Recovering Projective Transformations between Binary Shapes

Jozsef Nemeth

Department of Computer Algorithms and Artificial Intelligence, University of Szeged
H-6701 Szeged, PO. Box 652., Hungary, Fax: +36 62 546-397
nemjozs@inf.u-szeged.hu

Abstract. Binary image registration has been addressed by many authors recently however most of the proposed approaches are restricted to affine transformations. In this paper a novel approach is proposed to estimate the parameters of a general projective transformation (also called homography) that aligns two shapes. Recovering such projective transformations is a fundamental problem in computer vision with various applications. While classical approaches rely on established point correspondences the proposed solution does not need any feature extraction, it works only with the coordinates of the foreground pixels. The two-step method first estimates the perspective distortion independently of the affine part of the transformation which is recovered in the second step. As experiments on synthetic as well on real images show that the proposed method is less sensitive to the strength of the deformation than other solutions. The efficiency of the method has also been demonstrated on the traffic sign matching problem.

1 Introduction

In most of the image processing applications a key step is the registration of images *i.e.* the estimation of the transformation which aligns one image to the other (see [17] for a good survey). The overlapped images can be then combined or compared. The estimation of the parameters of a projective transformation (also known as planar homography) between two views of the same planar object has a fundamental importance in computer vision.

Classical landmark based (or correspondence based) methods usually trace back the problem into the solution of a system of linear equations set up using the coordinates of point pairs [7]. These point pairs are usually established by matching the intensity value patterns around the points [9]. On the other hand featureless methods estimate the transformation parameters directly from image intensity values over corresponding regions [10].

In many cases, however, the images do not contain sufficient variety of graylevel values (*e.g.* images of traffic signs or letterings), or suffered from intensity value distortions (*e.g.* X-ray images). Although there are some time consuming methods to cope with brightness change across image pairs [8], these conditions make the classical brightness-based methods unreliable. In [5], Francos *et al.* propose a method for the estimation of a homeomorphism between graylevel images. They showed how to

transform the problem into the solution of a linear system of equations, however they assumed that the intensity values differ only by a zero mean Gaussian noise.

When the segmentations are available it is reasonable to solve the registration problem using the binary versions of the images [13, 6]. Most of the current approaches are restricted to affine transformations. For example Domokos *et al.* showed that it is possible to trace back the affine matching problem to an exactly solvable polynomial system of equations [2]. Affine moments and invariants can also been used to recover linear transformation [15]. In [16] Yezzi *et al.* proposed a variational framework that uses active contours to simultaneously segment and register features from multiple images. Belongie *et al.* proposed a nonlinear shape matching algorithm in [1]. The method first establish point correspondences between the binary shapes using a novel similarity metric, called *shape context*, which consists in constructing a log-polar histogram of surrounding edge pixels. Then it uses the generic thin plate spline model to align the shapes.

In [11], Nemeth *et al.* proposed a method to estimate the parameters of projective transformations between binary shapes and later it has been extended by Domokos *et al.* [3] to more general nonlinear transformations *e.g.* polynomial and thin plate spline. This method has been proved to be efficient in case of many real applications *e.g.* matching handwritten characters or aligning multimodal prostate images. Although this approach proved to be very robust against the strength of the deformation, in some cases (*e.g.* when the shapes are rotated more than 90 degrees), it could not find the right solution due to the iterative minimization involved.

In this paper we propose a novel method to estimate the parameters of projective transformations between shapes. The perspective and the affine parts of the transformation are recovered in two sequential steps. It does not need any feature extraction or established correspondences, it works only with the coordinates of the foreground pixels. The performance of the method has been examined on synthetic as well as on real images.

2 The registration method

We are looking for a two dimensional projective transformation (also called *planar homography*) $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $\varphi(\mathbf{x}) = [\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x})]^T$ that aligns a pair of binary shapes, so that for any corresponding point pair $\mathbf{y} = [y_1, y_2]^T$ and $\mathbf{x} = [x_1, x_2]^T$ on the *template* and *observation* shapes:

$$\mathbf{y} = \varphi(\mathbf{x}). \quad (1)$$

A projective transformation in 2D is given by

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

where h_{ij} are the elements of the $\mathbf{H}_{3 \times 3}$ matrix (for more details on planar homography transformations see [7]). Since $\mathbf{H}$ is defined up to scale (it has only 8 degree of

