University of Szeged
PhD School in Computer Science

Geometric and Visual Reconstruction of Binary Shapes

Summary of the PhD Dissertation

József Németh

Supervisors:

Dr. János Csirik
Dr. Zoltán Kató

Szeged
2015

Introduction

Binary shapes play important role in the field of image processing, due to that

1. the images of many real world objects are basically binary shapes, *e.g.* text characters, traffic signs, bones or implants on X-ray images, *etc.*,
2. in most of the image processing applications at some point of the processing pipeline the images are binarized (*i.e.*, segmented).

In many cases, the additional information that we are working with a binary valued image limits the number of possible solutions (the search space) thus it can help to obtain more accurate results. For example, the linear inverse problems like tomography and deconvolution are usually under-determined, and have many possible solutions. Knowing that we are seeking a binary image restricts the search space, and as a result in the case of tomography, binary images can be usually reconstructed accurately only from a few projections.

In many other cases, however, the lack of rich intensity information makes it difficult to deal with these binary images. For example, image registration techniques commonly work with previously established point pairs and these correspondences are usually obtained using the intensity patterns around these points. Thus in the case of binary images these methods can not obtain appropriate point correspondences. On the other hand, in such cases we do not need to deal with the intensity change between the images. Therefore many techniques have been presented previously to register binary images using statistics computed using only the point coordinates of the shapes.

This work is a summary of the author's research results in the reconstruction and in the geometric registration of binary shapes. Table 1 shows the connection between the thesis points and the publications of the author. The planar homography related results of the thesis point III. was also previously presented in [13].

	[9]	[10]	[12]	[11]	[4]	[8]
I.	•					
II.		•				
III.			•	•	•	
IV.						•

Table 1: Correspondence between the thesis points and the publications.

I. Discrete Tomography with Unknown Intensity Levels

Here we propose a binary tomography reconstruction method which uses a higher order statistics based discretization term to enforce binary solutions. The digital image of size

$h \times w$ is represented by the column vector $\mathbf{x} \in \{c_1, c_2\}^n$, where $n = hw$ is the total number of the pixels and c_1, c_2 are the intensity values of the background pixels and the foreground (object) pixels. Suppose that projections have been taken from k different angles. Let l_i the number of measurements in the i*th* projection vector. All projection data are represented by the column vector $\mathbf{p} \in \mathbb{R}^m$, where $m = \sum_{i=1}^{k} l_i$ is the total number of measurements. The projection acquisition process is modeled by the system of linear equations

$$\mathbf{W}\mathbf{x} = \mathbf{p}, \tag{1}$$

where the matrix $\mathbf{W} \in \mathbb{R}^{m \times n}$ describes the projection geometry. The most frequently used beam geometry types in tomography as well as in binary tomography include parallel beam, fan beam, and cone beam.

The reconstruction is performed by the minimization of the following objective functional

$$E(\mathbf{x}, \alpha, \mu) = F(\mathbf{x}) + \lambda S(\mathbf{x}) + \mu D(\mathbf{x}, \alpha), \tag{2}$$

in which three expected properties of the solution are formulated. The so called *data fidelity term* represents that $\mathbf{x}$ should satisfy the projections:

$$F(\mathbf{x}) = \frac{1}{2}\|\mathbf{W}\mathbf{x} - \mathbf{p}\|_2^2. \tag{3}$$

This term expresses that the solution should have projections close to the observed data $\mathbf{p}$ in the least squares sense and it is commonly used in the field of tomography. As usually in the case of linear inverse problems, minimizing Eq. (3) alone could lead to non-feasible solutions, due to noisy projection data. To regularize the solution, the objective functional contains a *smoothness prior term*

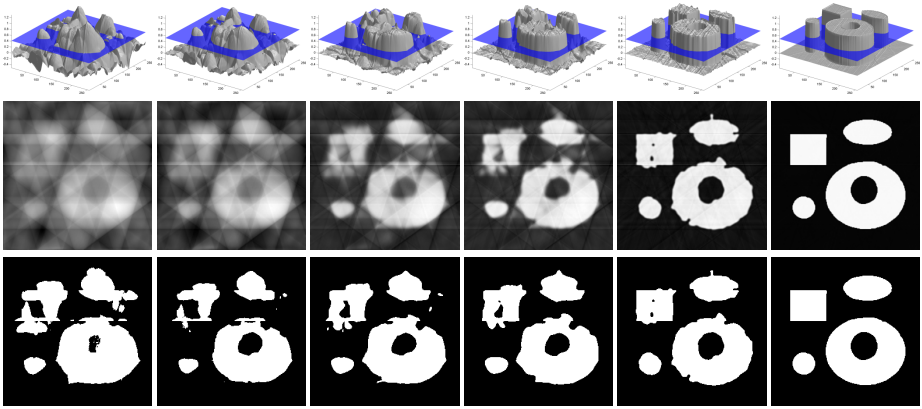
$$S(\mathbf{x}) = \frac{1}{2}\|L\mathbf{x}\|_2^2, \tag{4}$$

where L is the discrete Laplacian regularization matrix, *i.e.*, $L\mathbf{x}$ gives the same result as the 2-dimensional convolution with the Laplacian filter. This term penalizes solutions with high norm of second derivatives but allows the formation of edges. Setting its weight λ large enforces smooth regions even if the projections are noisy. Using only the data and the smoothness terms, the functional $F(\mathbf{x}) + \lambda S(\mathbf{x})$ is convex, and its minimum provides a continuous reconstruction.

