

Recovering Planar Homographies between 2D Shapes*

Jozsef Nemeth, Csaba Domokos, Zoltan Kato

Department of Image Processing and Computer Graphics, University of Szeged

H-6701 Szeged, PO. Box 652., Hungary, Fax: +36 62 546-397

{nemjzozs, dcs, kato}@inf.u-szeged.hu

Abstract

Images taken from different views of a planar object are related by planar homography. Recovering the parameters of such transformations is a fundamental problem in computer vision with various applications. This paper proposes a novel method to estimate the parameters of a homography that aligns two binary images. It is obtained by solving a system of nonlinear equations generated by integrating linearly independent functions over the domains determined by the shapes. The advantage of the proposed solution is that it is easy to implement, less sensitive to the strength of the deformation, works without established correspondences and robust against segmentation errors. The method has been tested on synthetic as well as on real images and its efficiency has been demonstrated in the context of two different applications: alignment of hip prosthesis X-ray images and matching of traffic signs.

1. Introduction

Perspective images of planar scenes are usual in perception of man made environments. In such cases, a planar scene and its image are related by a plane to plane homography, also known as a plane projective transformation. Estimating the parameters of the transformation between two views of the same planar object is a fundamental problem in computer vision with various applications (*e.g.* registration [16], shape matching [12], image mosaicing). It is well known that the mapping between two such images is also a planar homography.

Basically there are two classical approaches: One of them is called *landmark based*, which consists in establishing correspondences between feature points and then the transformation parameters are obtained as a solution of a system of homogeneous equations set up using the coordinates of point pairs (see [16] for a good survey on clas-

sical techniques). On the other hand, *featureless* methods estimate the transformation parameters directly from image intensity values over corresponding regions [10]. Both approaches rely on the availability of rich radiometric information: Landmark based methods usually match local brightness patterns around salient points while featureless methods make use of intensity correlation between image patches. In the case of differently exposed images, researchers have proposed time consuming algorithms to cope with brightness changes across image pairs. Recently, Francos *et al.* proposed an elegant solution for the estimation of a homeomorphism [5] between graylevel images. Assuming that corresponding regions have been identified and the intensities within these region pairs differ only by a zero mean Gaussian noise, they transform the matching problem into the solution of a linear system of equations. In many cases, however, radiometric information may not be available (*e.g.* binary shapes), or it is either very limited (*e.g.* images of traffic signs, letterings) or heavily distorted (*e.g.* X-ray images). This makes classical brightness-based features unreliable thus challenging current matching techniques.

While many techniques exists for matching graylevel or color image pairs, the alignment of binary images (or shapes) received less attention. This problem is challenged by the lack of rich radiometric information on which classical approaches are based. On the other hand, most of the current approaches to binary registration are restricted to the affine group [1]. The extension of these methods to a general plane projective transformation is far from trivial because an affine transformation is linear in the plane while a planar homography is usually not. For example, in [1] Domokos *et al.* showed that it is possible to trace back the affine matching problem to an exactly solvable polynomial system of equations. Another approach is using only the contour points of continuous regions. In [12], a novel similarity metric, called *shape context*, is proposed which is then used to establish correspondences between contour points. The shape matching problem is then solved via the minimization of an energy functional constructed from the point matches.

*Partially supported by the Hungarian Scientific Research Fund (OTKA) – K75637 and the TÁMOP-4.2.2/08/1/2008-0008 program of the Hungarian National Development Agency.

In this paper, we propose a novel method to recover a planar homography between binary shapes. The proposed method does not need any feature extraction or established correspondences, it works only with the coordinates of the foreground pixels. The basic idea is to set up a system of nonlinear equations by integrating a set of nonlinear functions over the image domains and then solve it by classical *Levenberg-Marquardt algorithm*. We show that the method provides good results on a variety of real images where classical methods fail. Furthermore, we quantitatively analyze the performance and robustness of the algorithm on a large set of synthetic images.

2. Estimation of planar homographies

In a general setting, we are looking for a plane projective transformation which aligns a pair of binary images (or shapes) such that one image (called the *observation*) becomes similar to the second one (called the *template*). Let us denote the homogeneous coordinates of the *template* and *observation* points by $\mathbf{x}' = [x'_1, x'_2, x'_3]^T \in \mathbb{P}^2$ and $\mathbf{y}' = [y'_1, y'_2, y'_3]^T \in \mathbb{P}^2$ respectively. The identity relation for the aligning planar homography is then as follows

$$\mathbf{y}' = \mathbf{H}\mathbf{x}' \Leftrightarrow \mathbf{x}' = \mathbf{H}^{-1}\mathbf{y}', \quad (1)$$