freedom), one of its elements can be fixed (herein $h_{33} = 1$). Furthermore we represent the template and the observation shapes by their foreground regions $\mathcal{F}_t \subset \mathbb{R}^2$ and $\mathcal{F}_o \subset \mathbb{R}^2$. Thus we can simply write

$$\mathcal{F}_t = \varphi(\mathcal{F}_o). \quad (3)$$

The parameters h_{31} and h_{32} are responsible for the *perspective* distortion, while the others effect *affine* transformation (translation, rotation, scaling and shearing). These two parts of the transformation can be performed one after the other, so thus φ can be decomposed as follows:

$$\varphi = \varphi^a \circ \varphi^p \quad (4)$$

where $\varphi^p : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $\varphi^p(\mathbf{x}) = [\varphi_1^p(\mathbf{x}), \varphi_2^p(\mathbf{x})]^T$ is a nonlinear transformation:

$$\begin{aligned} \varphi_1^p(\mathbf{x}) &= \frac{x_1}{p_1 x_1 + p_2 x_2 + 1} \\ \varphi_2^p(\mathbf{x}) &= \frac{x_2}{p_1 x_1 + p_2 x_2 + 1}, \end{aligned} \quad (5)$$

resulting only perspective distortion. and $\varphi^a : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $\varphi^a(\mathbf{x}) = [\varphi_1^a(\mathbf{x}), \varphi_2^a(\mathbf{x})]^T$ is an affine transformation:

$$\begin{aligned} \varphi_1^a(\mathbf{x}) &= a_{11}x_1 + a_{12}x_2 + a_{13} \\ \varphi_2^a(\mathbf{x}) &= a_{21}x_1 + a_{22}x_2 + a_{23}, \end{aligned} \quad (6)$$

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

$$\mathcal{F}_t = (\varphi^a \circ \varphi^p)(\mathcal{F}_o) = \varphi^a(\varphi^p(\mathcal{F}_o)) \quad (7)$$

The proposed method estimates the p_i parameters of the perspective component φ^p and the a_i parameters of the affine component φ^a in two distinct steps, then using Eq. (4) we can get the h_{ij} parameters of φ as follows:

$$\begin{aligned} h_{11} &= a_{11} + p_1 a_{13}, \quad h_{12} = a_{12} + p_2 a_{13}, \\ h_{21} &= a_{21} + p_1 a_{23}, \quad h_{22} = a_{22} + p_2 a_{23}, \\ h_{13} &= a_{13}, \quad h_{23} = a_{23}, \quad h_{31} = p_1, \quad h_{32} = p_2 \end{aligned} \quad (8)$$

2.1 Step 1: Estimation of the perspective distortion

If Eq. (7) stands then there is only an affine transformation between $\mathcal{F}_t$ and $\varphi^p(\mathcal{F}_o)$, thus for any affine-invariant function $I : \mathbb{R}^2 \rightarrow \mathbb{R}$:

$$I(\mathcal{F}_t) = I(\varphi^p(\mathcal{F}_o)). \quad (9)$$

Note that the unknowns of this equation are the p_i parameters of φ^p . Moreover as we will show for given values of p_1 and p_2 using traditional moment-based affine invariants the right hand side of the equation can be efficiently estimated using only the jacobian

of φ^p , so it is not necessary to actually generate the image $\varphi^p(\mathcal{F}_o)$ which would be very time-consuming.

The basic idea of the proposed method is that given a set of independent affine invariant functions $I_i : \mathbb{R}^2 \rightarrow \mathbb{R}, i = 1 \dots n$ we obtain a system of equations:

$$I_i(\mathcal{F}_t) = I_i(\varphi^p(\mathcal{F}_o)). \quad (10)$$

The parameters of φ^p are obtained as the solution of this system of equations. It is clearly a highly nonlinear system and thus do not have exact solution. However as experimental results show it can be efficiently solved by a general nonlinear solver.

