

# A New Physically Motivated Warping Model for Form Drop-Out

Guy Rosman Asaf Tzadok Doron Tal  
IBM Corporation, Haifa Research Center  
rosman,asaf,doront@il.ibm.com

## Abstract

*Documents scanned by sheet-fed scanners often exhibit distortions due to the feeding and scanning mechanism. This paper presents a new model, motivated by the distortions observed in such documents. Numerical problems affecting the use of this model are addressed using an approximated model which is easier to estimate correctly. We demonstrate results showing the robustness and accuracy of this model on sheet-fed scanners output, and relate to existing techniques for registration and drop-out of structured forms.*

## 1. Introduction

A significant portion of the business data processed today comes from forms scanned by sheet-fed scanners. Paper in these scanners is fed into a line scanner using two feed-rollers. The rate of advance of the paper at each point is determined by the friction of the paper and the rollers. This force may change in various parts of the paper. Moreover, the resulting uneven velocity of the paper at both its sides may add a rotation component to the motion of the paper.

The resulting distortions make form drop-out [22] and analysis more difficult. These distortions can be handled by registering the scanned form to a document template, and de-warping the document image. Image warping and registration is a well developed field of image processing and computer graphics [2, 4, 16]. It is related to surface reconstruction techniques in computer vision [4] and interpolation techniques in computer graphics [17]. Specifically, we look at the following problem, similar to the problem defined in [17]:

**Regularized Image Deformation Problem:** Given a set of  $n$  point pairs,  $(\mathbf{p}_i, \mathbf{q}_i)$ ,  $\mathbf{p}_i \in \mathbb{R}^2$ ,  $\mathbf{q}_i \in \mathbb{R}^2$ ,  $i = 1..n$ , output a continuous function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  such that  $\forall i = 1..n$ ,  $\{f(\mathbf{p}_i)\}$  best approximate  $\{\mathbf{q}_i\}$  in some sense, and that  $f$  is well behaved in some sense.

The exact sense in which  $f$  approximates the given cor-

respondences, and the meaning of well behavior, determine the nature of the optimal solution for the problem. Note that even if we demand that  $f$  exactly matches the given point pairs, the problem is still ill-posed. For a general survey of registration methods, refer to [23]. Often,  $\mathbf{p}_i$  are points taken from an ideal template representing the document.

In several applicative domains, one can limit the behavior of the matching function  $f$  according to physical principles. These domains include medical imaging [20], flow dynamics [16], and multiple view geometry [8]. In document processing, several physically motivated methods for image de-warping have been developed [10], but most relevant works do not address document feeder scanning.

A large body of works discusses skew detection and correction (see for example [13, 18]). Most of these methods, however, assume the existence of text lines or similar visual cues. In addition, their supported model of transformation is limited, and is mostly useful for preprocessing text images before recognition. It is not adequate for full registration aimed at further processing of structured documents. Often, these methods do not address the global distortions caused by sheet-fed scanners, but rather compensate for small vibrations [21], or deal with the case of scanning thick bound volumes [24]. Furthermore, many of these assume the scanned documents consist of large textual areas [18, 19]. This is due to the assumption that the document has distinct statistical properties which favor the horizontal direction. These can be horizontally aligned interest points, or long lines of text affecting the distribution of the pixels. In general, these assumptions may be incorrect.

In this paper we present a new model for registering a scanned document unto a document template, for use in form drop-out [22] and structured documents layout understanding. In Section 2 we discuss the problem of registration and warping, and detail several algorithms that can be used to document image registration. Section 3 develops a new model approximating the distortions introduced by the feeding mechanism. Section 4 describes one specific way in which the ideas and algorithms can be implemented. Section 5 details the resulting benefits of using these methods in

an actual form processing application. Section 6 concludes the paper and points at possible future work.

## 2. Registration and Warping

As mentioned earlier, the general problem of image de-warping is ill-posed. We experimented with several definitions for the penalty imposed on the function due to the matching error and the regularization required of the solution.

After the registration phase, a de-warping phase is taking place. The alignment of a sensed image and a reference image is common in many fields (such as satellite imagery understanding [2], medical imaging [11] and particle image velocimetry [16]). In our target application of form drop-out, the removal of template data from the scanned form is the motivation. The type of warping function should be chosen according to several requirements: computational complexity, density and regularity of the correspondences, accuracy of the correspondences, and *a priori* knowledge about the acquisition process and expected image degradations (e.g., scanner dependent distortion).

