline and $S'(p)$ its first derivative at parameter value p , i.e. a tangent vector of S at the point $S(p)$. Clearly, the value of \mathcal{V} is known at each point on the spline:

$$\mathcal{V}(S(p)) = \frac{S'(p)}{\|S'(p)\|} \quad (1)$$

for all valid parameter values p ; $\|\cdot\|$ denoting the norm of a vector. For evaluation of \mathcal{V} at points that do not lie on a baseline, note that equation 1 maps points to unit vectors. Thus an alternative representation of the field is the following map $\mathcal{V}^* : \mathbb{R}^2 \rightarrow [0..2\pi]$ of points to scalars:

$$\mathcal{V}^*(S(p)) = \arccos\left(\frac{S'(p) \circ (1 \ 0)}{\|S'(p)\|}\right) \quad (2)$$

where \circ is the inner product. Clearly, the two representations are related by

$$\mathcal{V}(S(p)) = \begin{pmatrix} \cos[\mathcal{V}^*(S(p))] \\ -\sin[\mathcal{V}^*(S(p))] \end{pmatrix}. \quad (3)$$

This allows to treat the problem of evaluating \mathcal{V}^* for a point that does not lie on a baseline as a three dimensional (i.e. “surface”) interpolation problem with scattered data. The known “datapoints” are angles of the baseline spline tangents at random locations distributed over the image. The values to be interpolated are unknown tangent angles at arbitrary locations. Since only local orientation features are required to interpolate the vector field, extraction of full baselines is not necessary for the method. We are currently experimenting with purely local orientation detectors.

There are several classic solutions for the interpolation problem; for a survey see [1]. We achieved convincing results using a Delaunay triangulation of the known datapoints and barycentric interpolation of the unknown. The triangulation can also be used for eliminating defective datapoints: Let (x, y, φ) be a known data point in the model and let P be the set of all datapoints (x, y, φ) is connected to in the triangulation. Then a heterogeneity measure $h_{xy\varphi}$ can be defined for each point:

$$h_{xy\varphi} = \frac{1}{|P|} \cdot \sum_{(x', y', \varphi') \in P} \frac{(\varphi - \varphi')^2}{(x - x')^2 + (y - y')^2} \quad (4)$$

A point is removed from the model if its heterogeneity exceeds an experimentally established threshold.

The warping mesh we use is a grid of quadrangles, approximating thereby the nonlinear distortion by a large number of tiled linear ones. Once the datapoints have been obtained the mesh can be derived by following flow lines of the vector field using the interpolation scheme described above. Formally, a flow line of a two-dimensional vector field \mathcal{V} is a parametric curve $\sigma : \mathbb{R} \rightarrow \mathbb{R}^2$ satisfying

$$\sigma'(t) = \mathcal{V}(\sigma(t)). \quad (5)$$

If the analytic expression of the field is unknown a flow line can be approximated numerically by the recursion

$$P_i = P_{i-1} + \delta \mathcal{V}(P_{i-1}) \quad i > 0 \quad (6)$$

where P_i is the i th point on the approximated flow line and δ is the step length, defining the precision of the approximation. To generate flow lines, equation 6 is evaluated from a starting point P_0 until some terminal condition is met. The path length $\Delta_{i,j}$ of the flow line between points P_i and P_j is given by

$$\Delta_{i,j} = \delta|i - j|. \quad (7)$$

Hence, to obtain a mesh with $m \times n$ points, n flow lines f_1, \dots, f_n are generated with a reasonably small δ ; the choice of starting and endpoints is addressed below. For each line f_i , m points $P_{1,i}, \dots, P_{m,i}$ equidistantly distributed along the line are stored.

To compute the warping, each of $m \times n$ quadrangular areas in the original image with corner points

$$\langle P_{i,j}, P_{i+1,j}, P_{i+1,j+1}, P_{i,j+1} \rangle$$

is projected to a rectangle in the corrected image with corner points

$$\langle (ci, dj), ((c+1)i, dj), ((c+1)i, [d+1]j), (ci, [d+1]j) \rangle$$

where c and d are scaling factors. See figure 1 for an illustration of a warping mesh. The projection is a homography whose matrix can be computed from the four pairs of corresponding points with standard methods; see e.g. [4, 5]. Figure 2 shows the corrected image corresponding to the mesh in figure 1. Figure 5 shows another warping example.