Although any set of affine invariant functions could be appropriate, herein we use affine moment invariants [4], because it allows a very efficient numerical estimation of the system of equations in Eq. (10). The left hand sides of the system of equations Eq. (10) do not depend on the parameters of φ^p so 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

$$m_{rs}(\mathcal{F}) = \int_{\mathcal{F}} x_1^r x_2^s d\mathbf{x}. \quad (11)$$

The affine moment invariants I_i of a shape are rely on the so called central moments that are defined as follows:

$$\mu_{rs}(\mathcal{F}) = \int_{\mathcal{F}} (x_1 - c_1)^r (x_2 - c_2)^s d\mathbf{x} \quad (12)$$

where the coordinates of the center of mass of the shape is given by using the geometric moments:

$$c_1 = \frac{m_{10}(\mathcal{F})}{m_{00}(\mathcal{F})} \quad \text{and} \quad c_2 = \frac{m_{01}(\mathcal{F})}{m_{00}(\mathcal{F})}. \quad (13)$$

The affine moment invariants $I_i(F)$ then are obtained using these central moments. For example the first two affine moment invariants are given as follows:

$$\begin{aligned} I_1 &= (\mu_{20}\mu_{02} - \mu_{11}^2)/\mu_{00}^4 \\ I_2 &= (-\mu_{30}^2\mu_{03}^2 + 6\mu_{30}\mu_{21}\mu_{12}\mu_{03} - 4\mu_{30}\mu_{12}^3 \\ &\quad - 4\mu_{21}^3\mu_{03} + 3\mu_{21}^2\mu_{12}^2)/\mu_{00}^{10}. \end{aligned} \quad (14)$$

For more on affine moment invariants see [4].

Given fixed parameters of φ^p we show how to compute the right hand side of the equations Eq. (10) avoiding the generation of the image $\varphi^p(\mathcal{F}_o)$ by making use of the Jacobian J_{φ^p} of the transformation. For a shape $\mathcal{F}$ that is distorted by φ^p the geometric moment can be estimated as follows:

$$m_{rs}(\varphi^p(\mathcal{F})) = \int_{\mathcal{F}} [\varphi_1^p(\mathbf{x})]^r [\varphi_2^p(\mathbf{x})]^s J_{\varphi^p}(\mathbf{x}) d\mathbf{x} \quad (15)$$

where the Jacobian of the perspective distortion is given by

$$J_{\varphi^p}(\mathbf{x}) = \frac{1}{(p_1x_1 + p_2x_2 + 1)^3},$$

(for more details on the usage of the Jacobian and mathematical derivation of such equations see [5] and [3]). On the perspectively distorted shape $\varphi^p(\mathcal{F})$ the central moments are given by

$$\mu_{rs}(\varphi^p(\mathcal{F})) = \int_{\mathcal{F}} [\varphi_1^p(\mathbf{x}) - c_1]^r [\varphi_2^p(\mathbf{x}) - c_2]^s J_{\varphi^p}(\mathbf{x}) d\mathbf{x}, \quad (16)$$

where

$$c_1 = \frac{m_{10}(\varphi^p(\mathcal{F}))}{m_{00}(\varphi^p(\mathcal{F}))} \quad \text{and} \quad c_2 = \frac{m_{01}(\varphi^p(\mathcal{F}))}{m_{00}(\varphi^p(\mathcal{F}))}. \quad (17)$$

For fixed values of the parameters p_1 and p_2 the affine moment invariants $I(\varphi^p(F))$ in the right hand side of the system Eq. (10) can be obtained using the central moments in Eq. (16) that can be estimated using only the foreground points of the shape $\mathcal{F}$. Thus we avoid to generate the $\varphi^p(\mathcal{F})$ images which would be very time consuming.

2.2 Step 2: Estimation of the affine transformation

After the perspective distortion is recovered (*i.e.* its parameters p_1 and p_2 are determined) the affine transformation φ^a should be estimated between shapes $\mathcal{F}_t$ and $\varphi^p(\mathcal{F}_o)$. For that purpose we used the approach proposed in [2] and, as in the previous section, in order to avoid the generation of $\varphi^p(\mathcal{F}_o)$ we modified the method by making use of the Jacobian J_{φ^p} of the perspective part. The following system of equations

$$\int_{F_t} y_k^n d\mathbf{y} = J_{\varphi^a} \sum_{i=1}^n \binom{n}{i} \sum_{j=0}^i \binom{i}{j} a_{k1}^{n-i} a_{k2}^{i-j} a_{k3}^j \int_{F_o} \varphi_1^p(\mathbf{x})^{n-i} \varphi_2^p(\mathbf{x})^{i-j} J_{\varphi^p}(\mathbf{x}) d\mathbf{x} \quad (18)$$

for $n = 1, 2, 3$ and $k = 1, 2$ can be written for the a_{ij} parameters of the affine transformation. This system contains six polynomial equations up to order three which is enough to solve for all 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:

$$J_{\varphi^a} = \frac{\int_{\mathcal{F}_t} d\mathbf{y}}{\int_{\mathcal{F}_o} J_{\varphi^p}(\mathbf{x}) d\mathbf{x}} \quad (19)$$

However the system Eq. (18) may have many solutions, we can select the real root which gives the same determinant as what we computed in Eq. (19). Note that the solution is not unique if the shape is affine symmetric.