where $\mathbf{H} = \{h_{ij}\}$ is the unknown 3×3 transformation matrix that we want to recover. Note that \mathbf{H} has only 8 degree of freedom thus one of its 9 elements can be fixed (we set $h_{33} = 1$)¹. As usual, the inhomogeneous coordinates $\mathbf{y} = [y_1, y_2]^T \in \mathbb{R}^2$ of a homogeneous point \mathbf{y}' are given by

$$\begin{aligned} y_1 &= \frac{y'_1}{y'_3} = \frac{h_{11}x_1 + h_{12}x_2 + h_{13}}{h_{31}x_1 + h_{32}x_2 + 1} \equiv \varphi_1(\mathbf{x}) \\ y_2 &= \frac{y'_2}{y'_3} = \frac{h_{21}x_1 + h_{22}x_2 + h_{23}}{h_{31}x_1 + h_{32}x_2 + 1} \equiv \varphi_2(\mathbf{x}), \end{aligned} \quad (2)$$

where $\varphi_i : \mathbb{R}^2 \rightarrow \mathbb{R}$. Indeed, planar homography is a linear transformation in the projective plane \mathbb{P}^2 , but it becomes nonlinear within the Euclidean plane \mathbb{R}^2 . The nonlinear transformation corresponding to \mathbf{H} is denoted by $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $\varphi(\mathbf{x}) = [\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x})]^T$, and the identity relation in Eq. (1) becomes

$$\mathbf{y} = \varphi(\mathbf{x}) \Leftrightarrow \mathbf{x} = \varphi^{-1}(\mathbf{y}), \quad (3)$$

where $\varphi^{-1} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is the inverse transformation corresponding to the inverse homography \mathbf{H}^{-1} . The main challenge here is that we can only observe the inhomogeneous coordinates \mathbf{x} and \mathbf{y} . As a consequence, we cannot handle the problem as a linear one, we must use the nonlinear relation φ .

¹Theoretically h_{33} could be 0 or small. However, we normalize the coordinates of the input shapes into $[-0.5, 0.5]$, center of mass being the origin. Thus if $h_{33} = 0$ or small then \mathbf{H} would map the origin into infinity or a very distant point, which is quite unlikely to be observed in practice.

2.1. Direct solution

Classical landmark-based approaches would now identify at least 4 point pairs and solve the system of linear equations obtained from Eq. (3). However, we are interested in a direct solution without solving the correspondence problem. For that purpose, let shapes be represented by their characteristic function $\chi : \mathbb{R}^2 \rightarrow \{0, 1\}$, where 0 and 1 correspond to the background and foreground respectively. If we denote the *template* by χ_t and the *observation* by χ_o , then Eq. (3) implies

$$\chi_o(\mathbf{y}) = \chi_o(\varphi(\mathbf{x})) = \chi_t(\mathbf{x}). \quad (4)$$

Since we do not have established point pairs, we cannot directly use Eq. (3)–(4). However, we can multiply these equations and then integrate out individual point correspondences

$$\int_{\mathbb{R}^2} \mathbf{y} \chi_o(\mathbf{y}) d\mathbf{y} = \int_{\mathbb{R}^2} \varphi(\mathbf{x}) \chi_t(\mathbf{x}) |J_\varphi(\mathbf{x})| d\mathbf{x}, \quad (5)$$

where the integral transformation $\mathbf{y} = \varphi(\mathbf{x})$, $d\mathbf{y} = |J_\varphi(\mathbf{x})| d\mathbf{x}$ has been applied. The Jacobian determinant $|J_\varphi| : \mathbb{R}^2 \rightarrow \mathbb{R}$ is given by

$$|J_\varphi(\mathbf{x})| = \begin{vmatrix} \frac{\partial \varphi_1}{\partial x_1} & \frac{\partial \varphi_1}{\partial x_2} \\ \frac{\partial \varphi_2}{\partial x_1} & \frac{\partial \varphi_2}{\partial x_2} \end{vmatrix} = \frac{|\mathbf{H}|}{(h_{31}x_1 + h_{32}x_2 + 1)^3}. \quad (6)$$

Since χ_t and χ_o take only values from $\{0, 1\}$, multiplying Eq. (3) and Eq. (4) essentially restricts the integrals in Eq. (5) to the foreground regions $\mathcal{F}_t = \{\mathbf{x} \in \mathbb{R}^2 | \chi_t(\mathbf{x}) = 1\}$ and $\mathcal{F}_o = \{\mathbf{y} \in \mathbb{R}^2 | \chi_o(\mathbf{y}) = 1\}$. Thus we obtain a finite integral equation

$$\int_{\mathcal{F}_o} \mathbf{y} d\mathbf{y} = \int_{\mathcal{F}_t} \varphi(\mathbf{x}) |J_\varphi(\mathbf{x})| d\mathbf{x}. \quad (7)$$

This equation implies that the finite domains \mathcal{F}_t and \mathcal{F}_o are also related by the unknown transformation φ , *i.e.* we match the shapes as a whole instead of point correspondences. In fact, Eq. (7) is a nonlinear system of two equations for $y_i, \varphi_i(\mathbf{x})$ ($i = 1, 2$). Unfortunately, two equations alone are not enough to solve for 8 unknowns.