Binary solutions are imposed by the *discretization term*

$$D(\mathbf{x}, \alpha) = n\frac{\|\mathbf{x} - \alpha\mathbf{1}_n\|_4^4}{(\|\mathbf{x} - \alpha\mathbf{1}_n\|_2^2)^2} - 1, \tag{5}$$

where α is the *discretization parameter*, $\mathbf{1}_n$ denotes the length-n column vector of ones, and $\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} x_i^p\right)^{1/p}$ is the general vector norm. This functional is minimized by binary images, if α is equal to the mean of the two intensity values, *i.e.*, there exists d for which $|x_i - \alpha| = d$, for every $i = 1, \ldots n$. In such cases, it has a value of 0, while it reaches its



$i=0$	$i=75$	$i=98$	$i=119$	$i=174$	$i=224$
$\mu=0$	$\mu\approx 414.8$	$\mu\approx 1.3\times 10^3$	$\mu\approx 3.7\times 10^3$	$\mu\approx 5.5\times 10^4$	$\mu\approx 6\times 10^5$
$\hat{\alpha}=0.4205$	$\hat{\alpha}=0.5239$	$\hat{\alpha}=0.4667$	$\hat{\alpha}=0.4706$	$\hat{\alpha}=0.4869$	$\hat{\alpha}=0.4970$
$D(\hat{\mathbf{x}},\hat{\alpha})=1.11$	$D(\hat{\mathbf{x}},\hat{\alpha})=0.74$	$D(\hat{\mathbf{x}},\hat{\alpha})=0.31$	$D(\hat{\mathbf{x}},\hat{\alpha})=0.15$	$D(\hat{\mathbf{x}},\hat{\alpha})=0.03$	$D(\hat{\mathbf{x}},\hat{\alpha})=0.00$

Figure 1: Demonstration of the convergence of the proposed method. The columns show intermediate reconstructions using 5 projections after different number of iterations. In the first row, 3-dimensional plots of the images are shown, in which the estimated mid-levels $\hat{\alpha}$ are indicated by horizontal planes. The second and third rows show the images and their $\hat{\alpha}$-thresholded versions. Below the images the corresponding discretization weight μ, the estimated mid-level $\hat{\alpha}$ and the value of the discretization term $D(\hat{\mathbf{x}},\hat{\alpha})$ can be found.

maximum value $n-1$ if all the pixel intensities are equal to α except one pixel. Thus it is bounded, and furthermore, two times continuously differentiable except at $\mathbf{x}=\alpha\mathbf{1}_n$. As a result, standard gradient based optimization techniques can be applied directly to minimize it. $D(\mathbf{x},\alpha)$ is closely related to the fourth order standardized moment (or kurtosis), except that the fourth and second moments are normalized using the mid-level α instead of the mean of the values.

Setting the weight μ of the discretization term large allows only binary solutions as it suppresses the other components of the energy functional. The following theorem states that for any $\xi>0$ a sufficiently high value of μ can be chosen such that the discretization level (*i.e.*, the value of the discretization term) will be at most ξ at the minimum point of the energy functional.

Theorem 1. *For any $\xi>0$, there exists $\mu(\xi)\in\mathbb{R}$, such that if $\mu\geq\mu(\xi)$ and $(\hat{\mathbf{x}},\hat{\alpha})=\arg\min_{\mathbf{x},\alpha} E(\mathbf{x},\alpha,\mu)$, then $D(\hat{\mathbf{x}},\hat{\alpha})\leq\xi$.*

The reconstruction problem for given $\xi>0$ and $\hat{\mu}\geq\mu(\xi)$ is defined as the following

minimization problem:

$$(\hat{\mathbf{x}}, \hat{\alpha}) = \arg\min_{\mathbf{x},\alpha} \; E(\mathbf{x}, \alpha, \hat{\mu}). \tag{6}$$

In the case of a binary image deconvolution method it was proposed, to optimize the value of the mid-level along with the solution and the point-spread-function in the same gradient descent based optimization process [6]. We found that in the case of the proposed method it is more efficient to estimate the mid-level directly to each intermediate solution $\hat{\mathbf{x}}$ based on the analysis of the discretization function. Thus we propose the following graduated optimization approach for solving the optimization problem in Eq. (6). Starting from an initial reconstruction $\hat{\mathbf{x}}$ obtained by the minimization of the objective functional without the discretization term (*i.e.*, with $\mu = 0$), in each iteration step

1. Estimate the mid-level $\hat{\alpha}$ for the current solution $\hat{\mathbf{x}}$.
2. Increase the weight of the discretization term (μ).
3. Refine the reconstruction $\hat{\mathbf{x}}$ by locally minimizing the objective functional using the new parameters $\hat{\alpha}$ and μ.

This method increasingly enforces binary solutions while iteratively approximates the mid-level (see Fig. 1). Note that the method does not require the estimation of the intensity values during this process neither the thresholding of the intermediate solutions. It was shown, that the objective functional Lipschitz-continuous, therefore it can be efficiently minimized by gradient based optimizers. The gray-intensity independence of the method was also shown, *i.e.*, that the algorithm can be implemented such that the reconstructed image does not depend on the values of the two intensities.