Putting together the projective transformation φ^p and the affine transformation φ^a using Eq. (8) we get the h_{ij} parameters of the aligning planar homography transformation φ .

2.3 Implementational issues

Although the two unknowns of the perspective part φ^p would necessitate only two equations *i.e.* two invariants in system Eq. (10), the independence of the chosen set of affine

Algorithm 1: Pseudo code of the proposed algorithm

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

Output: The h_{ij} parameters of the aligning planar homography

- 1 Choose a set of affine invariants $I_i : \mathbb{R}^2 \rightarrow \mathbb{R}$ ($i = 1, \dots, n$)
 - 2 The solution of the system of equations Eq. (10) gives the perspective parameters p_1 and p_2
 - 3 Solve the system of equations Eq. (18) to determine the affine parameters a_{ij}
 - 4 Use Eq. (8) to obtain the h_{ij} parameters of the aligning planar homography
-

moment invariants is not guaranteed. Accordingly we obtained better results with over-determined systems. The results shown in Section 3 was attained using the $\{I_3, I_4, I_5, I_6\}$ set of invariants. Since the system is clearly nonlinear we found that it could be solved efficiently using nonlinear optimization methods. We used the standard differential evolution [14] method in the tests presented in Section 3. The equations are constructed in the continuum but in practice we only have a limited precision digital image. Consequently, the integrals over the continuous domains $\mathcal{F}_t$ and $\mathcal{F}_o$ can only be approximated by discrete sums over the set of foreground pixels denoted by F_t and F_o .

The pseudo code of the proposed algorithm can be found in Algorithm 1 while Fig. 1 shows example result images after the two subsequent steps of the method.

Fig. 1. 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.

3 Experiments

The performance of the proposed method has been tested on a large synthetic dataset consist of 35 different template shapes and their projectively distorted versions as observations a total of ≈ 1100 images of size 300×300 . The applied projective transformations were randomly composed of $0.5, \dots, 1.5$ scalings; $-\frac{\pi}{4}, \dots, \frac{\pi}{4}$ rotations along the x and y axes and $-\pi, \dots, \pi$ along the z axis; $-1, \dots, 1$ 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.

Registration results were quantitatively evaluated using two kind of error measures. The first one (δ) measures the percentage of the non-overlapping area of the template

Template	Observation	Shape Context	Domokos <i>et al.</i>	Proposed

Fig. 2. Example images from the synthetic data set and registration results obtained by Shape Context [1], Domokos *et al.* [3] and the proposed method. The template and the registered observation were overlaid, overlapping pixels are depicted in gray whereas nonoverlapping ones are shown in black.

and the registered observation while ϵ measures the distance (in pixels) between the true φ and the estimated $\hat{\varphi}$ transformation:

$$\delta = \frac{|F_r \Delta F_t|}{|F_r| + |F_t|} \cdot 100\%, \quad \epsilon = \frac{1}{|F_o|} \sum_{\mathbf{x} \in F_o} \|\varphi(\mathbf{x}) - \hat{\varphi}(\mathbf{x})\|,$$

where F_r denote the set of foreground pixels of the registered template respectively. The summary of the results is shown in Table 1.

For comparison we examined the results of Domokos *et al.* [3] on the same dataset. It provided weaker results since could not cope with cases when the images were rotated more than 90 degrees. We have also compared the performance of our method to that of *Shape Context* [1], a more general 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).

The average running time of the Domokos *et al.* and the proposed method (C implementations) were 4.75 sec. and 12.86 sec. respectively. The runtime of the Matlab implementation of Shape Context (68.87 sec.) is not authoritative.

	$\delta(\%)$			ϵ (pixel)		
	m	μ	σ	m	μ	σ
Shape Context	10.10	13.36	11.02	-	-	-
Domokos <i>et al.</i>	10.54	14.31	13.99	151.60	122.72	87.21
Proposed	2.38	6.23	12.34	1.88	23.46	53.00

Table 1. Test results on the synthetic dataset of Shape Context, Domokos *et al.* and the proposed method. m , μ , and σ denote the median, mean, and deviation.