3.2 Choice of starting- and endpoints

The choice of starting and endpoints of the flow lines determines whether perspective distortion or only curvature is corrected by the algorithm. Note that plain curvature correction is sufficient for some applications. We found, for example, that OCR results are not significantly affected by perspective distortions as long as the lines are parallel to the X axis of the image.² If plain curvature detection is satisfactory, the flow lines can be started and ended at the image boundaries.

Otherwise we model the left bound of the leftmost text block and the right bound of the rightmost text block as straight lines; flow line evaluation is started on the former and ended when the flow line crosses the latter. For details of the bound modeling algorithm, see [7]. The choice of straight lines as bounds entails the assumption that the distortion of the V axis is approximately linear. With their ruled surface model, the authors of [9] show that this assumption is reasonable when correcting curved book pages.

²The claim holds for the OCR package “Abbyy Finereader Professional 8” which we used for evaluation.

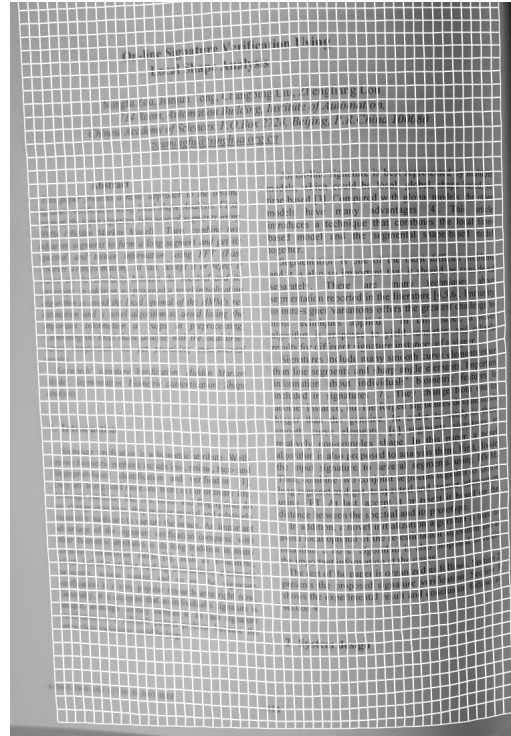


Figure 1. A warping mesh for a multi-column layout obtained by the vector field method.

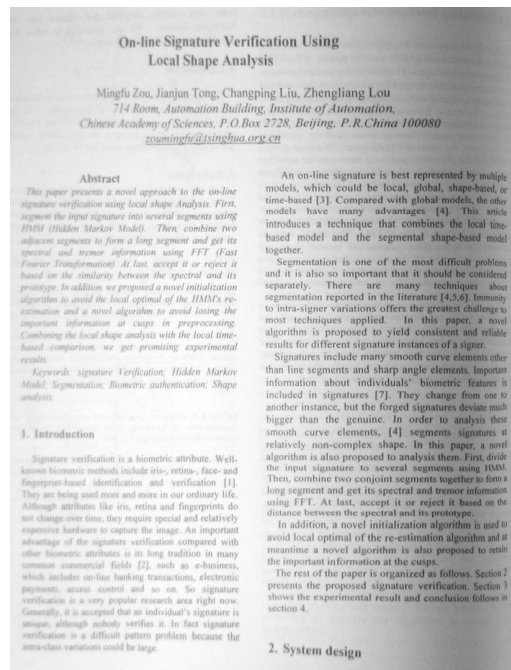


Figure 2. Document corrected with the mesh in fig. 1.

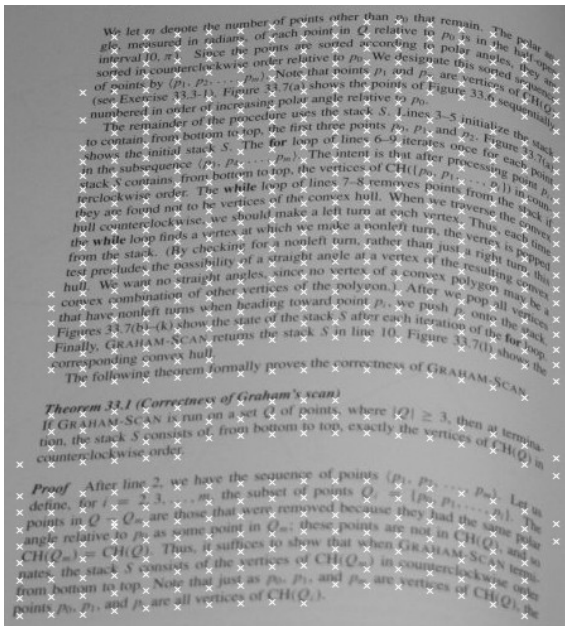


Figure 3. Candidate points found by the baseline detector.