The proposed method was tested on a synthetic database. Experiments showed that the method is robust to the applied projection noise. In comparison to the PDM-DART method [15], our algorithm provided competitive results, while it clearly outperformed an other convex programming based approach.

The method was also successfully applied in a real data experiment, where the homogeneous part of a gas pressure regulator was reconstructed using its 18 projections (see Fig. 2).

The Author's Contributions

The author introduces a novel higher order statistics based binary tomography reconstruction technique, which can be used in those cases, when the intensity values of the images are unknown. He proposes an objective functional in which a discretization term is applied to prescribe binary solutions. He propose to minimize the objective functional by a graduated non-convexity optimization approach, in which the weight of the discretization term is increased during the optimization process to gradually enforce binary solutions. He

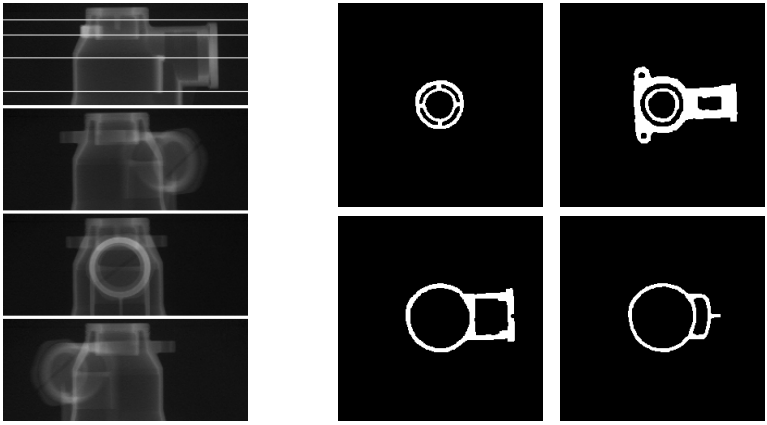


Figure 2: **First colum:** Four of the 18 projections of the homogeneous part of a gas pressure regulator. **Other columns:** Reconstruction results of the proposed method. The slices are indicated on the first projection image.

proposes to estimate the mid-level of the intensities directly to the intermediate solutions in each iteration step. This mid-level is estimated as the minimum of the discretization term. The author examines the convergence properties of the method and shows that the behaviour of the method is independent of the value of the intensities. He demonstrates the robustness of the algorithm against different strengths of projection noise. The author compares his algorithm to state-of-the-art methods and shows that his approach is a good alternative. He also successfully applies his algorithm to real projection data.

II. Binary Shape Deconvolution using Discrete Tomography

Our goal is to reconstruct an unknown image $f \in \mathbb{R}^{u \times v}$ from its blurred and noisy observation $g \in \mathbb{R}^{u \times v}$. The image degradation is modeled by the convolution of the image with the point-spread-function (PSF) and the addition of noise:

$$g = h * * f + n, \tag{7}$$

where $h \in \mathbb{R}^{p \times q}$ is the PSF, $**$ denotes the 2-dimensional convolution while $n \in \mathbb{R}^{u \times v}$ is the noise. We assume that the PSF is known. Our goal is to restore binary images, such as $b \in \{0,1\}^{u \times v}$. Since the scaling of the intensity values is unknown, we introduce this unknown scale factor s in the degradation model Eq. (7), thus $f = sb$.

Let the projection vector of an image f along an arbitrary angle θ denoted by $\check{f}^{\theta}$. We take advantage of that if $g = h * * f$ then $\check{g} = \check{h} * \check{f}$, where $*$ denotes the 1-dimensional convolution operator. The relationship between the projections of the original and the

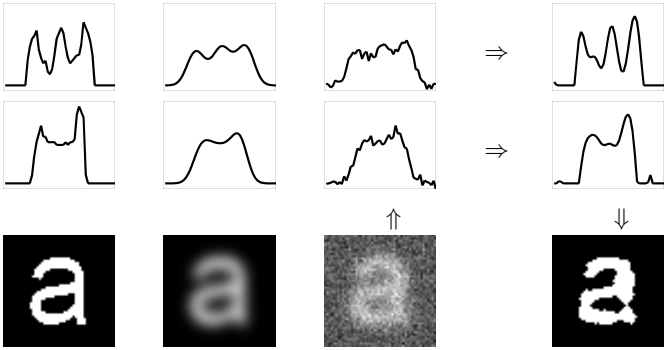


Figure 3: The degradation model and the basic idea of the proposed method. **First three columns:** the original image, the blurry image and the blurry and noisy image and their projections. **Fourth column:** the method restores the projections of the image then reconstructs the binary shape.