In order to generate more equations, let us remark that the identity relation in Eq. (3) remains valid when a function $\omega : \mathbb{R}^2 \rightarrow \mathbb{R}$ is acting on both sides of the equation [5, 1]. Indeed, for a properly chosen ω

$$\omega(\mathbf{y}) = \omega(\varphi(\mathbf{x})) \Leftrightarrow \omega(\mathbf{x}) = \omega(\varphi^{-1}(\mathbf{y})). \quad (8)$$

Intuitively, ω generates a consistent coloring of the binary shapes as shown in Fig. 1. The basic idea of the proposed approach is to generate enough linearly independent equations by making use of nonlinear ω functions. Clearly, each

$$\begin{aligned}
\varphi_1^{-1}(\mathbf{y}) &= x_1 = \frac{(h_{22} - h_{32}h_{23})y_1 - (h_{12} - h_{32}h_{13})y_2 + h_{23}h_{12} - h_{22}h_{13}}{(h_{32}h_{21} - h_{31}h_{22})y_1 - (h_{32}h_{11} - h_{31}h_{12})y_2 + h_{22}h_{11} - h_{21}h_{12}} \\
\varphi_2^{-1}(\mathbf{y}) &= x_2 = \frac{-(h_{21} - h_{31}h_{23})y_1 + (h_{11} - h_{31}h_{13})y_2 - (h_{23}h_{11} - h_{21}h_{13})}{(h_{32}h_{21} - h_{31}h_{22})y_1 - (h_{32}h_{11} - h_{31}h_{12})y_2 + h_{22}h_{11} - h_{21}h_{12}} \\
|J_{\varphi^{-1}}(\mathbf{y})| &= \frac{|\mathbf{H}|^2}{((h_{32}h_{21} - h_{31}h_{22})y_1 - (h_{32}h_{11} - h_{31}h_{12})y_2 + h_{22}h_{11} - h_{21}h_{12})^3}
\end{aligned} \tag{10}$$

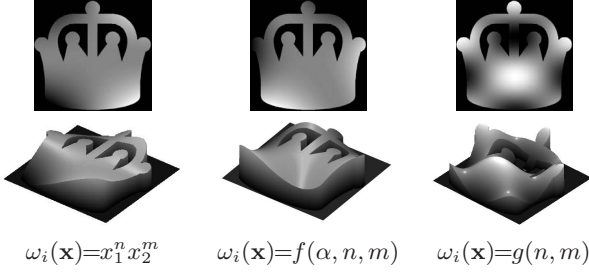


Figure 1. The effect of the ω_i functions. **Top**: the generated coloring of a binary shape for various functions, where $f(\alpha, n, m) = (x_1 \cos \alpha - x_2 \sin \alpha)^n (x_1 \sin \alpha + x_2 \cos \alpha)^m$ and $g(n, m) = \sin(n\pi x_1) \cos(m\pi x_2)$. **Bottom**: the corresponding volumes.

applied ω generates one equation. Let $\omega_i : \mathbb{R}^2 \rightarrow \mathbb{R}$ ($i = 1, \dots, \ell$) denote the set of adopted nonlinear functions. We thus obtain the following set of integral equations from Eq. (7) and Eq. (8)

$$\int_{\mathcal{F}_o} \omega_i(\mathbf{y}) d\mathbf{y} = \int_{\mathcal{F}_t} \omega_i(\varphi(\mathbf{x})) |J_{\varphi}(\mathbf{x})| d\mathbf{x}. \tag{11}$$

In order to solve for all unknowns, we need at least 8 equations, hence $\ell \geq 8$. Of course, since ω_i is also applied to the unknowns, the resulting equations will be nonlinear. Note however, that the generated equations contain no new information, they simply impose new linearly independent constraints. Indeed, from a geometric point of view, Eq. (7) simply matches the center of mass of the *template* and *observation* while the new equations of Eq. (11) match the volumes over the shapes constructed by the nonlinear functions ω_i (see Fig. 1). The parameters of the aligning planar homography are then simply obtained as the solution of the nonlinear system of equations in Eq. (11).