3.3 Baseline detection

As described in section 3.1, an explicit model of text baselines is an obvious source of the curvature information required for the algorithm. Our baseline detector is based on the fact that along the baselines there are sharp transitions from dark (ink) pixels to light (paper) pixels orthogonal to the baseline direction. Assuming that the depicted document is well-oriented in the image (i.e. that text lines roughly go from the left to the right), the Y component of the image gradient is a suitable measure for these transitions. Note that our method does not require a binarized image.

The algorithm splits up the Y-gradient image as obtained by a Sobel filter into vertical stripes. For each stripe, feature points likely to lie on a baseline are obtained by the vertical projection profile method. From a set of five projection angles ranging from -40 to 40 degrees the one with the largest profile variance is chosen; feature points are generated from the profile's local maxima. See figure 3 for an example. The feature points of incident stripes are connected. Connected points serve as control points for cubic spline models of the baselines.

Note that deciding which points to connect is not trivial: There may be erroneous feature points, the baselines themselves may have different lengths and the splines have to follow an unknown curvature. To cope with these issues, the algorithm first computes a prediction of line progression

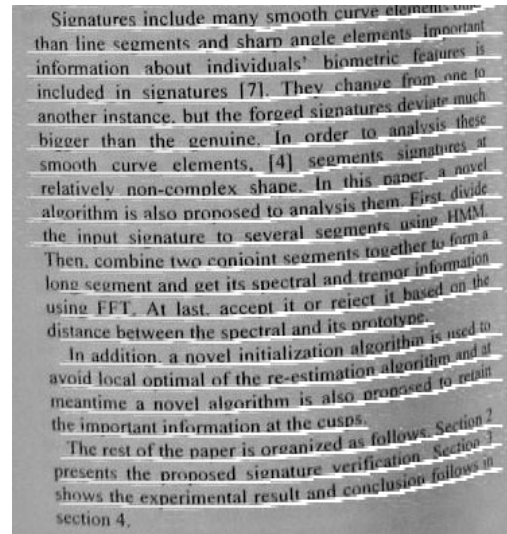


Figure 4. Baselines obtained by connecting candidate points.

for each line based on the part of the line that has already been modelled. For each line the ideal next point is determined based on this prediction. Lines are expanded by iteratively connecting them to their ideal next point, starting with the lines with the lowest prediction error. Thus, if two lines compete for the same point, the line with the smaller prediction error will be expanded and the other line, lacking a new point, ends. See figure 4 for a line connection result.

The result of baseline extraction is thus a set of cubic splines from which the data points for the vector field can easily be obtained by evaluating the first derivative of each spline at the desired locations.

3.4 Modelling the V-distortion

Since baselines are only one possible source of orientation information, nonlinear distortions of the vertical axis can also be corrected with our warping method. All we need are local orientation features describing the distortion of the V-axis. Note that the following is work in progress.

Our current method for finding these features presupposes straight text baselines, so the distortion of the horizontal axis has to be corrected with the method described above in a first pass. The image is then binarized and the individual lines are separated by the vertical projection profile method. On each line, we use the direction of the vertical strokes of characters like "f", "l", "h", "H", "M" as orientation features. They are obtained by radial sampling. Italic type, however, as well as some characters (e.g. "W", "Z") yield outliers.