degraded images can be expressed as

$$\check{g} = \check{h} * \check{f} + \check{n}. \tag{8}$$

For a direction set $\Omega = \{\theta_i | i = 1, \ldots, k\}$ the proposed method restores each projection vector $\check{f}^{\theta_i}$ from the projection vectors $\check{g}^{\theta_i}$ and then as a second step reconstructs f from its restored projections (see Fig. 3). Convolution is a linear operation, thus Eq. (8) can be written as

$$\check{g} = H\check{f} + \check{n}, \tag{9}$$

where the H matrix represents the convolution operation with the PSF $\check{h}$. Since $\check{n}$ is unknown, this system of linear equations is ill-posed and it requires regularization which penalizes solutions of large norm. The standard version of the Tikhonov regularization takes the form

$$\check{f}_\lambda = \arg\min_{\check{f}} \|H\check{f} - \check{g}\|_2^2 + \lambda^2 \|\check{f}\|_2^2, \tag{10}$$

where λ is a positive constant, the regularization parameter, that controls the smoothness of the solution. An explicit solution for a given λ is given by

$$\check{f}_\lambda = (H^T H + \lambda^2 I)^{-1} H^T \check{g}, \tag{11}$$

where I denotes the identity matrix. To determine a suitable value of the regularization parameter λ we used the L-curve method, which is the log-log plot of the norm of the residual and the norm of the solution for different regularization parameters:

$$\mathcal{L} = \{(\log_2 \|H\check{f}_\lambda - \check{g}\|_2^2, \log_2 \|\check{f}_\lambda\|_2^2), \lambda \geq 0\}. \tag{12}$$

To obtain an optimal trade-off between the two values, the L-curve method proposes to choose λ^* which maximizes the curvature κ_λ of the curve $\mathcal{L}$.

To eliminate the negative elements of the solution we apply a further, generalized Tikhonov regularization based iterative method which iteratively enforces non-negative values.

The discrete tomography reconstruction requires the $\check{b}^{\theta_i} = \check{f}^{\theta_i}/s$ binary projections. The scale factor s can not be determined explicitly. However, its upper and lower bounds can be estimated, thus we can define a set S of its possible values. For each $s \in S$ we obtain the vectors $\check{b}^{\theta_i} = \left[\check{f}^{\theta_i}/s\right]$ as estimations of the projections of the unknown binary image b.

We consider the reconstruction from the vertical and horizontal projection vectors $\check{b}^0$ and $\check{b}^{\pi/2}$. It is well known, that usually two projections are not enough to reconstruct binary images, *i.e.*, there can be many binary images that have the same projections while in some other cases there is no binary image that satisfies the projections. We define the tomographic equivalence class:

$$U = U(\check{b}^0, \check{b}^{\pi/2}) = \{z \in \{0,1\}^{u \times v} : \check{z}^0 = \check{b}^0, \check{z}^{\pi/2} = \check{b}^{\pi/2}\}. \tag{13}$$

While generally $|U| > 1$, we are interested in a solution which is the most similar to the input image g. For that purpose a model image m has been created from the input image g by removing noise of variance σ_n^2 (which has been estimated after the estimation of the projections). We show how to find such a solution using a *minimum cost maximal flow* algorithm. The expectation that the solution z_s should be similar to the model image means that on those (x, y) positions where $z(x, y) = 1$ it is most likely that the model image has high grayscale values. This can be formulated by the following minimization problem:

$$z_s = \underset{z \in U}{\arg\min} \left(-\sum_{x,y} z(x,y) m(x,y) \right). \tag{14}$$

This way the reconstruction problem is traced back to the *minimum cost maximal flow* (MCMF) problem [1]. In this network the *supply* and *demand* nodes are representing the projection vectors while the edges are representing the image pixels. The edge $S_x \to T_y$ corresponding to the (x, y) pixel has a flow capacity equal to 1 and a flow cost equal to $-m(x, y)$. The minimum cost maximal flow can be found in polynomial time [14] and it determines the solution z_s of the discrete tomography problem. The pixel $z_s(x, y)$ gets a value of 1, if and only if the flow passes through the edge $S_x \to T_y$. For different scale factors $s \in S$ the method finds different binary solutions z_s. To choose an optimal solution z, the method compares each z_s to the input image g in least-squares sense.

To examine the performance of the method we created a synthetic dataset of images of 62 alphanumeric characters (of size 59×59 pixels) and their 1550 degraded versions. Each image was blurred by different Gaussian filters and white noises of different levels was added

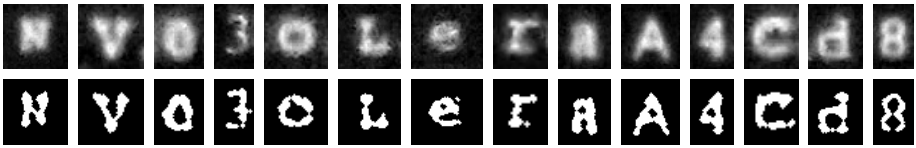


Figure 4: Example results on letters extracted from out-of-focus document images. The original images are shown in the first row while results of the proposed method are in the second row.

to each blurred image. Comparisons to the widely used Lucy-Richardson method showed that our algorithm provides competitive results. The performance of the method was also examined on out-of-focus character images taken with real camera. Some restoration results can be seen in Fig. 4.

The Author's Contributions

The author proposes a binary tomography based approach for the deconvolution of binary images. He proposes to deconvolve the projections of the blurred images using a Tikhonov regularization based approach, in which the optimal value of the regularization parameter is found by the L-curve method. To reconstruct the shapes from their deblurred projections, he applies a maximum-flow based binary tomography method. He shows in comparative tests that his method provides more reliable results then another widely-used method. He also presents results on real out-of-focus images.