3.1 Traffic signs

The recognition of traffic signs has an increasing importance in the car industry. The orientation of the sign is also important to decide whether it should be taken into consideration and it can be determined by registering the projectively distorted sign to a template. Herein we used classical thresholding however automatic detection and segmentation is also possible [12]. Fig. 3 shows some registration results of the proposed method. Recently, in [11] it has been shown that Shape Context[1] and SIFT[9] are unsuitable to extract point correspondences between traffic sign image pairs.

Fig. 3. Registration results on traffic signs. The images used as *observations* are shown in the first row, and below them the corresponding *templates* with the overlaid contours of the registration results.

4 Conclusion

We have proposed a novel approach to recover projective transformations of planar shapes. The two-step solution first estimates the perspective distortion independently

of the affine part of the transformation which is recovered in the second step. While classical image registration algorithms use feature correspondences our method works with invariant moments estimated using the point coordinates of the whole object. The efficiency of our method has been demonstrated on synthetic dataset as well as on real traffic sign images. Comparative test showed that our algorithm outperform other methods, especially in the case of strong deformations *e.g.* when the shapes are rotated more than 90 degrees. The examination of the robustness of the method against different level of segmentation errors will be the subject of future research.

Acknowledgements

This research was supported by the European Union and the European Regional Development Fund under the grant agreements TÁMOP-4.2.1/B-09/1/KONV- 2010-0005 and TÁMOP-4.2.2/B-10/1-2010-0012.

References

1. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape context. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(4), 509–522 (April 2002)
2. Domokos, C., Kato, Z.: Parametric estimation of affine deformations of planar shapes. *Pattern Recognition* 43(3), 569–578 (March 2010)
3. Domokos, C., Nemeth, J., Kato, Z.: Nonlinear shape registration without correspondences. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34(5), 943–958 (May 2012)
4. Flusser, J., Suk, T., Zitová, B.: *Moments and Moment Invariants in Pattern Recognition*. Wiley & Sons (October 2009)
5. Francos, J., Hagege, R., Friedlander, B.: Estimation of multidimensional homeomorphisms for object recognition in noisy environments. In: *Proceedings of Conference on Signals, Systems and Computers*. vol. 2, pp. 1615–1619. Pacific Grove, California, USA (November 2003)
6. Guo, H., Rangarajan, A., Joshi, S., Younes, L.: Non-rigid registration of shapes via diffeomorphic point matching. In: *Proceedings of International Symposium on Biomedical Imaging: From Nano to Macro*. vol. 1, pp. 924–927. IEEE, Arlington, VA, USA (April 2004)
7. Hartley, R.: In defense of the eight-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(6), 580–593 (June 1997)
8. Kaneko, S., Satohb, Y., Igarashi, S.: Using selective correlation coefficient for robust image registration. *Pattern Recognition* 36, 1165–1173 (May 2003)
9. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), 91–110 (November 2004)
10. Mann, S., Picard, R.W.: Video orbits of the projective group a simple approach to featureless estimation of parameters. *IEEE Transactions on Image Processing* 6(9), 1281–1295 (September 1997)
11. Nemeth, J., Domokos, C., Kato, Z.: Recovering planar homographies between 2d shapes. In: *Proceedings of International Conference on Computer Vision*. pp. 2170–2176. IEEE, IEEE, Cairo, Egypt (November 2009)
12. Paulo, C.F., Correia, P.L.: Automatic detection and classification of traffic signs. In: *Proc. of Workshop on Image Analysis for Multimedia Interactive Services*. pp. 11–14. Santorini, Greece (June 2007)

13. Simonson, K.M., Drescher, S.M., Tanner, F.R.: A statistics-based approach to binary image registration with uncertainty analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29, 112–125 (January 2007)
14. Storn, R., Price, K.: Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization* 11(4), 341–359 (December 1997)
15. Suk, T., Flusser, J.: Affine normalization of symmetric objects. In: *Proceedings of International Conference on Advanced Concepts for Intelligent Vision Systems*. vol. 3708, pp. 100–107. Springer, Antwerp, Belgium (September 2005)
16. Yezzi, A., Zöllei, L., Kapur, T.: A variational framework for joint segmentation and registration. In: *Proceedings of IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*. pp. 44–51. IEEE, Kauai, HI, USA (December 2001)
17. Zitová, B., Flusser, J.: Image registration methods: A survey. *Image and Vision Computing* 21(11), 977–1000 (October 2003)