2.2. Numerical implementation

We have constructed our equations in the continuum but in practice we only have a limited precision digital image. This means that the integrals in Eq. (11) can only be *approximated* by a discrete sum over the foreground pixels introducing an inherent, although negligible error into our computation. The continuous domains \mathcal{F}_t and \mathcal{F}_o are represented as finite sets of foreground pixels denoted by

$\{\mathbf{X}^n\}_{n=1}^N$ and $\{\mathbf{Y}^m\}_{m=1}^M$, where N and M are the numbers of the foreground pixels of the *template* and *observation* respectively. Then Eq. (11) can be approximated by the following equation for $i = 1, \dots, \ell$

$$t_1 t_2 \sum_{m=1}^M \omega_i(\mathbf{Y}^m) = s_1 s_2 \sum_{n=1}^N \omega_i(\varphi(\mathbf{X}^n)) |J_{\varphi}(\mathbf{X}^n)| \tag{12}$$

Depending on the ω_i functions, the summed values can be highly varied causing numerical instability. One way to avoid that is normalizing image coordinates to an appropriate interval (e.g. $[-0.5, 0.5]$), and choosing ω_i with a limited range (e.g. $[-1, 1]$). Note that normalization involves a non-uniform scaling (s_1, s_2) and (t_1, t_2) taken into account in Eq. (12)–(13). Moreover, each equation can also be written in three alternative forms by making use of the inverse transformation $\varphi^{-1}(\mathbf{y})$ and the reverse integral transformation $\mathbf{x} = \varphi^{-1}(\mathbf{y})$, $d\mathbf{x} = |J_{\varphi^{-1}}(\mathbf{y})| d\mathbf{y}$:

$$\begin{aligned}
s_1 s_2 \sum_{n=1}^N \omega_i(\mathbf{X}^n) &= t_1 t_2 \sum_{m=1}^M \omega_i(\varphi^{-1}(\mathbf{Y}^m)) |J_{\varphi^{-1}}(\mathbf{Y}^m)| \\
s_1 s_2 \sum_{n=1}^N \omega_i(\mathbf{X}^n) |J_{\varphi}(\mathbf{X}^n)| &= t_1 t_2 \sum_{m=1}^M \omega_i(\varphi^{-1}(\mathbf{Y}^m)) \\
s_1 s_2 \sum_{n=1}^N \omega_i(\varphi(\mathbf{X}^n)) &= t_1 t_2 \sum_{m=1}^M \omega_i(\mathbf{Y}^m) |J_{\varphi^{-1}}(\mathbf{Y}^m)|,
\end{aligned} \tag{13}$$

where φ^{-1} and $|J_{\varphi^{-1}}|$ are defined in Eq. (10). Note that the above equations are equivalent to Eq. (12), including them into the system is redundant in the mathematical sense. However, in practice they play an important role in ensuring numerical stability of the final solution. Therefore in our implementation, we use all four equations for each ω_i .

In our experiments, we have empirically tested a number of potential $\{\omega_i\}$ sets and found the following set satisfactory for all test cases:

$$\omega_i(\mathbf{x}) = (x_1 \cos \alpha - x_2 \sin \alpha)^n (x_1 \sin \alpha + x_2 \cos \alpha)^m \tag{14}$$

$i = 1, \dots, 12$, using all combinations for $\alpha \in \{0, \frac{\pi}{6}, \frac{\pi}{3}\}$ and $(n, m) \in \{(1, 2), (2, 1), (1, 3), (3, 1)\}$.

It is clear that Eq. (12)–(13) are nonlinear. However it can be solved efficiently by *Levenberg-Marquardt algorithm* (an iterative least-squares minimization algorithm). Initial parameters can be set to the identity transformation (i.e. $h_{11} = h_{22} = h_{33} = 1$, the others equals to 0). Since in practice the *template* and the *observation* often have significantly different sizes, the solution can be improved by initializing scaling as $h_{11} = h_{22} = \sqrt{M/N}$. The solution of Eq. (12)–(13) gives the parameters of the transformation and the *registered* image is obtained by applying it to the *template*. See Algorithm 1.

Algorithm 1: Pseudo code of the proposed algorithm

Input : Binary images: *template* and *observation*

Output: Parameters of the aligning planar homography

- 1 Choose a set of $\omega_i : \mathbb{R}^2 \rightarrow \mathbb{R}$ ($i = 1, \dots, \ell \geq 8$)
 - 2 Find a least-squares solution of the system of equations Eq. (12)–(13) using *Levenberg-Marquardt* algorithm. Initialize the solver by $h_{11} = h_{22} = \sqrt{M/N}$, $h_{33} = 1$, and 0 elsewhere
 - 3 The solution gives the parameters of the transformation
-

3. Experimental results

The proposed method has been tested on a database of 37 different shapes and their transformed versions, a total of ≈ 1500 images of size 256×256 . The applied plane projective transformations of the normalized *template* (i.e. within $[-0.5, 0.5]$) were randomly composed of 0.5, \dots , 1.5 scalings; $-\frac{\pi}{4}, \dots, \frac{\pi}{4}$ rotations along the three axes; $-1, \dots, 1$ translations along both x and y axis and 0.5, \dots , 2.5 along the z axis; and a random focal length chosen from the $[0.5, 1.5]$ interval. Intuitively, these are projective transformations mapping a *template* shape from a plane placed in the three dimensional space to the xy plane. Some typical examples of these images can be seen in Fig. 2.