III. Nonlinear Registration of Binary Shapes

We introduce a general binary registration framework. Suppose that $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ is a diffeomorphism, *i.e.*, a differentiable and invertible transformation and its inverse is also differentiable. Our goal is to estimate its parameters with which the transformation aligns the input shapes. Let the corresponding point coordinates on the *template* and *observation* shapes denoted by $\mathbf{x} = [x_1, x_2]^T \in \mathbb{R}^2$ and $\mathbf{y} = [y_1, y_2]^T \in \mathbb{R}^2$. Then

$$\mathbf{y} = \varphi(\mathbf{x}) \quad \Leftrightarrow \quad \mathbf{x} = \varphi^{-1}(\mathbf{y}). \tag{15}$$

This relation remains valid if we apply a function $\omega : \mathbb{R}^2 \to \mathbb{R}^n$, on both sides of the equation [3, 12, 11]:

$$\omega(\mathbf{y}) = \omega(\varphi(\mathbf{x})) \quad \Leftrightarrow \quad \omega(\mathbf{x}) = \omega(\varphi^{-1}(\mathbf{y})). \tag{16}$$

Integrate on both sides we obtain the following equation:

$$\int_{\mathcal{F}_o} \omega(\mathbf{y}) d\mathbf{y} \quad = \quad \int_{\mathcal{F}_t} \omega(\varphi(\mathbf{x}))\, |J_\varphi(\mathbf{x})|\, d\mathbf{x}. \tag{17}$$

where $\mathcal{F}_t$ and $\mathcal{F}_o$ are the foreground regions of the *template* and *observation* shapes and $|J_\varphi| : \mathbb{R}^2 \to \mathbb{R}$ is the Jacobian of the transformation, given by:

$$|J_\varphi(\mathbf{x})| = \begin{vmatrix} \frac{\partial \varphi_1}{\partial x_1} & \frac{\partial \varphi_1}{\partial x_2} \\ \frac{\partial \varphi_2}{\partial x_1} & \frac{\partial \varphi_2}{\partial x_2} \end{vmatrix}. \tag{18}$$

The basic idea of the proposed method is to generate a sufficient number of equations using a set of linearly independent ω functions. Since an equation set up using $\omega : \mathbb{R}^2 \to \mathbb{R}^n$ $(n > 1)$ can be decomposed into n equations, we can assume that $n = 1$ without any restriction.

Suppose that φ has k parameters and let $\omega_i : \mathbb{R}^2 \to \mathbb{R}$, $(i = 1, \ldots, \ell)$ the set of applied functions. To solve for all unknowns, we need at least k equations, hence $\ell \geq k$. Thus we can obtain the following system of equations

$$\int_{\mathcal{F}_o} \omega_i(\mathbf{y}) d\mathbf{y} = \int_{\mathcal{F}_t} \omega_i\big(\varphi(\mathbf{x})\big)\, |J_\varphi(\mathbf{x})|\, d\mathbf{x}, \qquad i = 1, \ldots, \ell, \tag{19}$$

where each ω_i function provides one new equation. Each applied ω_i function can be considered as a coloring of the shape and the integrals in Eq. (19) give the volumes of the ω_i functions over the area of the shape. The equations thus match these volumes. Each equation provides additional constraints and the solution of this system gives the estimation of the parameters of the aligning transformation.

We applied our registration framework on different deformation classes:

1. **Planar homography** is the projective transformation between the images of the same planar object. It plays an important role in computer vision.

2. **Polynomial transformations** are often used to approximate other transformation models and general deformations.

3. **Thin Plate Spline** (TPS) model is widely used to approximate general non-rigid deformations.

To avoid extreme numerical values of the integrals, we applied some normalization on both the pixel coordinates and the ω_i functions. For that purpose, we normalized the coordinates of both shapes into the square $[-0.5, 0.5] \times [-0.5, 0.5]$ and chose ω_i with a range limited to an interval (*e.g.* $[-1, 1]$). Despite of these normalizations, the integrals in our equations could take values of different range of magnitude yielding a different contribution from each equation to the objective function value in the optimization process. Thus we normalized the equations by dividing the integrals with an appropriate constant, which is given by the integral of the applied ω_i function on a circle with center in the origin and a radius $\frac{\sqrt{2}}{2}$:

$$N_i = \int_{\|\mathbf{x}\| \leq \frac{\sqrt{2}}{2}} |\omega_i(\mathbf{x})| d\mathbf{x}, \tag{20}$$

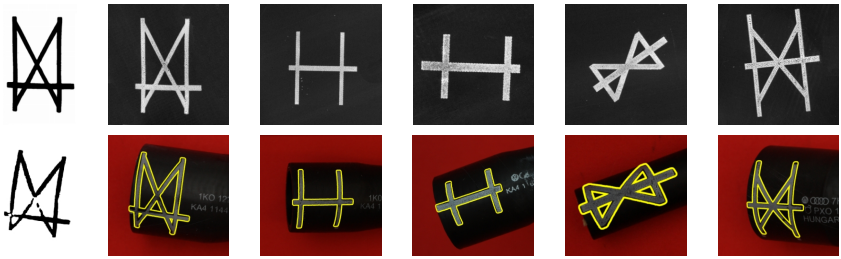


Figure 5: Registration results of printed signs. **Top:** planar *templates*. **Bottom:** the corresponding *observations* with the overlaid contour of the registration results. The first image pair shows the segmented regions used for registration. Note the typical segmentation errors. (Images provided by *ContiTech Fluid Automotive Hungária Ltd.*)

and the normalized version of Eq. (19) becomes

$$\frac{\int_{\mathcal{F}_o} \omega_i(\mathbf{y}) d\mathbf{y}}{N_i} = \frac{\int_{\mathcal{F}_t} \omega_i\big(\varphi(\mathbf{x})\big)\, |J_\varphi(\mathbf{x})|\, d\mathbf{x}}{N_i}, \quad i = 1, \ldots, \ell. \tag{21}$$