Given a match between two sets of image features, methods warping one image into another can be divided into two categories. Global methods use all of the correspondences to estimate the warping function at each location. Local methods treat the image as a composition of patches, resulting in a warping that depends only on local correspondences. Another possible classification is according to the warping of the corresponding points themselves. Some of the warping methods guarantee that the known points are wrapped exactly (i.e.,  $f(\mathbf{p}_i) = \mathbf{q}_i$ ), whilst approximation methods try to find the best trade-off between the accuracy of this matching and other requirements made of the warping function.

### 2.1. Triangulation Based Warping

Triangulation [15] is often used in computer graphics to approximate a surface based on sampled points. The resulting function matches the given points exactly, and it is continuous, but it is not continuously differentiable. This method suffers from two major drawbacks. The first is fold-over of the image. Another drawback is the lack of a natural extension of the warping outside the correspondences' convex hull. We address these two issues in Section 4. On the other hand, due to the local and piecewise linear nature of the transformation, the warping can be done very fast.

Triangulation in itself is merely a member of the larger family of local deformation models. Other local methods can be applied to the problem of image deformations. These include spline-based deformations [12], local quadratic de-

formations (such as [7]), RBF-based deformations [17] and others.

### 2.2. Polynomial-Based Warping

A serious limitation of warping using triangulation and other local methods is the small support set for each local model. This limitation tends to make the algorithm vulnerable to outliers. In addition, the registration may be noisy and global fitting of the points may improve accuracy. We therefore turn to methods that use the entire data set to produce a model of the deformation.

One of these methods is a regularized version of the polynomial interpolation algorithm. Polynomial interpolation has long been used for image registration [2]. We approximate by a polynomial the displacement functions between each of the coordinates of  $\mathbf{p}_i$  and  $\mathbf{q}_i$ , in both the  $x$  and  $y$  axes.

The basic method of polynomial approximation consists of solving for the polynomial coefficients  $\mathbf{c}$

$$\mathbf{V}\mathbf{c} = \mathbf{f}$$

in the least-square sense, where  $\mathbf{f}$  is the vector of displacement values between registered point pairs along one of the coordinate axes, and  $\mathbf{V}$  is the Vandermonde matrix with elements  $(\mathbf{V})_{i,j+Nk} = x_i^j y_i^k$ , where  $x_i, y_i$  denote the  $x$  and  $y$  coordinates of point  $\mathbf{p}_i$ , respectively, and  $N$  is the degree of the polynomial.

In practice, we limit ourselves to polynomials of a low degree. This makes the effect of the Runge phenomenon [3] less significant. In addition we add regularization to the interpolation problem: Over a set of points  $(x_j, y_j), j = 1, 2, \dots, M$  we add a regularization term for the derivatives of the approximated function, in the least-squares sense. Assuming the set  $(x_j, y_j)$  is dense enough in the image area, we obtain a regularization approximating a Tikhonov regularization.

The regularization is performed by computing the expressions for the second derivatives at the set of points  $(x_j, y_j)$  in terms of the elements of  $\mathbf{c}$ , adding them as additional rows to the Vandermonde matrix  $\mathbf{V}$  to form an extended Vandermonde matrix. Since the set of regularization equations is a homogeneous system, it can be multiplied by a positive scalar to establish the weighting of the regularization term in the functional.

This regularization can be related to the thin-plate model of the surface reconstruction problem in computer vision [6]. A slightly different approach of a Tikhonov regularization has been attempted, under different conditions, in [5]. Other possible extensions to this idea can be done, including regularization based on the high pass response energy, matching points to lines and reweighting the various constraints.

### 3. A Physically Motivated Model For Image Warping

We now describe a transformation model approximating the deformations caused by the feeding and scanning process. We assume that the template document is available in an undistorted manner. This can be done by careful scanning of the template document using a flatbed scanner or by processing the template document before using its image.

We assume that over the scanning of the deformed region, the distance along the paper between both tractors remains the same. This stems from the assumption that the paper is inserted while stressed between the two tractors, and remains so without slipping in a direction orthogonal to the scanning direction.