In order to quantitatively evaluate registration results, we defined two kind of error measures. The first one (denoted by ϵ) measures the distance between the true φ and the estimated $\hat{\varphi}$ transformation. Intuitively, ϵ shows the average transformation error per pixel. Another measure is the absolute difference (denoted by δ) between the *observation* and the *registered* image.

$$\epsilon = \frac{1}{N} \sum_{n=1}^N \|\varphi(\mathbf{X}^n) - \hat{\varphi}(\mathbf{X}^n)\|, \quad \delta = \frac{|R \Delta \{\mathbf{Y}^m\}_{m=1}^M|}{|R| + M},$$

where Δ is the symmetric difference and R denotes the set of pixels of the *registered* shape. A summary of these results

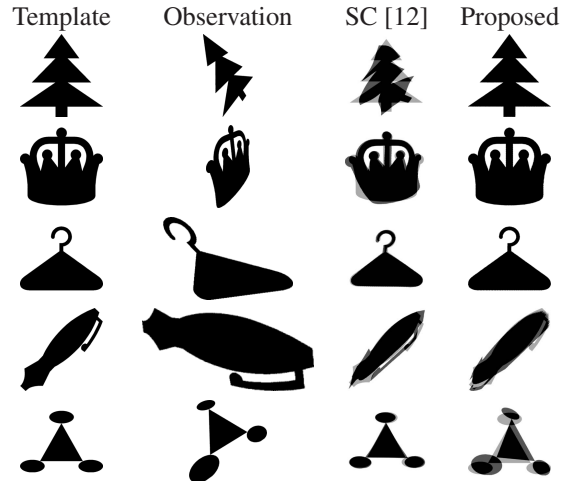


Figure 2. Example images from the synthetic data set and registration results. The third column shows the results obtained by *Shape Context* [12] while the last column contains the results of our method. The *template* and the registered *observation* were overlaid, overlapping pixels are depicted in black whereas non-overlapping ones are shown in gray.

	Runtime (sec.)		δ (%)		ϵ (pixel)	
	SC [12]	Prop.	SC [12]	Prop.	SC [12]	Prop.
m	98.72	26.38	2.69	0.11	-	0.10
μ	102.78	37.41	4.41	0.76	-	3.67
σ	28.26	49.01	4.79	3.72	-	21.91

Table 1. Test results and comparison of the proposed method to *Shape Context* [12] (m denotes the median while μ and σ denotes the mean and deviation of the values).

missing pixels (%)	5	15	20	25
δ (%)	1.36	4.00	5.38	6.72
ϵ (pixel)	1.08	3.29	4.42	5.40
size of occlusion (%)	1	2.5	5	10
δ (%)	1.46	3.36	5.62	9.86
ϵ (pixel)	1.98	4.57	7.97	13.90

Table 2. Median of error measures versus amount of randomly removed pixels (top) and size of random occlusions (bottom).

is presented in Table 1. We have also compared the performance of our method to that of *Shape Context* [12], a novel shape matching algorithm developed for nonlinear registration of binary images. For testing, we used the program provided by the authors, its parameters were set empirically to their optimal value ($\beta_{init} = 30$, $n_{iter} = 30$, annealing rate $r = 1$).

In practice, segmentation never produces perfect shapes. Therefore we have also evaluated the robustness of the proposed approach on the same dataset against two types of missing data. In the first test, we randomly removed

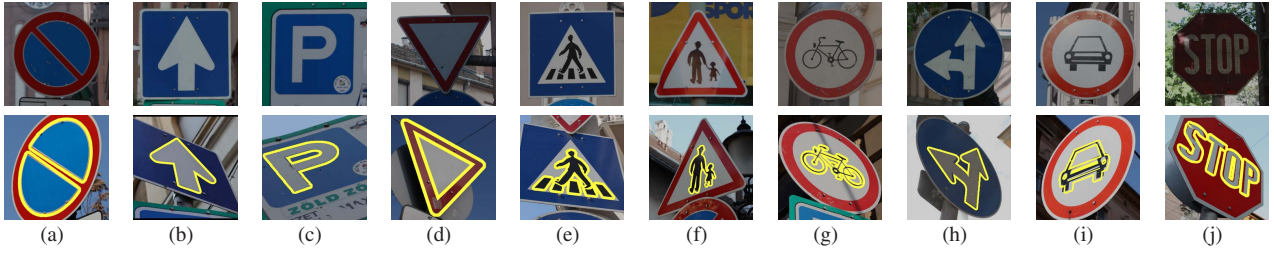


Figure 3. Registration results on traffic signs. The images used as *templates* are shown in the first row, and below them the corresponding *observations* with the overlaid contour of the registration results (the average δ error was 12.47% on these images).