The system of equations in Eq. (21) has been constructed in the continuous space, in practice, however, we have limited resolution digital images. Therefore the integrals can only be approximated by finite sums over the foreground pixels. Let us denote by F_t and F_o the finite pixel sets corresponding to the continuous pixel regions $\mathcal{F}_t$ and $\mathcal{F}_o$. Then Eq. (21) can be approximated by the following system of equations:

$$\frac{1}{N_i} \sum_{\mathbf{y} \in F_o} \omega_i(\mathbf{y}) = \frac{1}{N_i} \sum_{\mathbf{x} \in F_t} \omega_i\big(\varphi(\mathbf{x})\big)\, |J_\varphi(\mathbf{x})|\,, \quad i = 1, \ldots, \ell. \tag{22}$$

The parameters of the aligning transformations are obtained as the solution of this system of equations. While the equations are nonlinear, we found that it can be solved efficiently by *Levenberg-Marquardt algorithm* [7] in least-squares sense. In the case of planar homography, the equations can be written in three alternative forms by making use of the corresponding inverse transformation. The additional equations provide additional constraints and thus help the optimizer to find the optimal solution.

We examined the performance of the method using different $\{\omega_i\}$ function sets. We considered different power, polynomial, and trigonometric function sets. Based on the results we recommend to use low order polynomials for computational efficiency. In the case of planar homography, we showed that the performance of the algorithm can be improved if we apply the framework to the Taylor series expansion of the transformation.

We tested the performance of the algorithm on large synthetic datasets for each transformation models. We also compared the proposed method to the Shape Context [2] method, and found that it provides more accurate alignments. We also examined the robustness

of the proposed method against 4 different type of segmentation errors on the synthetic dataset of planar homography. Results showed that the proposed method is quite robust in those cases where the segmentation errors are uniformly distributed on the shapes. In those cases, however, when the segmentation errors are consisted of larger continuous regions, like occlusion and disocclusion, it provides less accurate results compared to the Shape Context method.

The proposed method has been applied in the case of different real applications, like the registration of handwritten characters (TPS model), traffic signs (planar homography), and in an industrial quality inspection problem (problem specific transformation model, see Fig. 5).

The Author's Contributions

The author addresses the problem of nonlinear registration of binary shapes. A general registration framework is used which traces back the registration problem to a system of nonlinear equations. He applies the framework to different nonlinear transformation classes such as planar homography, polynomial transformations, and thin plate splines. He goes into the implementational details and shows that the equations can be written in three alternative forms which improves the registration results. In the case of planar homography, the author shows that the performance of the algorithm can be improved if it is applied to the Taylor series expansion of the transformation. The author proposes different ω function sets to construct the system of equations and compares them. He proposes to normalize the equations to guarantee equal contribution to the objective functional. The author compares the method to other state-of-the-art methods on synthetic datasets, and examines the robustness of the algorithm against different types of segmentation errors. The author shows that the method can be applied in different real world applications.

IV. Affine Invariants Based Projective Registration of Binary Shapes

We introduce a two-step method to estimate the parameters of a planar homography transformation that aligns two binary shapes. The parameters of such a transformation are the H_{ij} elements of the 3×3 matrix $\mathbf{H}$, while $H_{33}=1$ is fixed. The H_{31} and H_{32} parameters of the transformation are responsible for the *perspective* distortion, while the others generate an *affine* transformation. The transformation can be decomposed as follows:

$$h = h^a \circ h^p \tag{23}$$

where $h^p : \mathbb{R}^2 \to \mathbb{R}^2$, $h^p(\mathbf{x}) = [h_1^p(\mathbf{x}), h_2^p(\mathbf{x})]^T$ is a nonlinear transformation:

$$\begin{aligned} h_1^p(\mathbf{x}) &= \frac{x_1}{p_1 x_1 + p_2 x_2 + 1} \\ h_2^p(\mathbf{x}) &= \frac{x_2}{p_1 x_1 + p_2 x_2 + 1}, \end{aligned} \tag{24}$$

resulting only perspective distortion, and $h^a : \mathbb{R}^2 \to \mathbb{R}^2$, $h^a(\mathbf{x}) = [h^a_1(\mathbf{x}), h^a_2(\mathbf{x})]^T$ is an affine transformation:

$$\begin{aligned} h^a_1(\mathbf{x}) &= a_{11}x_1 + a_{12}x_2 + a_{13} \\ h^a_2(\mathbf{x}) &= a_{21}x_1 + a_{22}x_2 + a_{23}, \end{aligned} \tag{25}$$

Thus we can write the relationship between the shapes as follows:

$$\mathcal{F}_\mathbf{t} = (h^a \circ h^p)(\mathcal{F}_\mathbf{o}) = h^a(h^p(\mathcal{F}_\mathbf{o})). \tag{26}$$

The proposed method estimates the p_i parameters of the perspective component h^p and the a_i parameters of the affine component h^a in two distinct steps (see Fig. 6), then using Eq. (23) we get the H_{ij} parameters of h.