Based on these assumptions, it is natural to consider a model similar to the transformation from polar to Cartesian coordinates around some center point. Usually the center point is far from the document area (i.e., the document is only slightly distorted because of the large radius)

$$\begin{pmatrix} \tilde{x}(r, \theta) \\ \tilde{y}(r, \theta) \end{pmatrix} = \begin{pmatrix} x_c \\ y_c \end{pmatrix} + \alpha_r(r - r_c) \begin{pmatrix} \cos(\alpha_\theta(\theta - \theta_c)) \\ \sin(\alpha_\theta(\theta - \theta_c)) \end{pmatrix}$$

The parameters of this model are the center of the polar transformation  $x_c$ , the scales  $\alpha_r, \alpha_\theta$  of the polar coordinates, and the displacement of the polar coordinates  $r_c, \theta_c$ . An example of such a transformation was shown in the context of computer graphics and solid deformations [1].

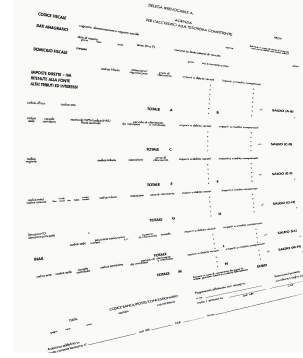
In the asymptotic case where  $r_c \rightarrow \infty$ ,  $\alpha_\theta = \frac{-1}{r_c}$ ,  $\alpha_r = 1$ ,  $x_c = r_c$ ,  $y_c = 0$ ,  $\theta_c = 0$  we obtain

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} = \begin{pmatrix} x_c \\ y_c \end{pmatrix} + \alpha_r(r - r_c) \begin{pmatrix} 1 \\ \alpha_\theta(\theta - \theta_c) \end{pmatrix} = \begin{pmatrix} r \\ \theta \end{pmatrix} \quad (1)$$

Therefore, as private cases of this model, we have the identity transformation, and in fact, all Euclidean transformations. The results of the deformations created by this model, exaggerated to be visible, are shown in Figure 1. Given this transformation model, we can try to estimate its parameters in the least squares sense. The search for the optimal parameters of this deformation model, however is not so simple. Since typical documents have a small distortion, we require a large radius  $r_c$ . For an image without any warping, we have  $r_c \rightarrow \pm\infty$ , which is ambiguous. In addition, the problem is ill-conditioned by nature, as can be seen by the condition number of the Hessian matrix of the error function, which reaches  $10^6$ .

Looking for a model that converges faster, we use the following approximation of the polar model:

$$T_P \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{R}_\theta \mathbf{S} \mathbf{P}_{\alpha, \beta} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \Delta x_c \\ \Delta y_c \end{pmatrix} \quad (2)$$



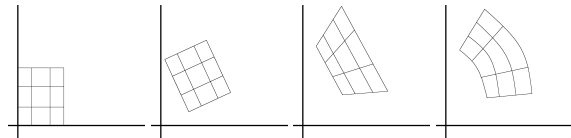
**Figure 1. An exaggerated example of a polar distortion**

where  $\mathbf{R}_\theta$  is the rotation matrix at an angle of  $\theta$ ,  $\mathbf{S}$  is a matrix scaling both coordinates separately, and  $\mathbf{P}_{\alpha, \beta}$  operates on  $(x, y)$  such that

$$\mathbf{P}_{\alpha, \beta} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x(1 + \alpha(y - K)^2) \\ y(1 + \beta x) \end{pmatrix} \quad (3)$$

where  $K$  is the center of the parabolic distortion and  $\alpha, \beta$  are scale factors for the distortions introduced.

Thus, this model consists of a Euclidean transformation and scaling, but also allows a trapezoid deformation of the original  $y$  axis, and a parabolic deformation of the original  $x$  axis. The steps involved in transforming the template are illustrated in Figure 2.



**Figure 2. From left to right: the truncated polar model can be view as a shift, a scale and a rotation operations, along a trapezoid and parabolic distortion**

To see the relation between this model and the polar model of Equation 1, we look at the case of a small distortion due to a large radius  $r_c$  and small angular proportional factor  $\alpha_r = \frac{1}{r_c}$ . Using the Taylor series expansion for the trigonometric functions, we obtain a model which is, up to redefinition of constants, the same as the model shown in Equation 2. This model resembles the one given by Kanungo et. al. [9], but also handles some of the distortions introduced by the feeding mechanism of the scanner.