5%, ..., 25% of the foreground pixels from the *observation* before registration, while in the second case we occluded continuous square shaped regions of size equal to 1%, ..., 10% of the shape. Table 2 shows the quantitative error measures versus the percentage of segmentation errors. Our method provides good results up to as high as 20% missing pixels and 5% occlusions.

3.1. Real images

The performance of our method has also been evaluated on real images in the context of two important application areas.

Traffic sign matching. Traffic sign detection and recognition is an important application area where one of the key tasks is the matching of a projectively distorted sign with a *template*. Herein, we have used classical thresholding and some morphological operations for segmentation but automatic detection/segmentation is also possible [11]. Each *template-observation* image pair contains pictures of different signboards. Fig. 3 shows some of our registration results. The main challenges were strong deformations, segmentation errors and variations in the style of the depicted objects. For example, the *observations* in Fig. 3(h)–3(j) do not contain exactly the same shape as the object on the image used as *template*. In particular, the *STOP* sign in Fig. 3(i) uses different fonts. In spite of these difficulties, our method was able to recover a quite accurate aligning transformation.

Registration of medical X-ray images. Hip replacement is a surgical procedure in which the hip joint is replaced by a prosthetic implant. In current practice, clinicians assess loosening of the implant by inspecting a number of post-operative X-ray images taken over a period of time. Obviously, such an analysis requires the registration of X-ray images as shown in Fig. 4. Since one is looking for deformations of the bone surrounding the implant, alignment must be based on the implant as it is the only imaged part which is guaranteed to remain unchanged from one image to the other. There are two main challenges here: One is

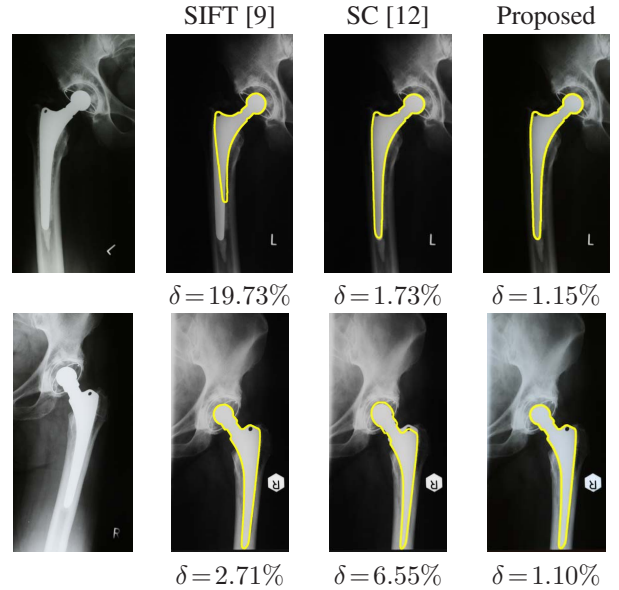


Figure 4. Registration results on hip prosthesis X-ray images. The overlaid contours show the aligned contours of the corresponding images on the left. Images in the second column show the registration results obtained by *SIFT* [9]+*homest* [8], in the third column the results of *Shape Context* [12]+*homest* [8], while the last column contains the results of the proposed method.

the highly non-linear radiometric distortion which makes any graylevel-based method unstable. Fortunately, the segmentation of the prosthetic implant is quite straightforward, herein we used active contours to segment the implant. The second problem is that the true transformation is not a plane projective one, it also depends on the position of the implant in 3D space. Indeed, there is a rigid-body transformation in 3D space between the implants, which becomes a projective mapping between the X-ray images. Fortunately, the planar homography assumption is a good approximation here as the X-ray images are taken in a well defined *standard position* of the patient’s leg. Some registration results are presented in Fig. 4. To achieve ultimate performance, we have ran additional tests to determine the optimal parameters for the function set Eq. (14) and found $\alpha \in \{0, \frac{\pi}{6}, \frac{\pi}{3}\}$ and $(n, m) \in \{(1, 2), (2, 1), (2, 3), (3, 2)\}$.



Figure 5. Registration results on traffic signs obtained by *SIFT* [9]+*homest* [8] (first row) and *Shape Context* [12]+*homest* [8] (second row). The contours of the registered images are overlaid.