If Eq. (26) stands, then for any affine-invariant function $I : \mathbb{R}^2 \to \mathbb{R}$:

$$I(\mathcal{F}_\mathbf{t}) = I(h^p(\mathcal{F}_\mathbf{o})). \tag{27}$$

Given a set of independent affine invariant functions $I_i : \mathbb{R}^2 \to \mathbb{R}$, $i = 1 \dots n$, we obtain a system of equations:

$$I_i(\mathcal{F}_\mathbf{t}) = I_i(h^p(\mathcal{F}_\mathbf{o})). \tag{28}$$

The solution of this system of equations provides the parameters of h^p. It is clearly a highly nonlinear system and thus do not have exact solution. However our experiments showed that it can be efficiently solved by a general nonlinear solver.

We use affine moment invariants [5], because they allows efficient numerical estimation of the system of equations in Eq. (28). The left hand sides of the equations in Eq. (28) do not depend on the parameters of h^p so they can be estimated directly using the point coordinates of the template image. The geometric moment m_{rs} of order $(r+s)$ of a shape $\mathcal{F}$ is defined as

$$m_{rs}(\mathcal{F}) = \int_\mathcal{F} x_1^r x_2^s \mathrm{d}\mathbf{x}. \tag{29}$$

The affine moment invariants are rely on central moments:

$$\mu_{rs}(\mathcal{F}) = \int_\mathcal{F} (x_1 - c_1)^r (x_2 - c_2)^s \mathrm{d}\mathbf{x} \tag{30}$$

where the coordinates of the center of mass of the shape are given by:

$$c_1 = m_{10}(\mathcal{F})/m_{00}(\mathcal{F}) \quad \text{and} \quad c_2 = m_{01}(\mathcal{F})/m_{00}(\mathcal{F}). \tag{31}$$

For example the first two affine moment invariants are given as follows:

$$\begin{aligned} I_1 &= (\mu_{20}\mu_{02} - \mu_{11}^2)/\mu_{00}^4 \\ I_2 &= (-\mu_{30}^2\mu_{03}^2 + 6\mu_{30}\mu_{21}\mu_{12}\mu_{03} - 4\mu_{30}\mu_{12}^3 - 4\mu_{21}^3\mu_{03} + 3\mu_{21}^2\mu_{12}^2)/\mu_{00}^{10}. \end{aligned} \tag{32}$$

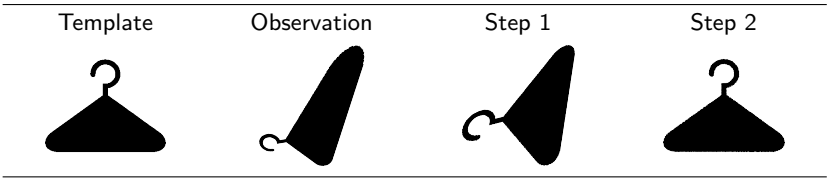


Figure 6: The registration process: The first step removes only the perspective distortion from the observation image while the second step restores the affine transformation and thus align it to the original template image.

Given fixed parameters of h^p we show how to compute the right hand sides of Eq. (28) by making use of the Jacobian $|J_{h^p}|$ of the transformation, thus avoiding the generation of the image $h^p(\mathcal{F}_o)$. For a shape $\mathcal{F}$ that is distorted by h^p the geometric moment can be estimated as follows:

$$m_{rs}(h^p(\mathcal{F})) = \int_{\mathcal{F}} [h_1^p(\mathbf{x})]^r \, [h_2^p(\mathbf{x})]^s \, |J_{h^p}(\mathbf{x})| \mathrm{d}\mathbf{x} \tag{33}$$

where the Jacobian of the perspective distortion is given by

$$|J_{h^p}(\mathbf{x})| = \frac{1}{(p_1 x_1 + p_2 x_2 + 1)^3}, \tag{34}$$

On the perspectively distorted shape $h^p(\mathcal{F})$ the central moments are given by

$$\mu_{rs}(h^p(\mathcal{F})) = \int_{\mathcal{F}} [h_1^p(\mathbf{x}) - c_1]^r \, [h_2^p(\mathbf{x}) - c_2]^s \, |J_{h^p}(\mathbf{x})| \mathrm{d}\mathbf{x}, \tag{35}$$

where

$$c_1 = m_{10}(h^p(\mathcal{F}))/m_{00}(h^p(\mathcal{F})) \quad \text{and} \quad c_2 = m_{01}(h^p(\mathcal{F}))/m_{00}(h^p(\mathcal{F})). \tag{36}$$

For fixed values of the parameters p_1 and p_2, the affine moment invariants $I(h^p(F))$ in the right hand side of the system Eq. (28) can be obtained using the central moments in Eq. (35) that can be estimated using only the foreground points of the shape $\mathcal{F}$. Thus we avoid to generate the $h^p(\mathcal{F})$ images which would be very time consuming.