Using this new model, we get a faster rate of convergence. Using a reasonable initial solution, we need only a few conjugate gradients [14] iterations to converge.

Next, we note that in practice, the speed with which the paper advances with respect to the feeder can change during the scan, due to paper elasticity, folds and wrinkles that affect friction, and stains. It is only natural to extend the model into a compound model, consisting of a handful of regions, where each region is modeled separately according to Equation 2. We refer to this model as a piecewise polar model. Even when using an arbitrary partition of the page, this model achieves a low error similar to the one exhibited by the triangulation model, while still using several points to approximate the transformation at each region.

#### 4. Implementation Considerations

Even though the physical model performs rather well, our application requires faster de-warping for each image. De-warping of the complete image is computationally costly if done on a per pixel basis. In order to avoid this, we use triangulation based warping using the correspondence points, while using the physical model to make sure we have a correct displacement value at each of these points. In addition, in order to make sure the triangulation is defined on the whole image domain, we create virtual points at the page boundary, using the physical model to establish the displacement values.

As for detecting the correspondences, we do so in a coarse to fine manner. In the first stage, a feature detector matches a consensus set of correspondences. The physical model is then used for a constrained matching of more features, namely the refinement stage.

The combined usage of triangulation and the physical model mitigates the disadvantages of each method. We take advantage of the exact nature of triangulation, as well as its low computational cost. By using the physical model we obtain three benefits. First, we increase the robustness to outliers. Second, the risk of fold-overs is reduced. Finally, our extrapolation can be naturally extended beyond the convex hull of the interpolation points.

#### 5. Results

We now show the results of the registration of the document template using each of these models. The results of the algorithms are shown in Figure 3, on portions from a scanned Italian tax form. White and black pixels denote pixels in which the template and the de-warped form agree. Blue and red areas denote areas that are black in the form only and in the template only, respectively.

Deformation models differ in the number of points needed to robustly estimate the model parameters, and the



**Figure 3. From left to right: the resulting image de-warping using the scale and shift, triangulation and polynomial (first row) and truncated polar and piecewise polar models (second row)**

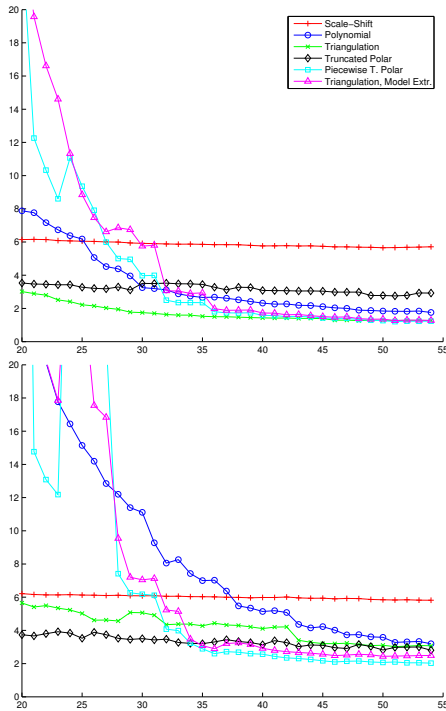
error achieved on the rest of the domain (the generalization error). These characteristics can be observed by using different subsets of the known points to learn the model, and testing the error on the remaining points.

In Figure 4, around 80 point correspondences were available, of which we use between 20 – 55 points to train the model. The RMS values of the error when estimating the rest of the points are displayed. This shows that the truncated polar model, if used in a piecewise manner, models the deformation as well as triangulation does. This is while still allowing a robust estimation using all the point correspondences available in the region, even in cases where correspondences were not highly accurate. At the bottom of Figure 4 we show the model accuracy when Gaussian noise is added to the coordinates of the point matches.

#### 6. Conclusions and Future Work

We have presented a new algorithm for the registration of structured forms, using physical motivation regarding the visual distortions caused by the sheet-feeding mechanism to document images. Since this model is tailored to the problem at hand, it can be made more accurate and more robust than existing methods. The results shown support this claim.