Comparisons. Since the grayscale versions of the images were available, it was possible to compare our method to a feature-correspondence based solution. For that purpose, we have used *homest* [8], which implements a kind of “gold standard” algorithm composed of [15, 6]. The point correspondences has been extracted by the *SIFT* [9] method. As input, we provided the masked signboard region for traffic sign matching and the prosthesis region for medical registration. Furthermore, we have also extracted point correspondences established by *Shape Context* [12]. Here, the input was the binary mask itself used for *SIFT* as well as for our method. Although the *SIFT* parameter called *dist-Ratio*, controlling the number of the extracted correspondences, has been manually fine-tuned, we could not get reliable results due to the lack of rich radiometric features. Fig. 4 shows two results on X-ray images while on traffic signs (see Fig. 5), *SIFT* could not find enough correspondences in about half of the cases. As for *Shape Context*-based correspondences, we got somewhat better alignments (an average δ of 35.04% for the traffic sign images).

4. Discussion

Although our method is not moment-based, it is interesting to analyze how Eq. (11) is related to moments. Image moments and invariants were introduced by Hu [7] for 2D pattern analysis. Since then, they became one of the most popular region-based descriptors because any shape can be reconstructed from its infinite set of moments [4]. Traditional two dimensional ($p + q$)th order moments of a function $\rho : \mathbb{R}^2 \rightarrow \mathbb{R}$ are defined as $m_{pq} = \int_{\mathbb{R}^2} x_1^p x_2^q \rho(\mathbf{x}) d\mathbf{x}$, where $p, q \in \mathbb{N}_0$. When ρ is an image function then these moments are also referred to as *image moments*. In the binary case, where objects are represented by their silhouette, ρ is a characteristic function yielding $m_{pq} = \int_{\mathcal{F}} x_1^p x_2^q d\mathbf{x}$, where $\mathcal{F} = \{\mathbf{x} \in \mathbb{R}^2 : \rho(\mathbf{x}) = 1\}$. This is often called a *shape moment*. Generally, orthogonal moments, such as Legendre [4] or Zernike moments [13], are numerically more stable than regular moments. We remark, however,

that orthogonal moments can always be expressed by a finite set of regular moments.

In this sense, we can recognize a 0th order *function moment* of ω_i in the left hand side of Eq. (11) (just like any function integral can be regarded as the 0th order moment of the function itself). Shape moments, on the other hand, use only polynomials of the coordinates. Obviously, just like Legendre, or Zernike moments, our function moments could also be expressed in terms of such shape moments whenever the adopted ω_i functions are polynomial. When ω_i is not a polynome then it should be expressed in terms of its Taylor series, which again results in an approximating polynome which in turn yields an infinite sum of shape moments. The right hand side of Eq. (11) is more complex as it includes the product of the unknown transformation $\varphi(\mathbf{x})$ and its Jacobian determinant $|J_{\varphi}(\mathbf{x})|$ which are not polynomes. Therefore, independently of the choice of ω_i , it could be expressed in terms of shape moments only by expanding it into a Taylor series.

It is thus clear that our system of equations outlined in Eq. (11) cannot be rewritten in terms of a finite set of classical *shape moments*, and hence not even in terms of orthogonal moments. This result corresponds to similar findings reported in [14, 3] in the context of projective invariants. What we propose in this paper is another approach, which –starting from the identity relation in Eq. (3)– builds up an elegant method to generate an arbitrary set of linearly independent equations.

Invariance vs. covariance. Moment invariants [7, 14] are extensively studied as they provide a powerful tool for shape matching. There is a well established theory on affine invariants [2] and recently important results on the existence of projective moment invariants has also been reported [3]. Herein, we are not interested in constructing invariants (*i.e.* a function that is immune to the action of the projective group). All we need is sufficiently many linearly independent equations containing the parameters of the planar homography $\varphi(\mathbf{x})$. For that purpose, we need *covariant* functions (*i.e.* pairs of functions that vary with the actual transformation). Nevertheless, our framework also offers a way to define projective invariants. For that purpose, the Jacobian $|J_{\varphi}(\mathbf{x})|$ and the parameters h_{ij} of $\varphi(\mathbf{x})$ appearing in the integrand of the right hand side of Eq. (11) has to be eliminated. One way to achieve this is to follow the *triangle method* proposed in [3] and set $\omega_i(\mathbf{x})$ to be the integral of the appropriate power of the area of all triangles with one common vertex \mathbf{x} .

Choice of ω functions. Theoretically any nonlinear function (as long as it satisfies Eq. (8)) could be applied for constructing the system in Eq. (11). In practice, however the registration result depends on the set of ω_i , mainly