After the perspective distortion is recovered (*i.e.*, its parameters p_1 and p_2 are determined) the affine transformation h^a should be estimated between the shapes $\mathcal{F}_t$ and $h^p(\mathcal{F}_o)$. For that purpose we used the approach proposed in [3] and in order to avoid the generation of $h^p(\mathcal{F}_o)$ we modified the method by making use of the Jacobian $|J_{h^p}|$ of the perspective part. The following system of equations can be written for the a_{ij} parameters of the affine transformation

$$\int_{F_t} y_k^n d\mathbf{y} = |J_{h^a}| \sum_{i=1}^{n} \binom{n}{i} \sum_{j=0}^{i} \binom{i}{j} a_{k1}^{n-i} a_{k2}^{i-j} a_{k3}^{j} \int_{F_o} h_1^p(\mathbf{x})^{n-i} h_2^p(\mathbf{x})^{i-j} |J_{h^p}(\mathbf{x})| d\mathbf{x}, \tag{37}$$

Figure 7: Registration results on traffic signs. The images used as *observations* are shown in the first row, and below them the corresponding *templates* with the overlayed contours of the registration results.

for $n = 1, 2, 3$ and $k = 1, 2$. This system contains six polynomial equations up to order three which is enough to solve for all the 6 unknowns.

The Jacobian of an affine transformation is constant over the whole plane, thus it can be simply estimated as the ratio of the areas of the shapes:

$$|J_{h^a}| = \frac{\int_{\mathcal{F}_t} d\mathbf{y}}{\int_{\mathcal{F}_o} |J_{h^p}(\mathbf{x})| d\mathbf{x}} \tag{38}$$

While the system Eq. (37) may have many solutions, we can select the real root which corresponds to the determinant what we computed in Eq. (38). Note that the solution is not unique if the shape is affine symmetric.

Putting together the projective transformation h^p and the affine transformation h^a, we get the aligning planar homography transformation h. In our experiments, we used the I_3, I_4, I_5, I_6 invariants and applied the *differential evolution* method to solve Eq. (28). Experiments on a synthetic dataset showed that the proposed method can outperform the method that was described in the previous section which can not cope with those cases when the images were rotated more than 90 degrees. The robustness of the proposed method against segmentation errors should be examined in a future work, especially since affine moment invariants are sensitive to such degradations. Nevertheless, we were able to obtain good registration results on several traffic sign image pairs (see Fig. 7).

The Author's Contributions

The author proposes an affine invariant moments based approach to estimate the parameters of a planar homography transformation between binary shapes. He shows how to

decompose the transformation into a perspective and an affine part and recover their parameters separately in two consecutive steps. He shows how to estimate the projective parameters of the transformation using affine moment invariants. As a second step, he applies an affine registration method to determine the remaining parameters of the transformation. The author compares the method to the algorithm presented in the previous thesis point and shows results obtained by the algorithm on real images.

References

[1] Kees Joost Batenburg. A network flow algorithm for reconstructing binary images from discrete x-rays. *Journal of Mathematical Imaging and Vision*, 27(2):175–191, 2007.

[2] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape context. *Transaction on Pattern Analysis and Machine Intelligence*, 24(4):509–522, April 2002.

[3] Csaba Domokos and Zoltan Kato. Parametric estimation of affine deformations of planar shapes. *Pattern Recognition*, 43:569–578, March 2010.

[4] Csaba Domokos, Jozsef Nemeth, and Zoltan Kato. Nonlinear shape registration without correspondences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5):943–958, May 2012.

[5] Jan Flusser, Tomás Suk, and Barbara Zitová. *Moments and Moment Invariants in Pattern Recognition*. Wiley & Sons, October 2009.

[6] Jeongtae Kim and Soohyun Jang. High order statistics based blind deconvolution of bi-level images with unknown intensity values. *Optics Express*, 18(12):12872–12889, June 2010.

[7] Donald W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, 11(2):431–441, 1963.

[8] Jozsef Nemeth. Recovering projective transformations between binary shapes. In *Proceedings of Advanced Concepts for Intelligent Vision Systems*, volume 7517, pages 374–383, Brno, Czech Republic, 2012.

[9] Jozsef Nemeth. Discrete tomography with unknown intensity levels using higher-order statistics. *Journal of Mathematical Imaging and Vision*, 2015. Online first, DOI: 10.1007/s10851-015-0581-0.

[10] Jozsef Nemeth and Peter Balazs. Restoration of blurred binary images using discrete tomography. In *Proceedings of Advanced Concepts for Intelligent Vision Systems*, volume 8192, pages 80–90, Poznan, Poland, 2013.

[11] Jozsef Nemeth, Csaba Domokos, and Zoltan Kato. Nonlinear registration of binary shapes. In *Proceedings of International Conference on Image Processing*, pages 1101–1104, Cairo, Egypt, November 2009. IEEE.

[12] Jozsef Nemeth, Csaba Domokos, and Zoltan Kato. Recovering planar homographies between 2D shapes. In *Proceedings of International Conference on Computer Vision*, pages 2170–2176, Kyoto, Japan, September 2009. IEEE.

[13] József Németh. Síkhomográfia paramétereinek becslése bináris képeken. Proceedings of National Scientific Students' Associations Conference, April 2009. Supervisors: Zoltán Kató and Csaba Domokos. Note: in Hungarian.

[14] James B. Orlin. A polynomial time primal network simplex algorithm for minimum cost flows. In *Proceedings of the Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '96, pages 474–481, Philadelphia, PA, USA, 1996. SIAM.

[15] Wim van Aarle, Kees Joost Batenburg, and Jan Sijbers. Automatic parameter estimation for the discrete algebraic reconstruction technique (DART). *IEEE Transactions on Image Processing*, 21(11):4608–4621, 2012.