However, to achieve its full potential, the variations in the feeding speed and friction should be addressed. This can be done through the use of a piecewise polar model, as we



**Figure 4. RMS values of the resulting geometric error, without noise (top) and with Gaussian noise,  $\sigma^2 = 4$  (bottom)**

have shown, and we intend to further study this extension, estimating the parameters in an optimal manner.

Other future directions include a fast interpolation method based on the models shown, and statistically robust approximation of model parameters. We also intend to investigate weighting schemes for these models.

## References

- [1] A. H. Barr. Global and local deformations of solid primitives. In *SIGGRAPH '84*, pages 21–30. ACM Press, 1984.
- [2] R. Bernstein. Digital image processing of earth observation sensor data. *IBM Journal of Research and Development*, 20(1):40–57, 1976.
- [3] A. Bjorck and G. Dahlquist. *Numerical Methods*. Prentice Hall Professional Technical Reference, 1990. Translator-N. Anderson.
- [4] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(6):567–585, 1989.
- [5] J. P. Boyd. Defeating the Runge phenomenon for equispaced polynomial interpolation via Tikhonov regularization. *Applied Mathematics Letters*, 5(6):57–59, 1992.
- [6] W. E. L. Grimson. A Computational Theory of Visual Surface Interpolation. *Royal Society of London Philosophical Transactions Series B*, 298:395–427, Sept. 1982.
- [7] B. Hamann. *Curvature Approximation for Triangulated Surfaces*, pages 139–153. Number 8 in Computing Supplement. Springer-Verlag, 1993.
- [8] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.
- [9] T. Kanungo, R. M. Haralick, H. S. Baird, W. Stuezle, and D. Madigan. Document degradation models: Parameter estimation and model validation. In *IAPR Workshop on Machine Vision Applications*, pages 552–7, 1994.
- [10] T. Kanungo, R. M. Haralick, and I. T. Phillips. Global and local document degradation models. In *International Conference on Document Analysis and Recognition*, pages 730–734, 1993.
- [11] D. Loeckx, F. Maes, D. Vandermeulen, and P. Suetens. Non-rigid image registration using a statistical spline deformation model. In *IPMI*, pages 463–474, 2003.
- [12] C. Loop. Smooth spline surfaces over irregular meshes. In *SIGGRAPH '94*, pages 303–310. ACM Press, 1994.
- [13] Y. Min, S.-B. Cho, and Y. Lee. A data reduction method for efficient document skew estimation based on hough transformation. In *Proceedings of the 1996 International Conference on Pattern Recognition*, volume 3, page 732, 1996.
- [14] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer-Verlag, 1999.
- [15] J. O'Rourke. *Computational Geometry in C*. Cambridge University Press, 1998.
- [16] P. Ruhnau, A. Stahl, and C. Schnörr. Variational estimation of experimental fluid flows with physics-based spatio-temporal regularization. *Measurement Science and Technology*, 2007. In Print.
- [17] D. Ruprecht and H. Müller. Image warping with scattered data interpolation. *IEEE Computer Graphics and Applications*, 15(2):37–43, 1995.
- [18] A. L. Spitz. Correcting for variable skew in document images. *International Journal of Document Analysis and Recognition*, 6(3):181–189, 2003.
- [19] A. Ulges, C. H. Lampert, and T. M. Breuel. Document image dewarping using robust estimation of curled text lines. In *Proceedings of the Eighth International Conference on Document Analysis and Recognition*, pages 1001–1005, 2005.
- [20] Y. Wang. Physical model based non-rigid registration incorporating statistical shape information, 1998.
- [21] G. Wolberg and R. P. Loce. Inverting input scanner vibration errors. In *Proceedings of the 1995 International Conference on Image Processing*, volume 3, pages 129–132, 1995.
- [22] B. Yu and A. K. Jain. A form dropout system. In *Proceedings of the 1996 International Conference on Pattern Recognition*, volume 3, page 701, 1996.
- [23] L. Zagorchev and A. A. Goshtasby. A comparative study of transformation functions for nonrigid image registration. *IEEE Transactions on Image Processing*, 15(3):529–538, 2006.
- [24] Z. Zhang and C. Tan. Recovery of distorted document images from bound volumes. In *Proceedings of the Sixth International Conference on Document Analysis and Recognition*, page 429, 2001.