because in practice our equations are always corrupted by errors arising from imperfect data (*e.g.* segmentation and discretization errors). Since the solution is obtained via least-squares minimization, our function set should be unbiased with respect to error calculation. Overall error of the system in Eq. (11) is obtained as a squared sum of equations. Since both sides of these equations contain an integral of the corresponding ω_i function over the shape's domain, the characteristics of $\sum_i \omega_i^2(\mathbf{x})$ clearly influence the overall error. In fact, for a particular \mathbf{x} , this summation gives the contribution of a point \mathbf{x} to the integral and thus the measured squared error. In order to ensure an equal contribution from each pixel, $\sum_i \omega_i^2(\mathbf{x})$ should be constant for all \mathbf{x} . Basically, this also guarantees that the algebraic error, which is actually minimized, corresponds to the geometric error (up to scaling). We found empirically, that the squared sum of ω_i functions in Eq. (14) provides a flat surface which approximates well the above requirements. The function sets $\omega_i(\mathbf{x}) = x_1^n x_2^m$ and $\omega_i(\mathbf{x}) = \sin(n\pi x_1) \cos(m\pi x_2)$ (with $i = 1, \dots, 12$ and $(n, m) \in \{(1,2), (2,1), (2,2), (1,3), (3,1), (2,3), (3,2), (3,3), (1,4), (4,1), (2,4), (4,2)\}$) have also been considered (see Fig. 1). However, looking at the characteristics of the surface of the square-summed functions, it turns out that they do not satisfy our flatness criterion. As a result, they yield erroneous alignment. For example, experimental results on the traffic sign images of Fig. 3 show that the mean δ error were 14.08% for the first set and 19.28% for the second one; as opposed to 12.47% for the set in Eq. (14).

5. Conclusions

We have proposed a novel approach for planar shape alignment. The fundamental difference compared to classical image registration algorithms is that our model works without any landmark, feature detection, or correspondences by adopting a novel idea where the transformation is obtained as a solution of a set of nonlinear equations. Experimental results show that the proposed method provides good alignment on both real and synthetic images. Furthermore, its robustness has been demonstrated on a large synthetic dataset as well as on real images. Although our method clearly dominates state of the art correspondence-based methods, it has to be noted that, being calculated from the whole object, our equations are sensitive to partial occlusions. On the other hand, a common limitation of classical approaches is that they assume a deformation close to identity in order to establish reliable correspondences. Therefore we see our contribution as a complementary method rather than a replacement for all previous registration algorithms. Its superiority can be fully exploited in applications where occlusion can be kept at a minimum (*e.g.* medical imaging or industrial inspection), while feature-based methods can be more efficient when occlusions are common (*e.g.* surveillance). A rigorous theoretical analysis on how to select an optimal $\{\omega_i\}$ set is an important issue,

which will be the subject of future research.

References

- [1] C. Domokos, Z. Kato, and J. M. Francos. Parametric estimation of affine deformations of binary images. In *Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing*, pages 889–892, Las Vegas, Nevada, USA, April 2008.
- [2] J. Flusser and T. Suk. Pattern recognition by affine moment invariants. *Pattern Recognition*, 1:167–174, January 1993.
- [3] J. Flusser and T. Suk. Projective moment invariants. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(10):1364–1367, October 2004.
- [4] A. Foulonneau, P. Charbonnier, and F. Heitz. Multi-reference shape priors for active contours. *Int. Journal of Computer Vision*, 81(1):68–81, January 2009.
- [5] J. Francos, R. Hagege, and B. Friedlander. Estimation of multidimensional homeomorphisms for object recognition in noisy environments. In *Proc. of Conf. on Signals, Systems and Computers*, volume 2, pages 1615–1619, Pacific Grove, California, USA, November 2003.
- [6] R. Hartley. In defense of the eight-point algorithm. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(6):580–593, June 1997.
- [7] M.-K. Hu. Visual pattern recognition by moment invariants. *IRE Trans. on Information Theory*, 8(2):179–187, February 1962.
- [8] M. Lourakis. homest: A C/C++ library for robust, non-linear homography estimation. software, www.ics.forth.gr/~lourakis/homest, 2008.
- [9] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision*, 60(2):91–110, November 2004.
- [10] S. Mann and R. Picard. Video orbits of the projective group a simple approach to featureless estimation of parameters. *IEEE Trans. on Image Processing*, 6(9):1281–1295, September 1997.
- [11] C. F. Paulo and P. L. Correia. Automatic detection and classification of traffic signs. In *Proc. of Workshop on Image Analysis for Multimedia Interactive Services*, pages 11–14, Santorini, Greece, June 2007.
- [12] J. M. S. Belongie and J. Puzicha. Shape matching and object recognition using shape context. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(4):509–522, April 2002.
- [13] M. R. Teague. Image analysis via the general theory of moments. *Journal of the Optical Society of America*, 70:920–930, August 1980.
- [14] L. van Gool, T. Moons, E. Pauwels, and A. Oosterlinck. Vision and Lie's approach to invariance. *Image and Vision Computing*, 13:259–277, May 1995.
- [15] O. F. Z. Zhang, R. Deriche and Q. T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, 78(1-2):87–119, October 1995.
- [16] B. Zitová and J. Flusser. Image registration methods: A survey. *Image and Vision Computing*, 21(11):977–1000, October 2003.