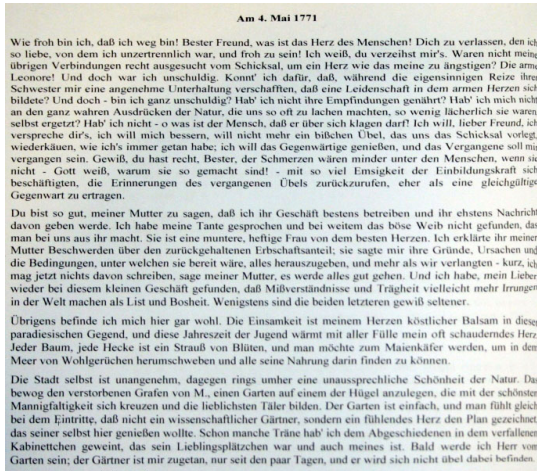
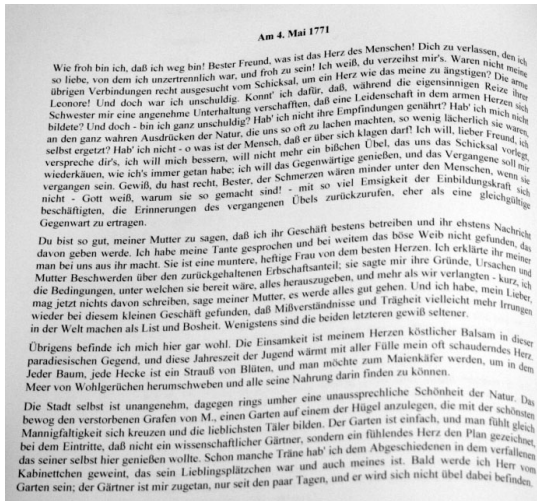


Figure 5. Top: Document from the OCR test set, recognition rate 68%. Bottom: Corrected document, OCR rate 90%.

4 Evaluation and conclusion

We described a new algorithm for correcting nonlinear distortions in camera based document processing. We introduced the idea of using local orientation features to interpolate a vector field from which a warping mesh is derived. Since the algorithm does not derive the mesh directly from text baselines it is robust over complex layouts. We also described an algorithm for baseline detection and sketched a method for correcting nonlinear distortions of a document's vertical axis with our approach.

Our algorithm achieves increases in OCR recognition rate of up to 25 percent on strongly curved documents. The software we used for evaluation was "Abby Finereader Professional 8". Our method outperformed the program's

internal curvature correction engine by up to 22 percent.

Concerning future research, we concentrate on reducing the number of defective orientation features which affect the warping quality. We investigate ways of avoiding erroneous features in our detectors as well as methods for identifying and eliminating them post-hoc. We are experimenting with local detectors for horizontal and vertical orientation in order to avoid baseline extraction in the future. Moreover, we work on improving the handling of unusual content like formulas, and we are developing methods for detecting and handling non-text elements such as images and illustrations.

References

- [1] I. Amidror. Scattered data interpolation methods for electronic imaging systems: A survey. *Journal of Electronic Imaging*, 11 (2):157–176, 2002.
- [2] H. Cao, X. Ding, and C. Liu. Rectifying the bound document image captured by the camera: A model based approach. In *Proceedings of the Seventh International Conference on Document Analysis and Recognition*, 2003.
- [3] K. B. Chu, L. Zhang, Y. Zhang, and C. L. Tan. A fast and stable approach for restoration of warped document images. In *Proceedings of the Eighth International Conference on Document Analysis and Recognition*, 2005.
- [4] D. Forsyth and J. Ponce. *Computer Vision – A Modern Approach*. Prentice Hall, 2003.
- [5] M. Harker and P. O’Leary. Computation of homographies. In *British Machine Vision Conference, Electronic Proceedings*, 2005.
- [6] M. Pilu. Undoing page curl distortion using applicable surfaces. In *Proceedings of the IEEE International Conference on Image Processing*, 2001.
- [7] D. C. Schneider. Correcting curvature and perspective for camera based document processing. Master thesis, Freie Universitaet Berlin, 2006.
- [8] C. Wu and G. Agam. Document image de-warping for text/graphics recognition. In *Proceedings Of Joint IAPR 2002 and SPR 2002*, 2002.
- [9] L. Zhang and C. L. Tan. Warped image restoration with applications to digital libraries. In *Proceeding of the Eighth International Conference on Document Analysis and Recognition*. IEEE, 2005.
- [10] Z. Zhang, C. L. Tan, and L. Fan. Estimation of 3d shape of warped document surface for image restoration. In *Proceedings of the International Conference on Pattern Recognition*, 2004.