Simultaneous Affine Registration of Multiple Shapes*

Csaba Domokos, Zoltan Kato

Image Processing and Computer Graphics Dept., University of Szeged H-6701 Szeged, PO. Box 652., Hungary

Abstract

The problem of simultaneously estimating affine deformations between multiple objects occur in many applications. Herein, a direct method is proposed which provides the result as a solution of a linear system of equations without establishing correspondences between the objects. The key idea is to construct enough linearly independent equations using covariant functions*, and then finding the solution simultaneously for all affine transformations. Quantitative evaluation confirms the performance of the method.*

1. Introduction

We consider the following general problem: Given a binary *template* image with a set of objects, and their affine distorted versions on the *observation* image, we want to establish the geometric correspondence between these images. The overall distortion is a global nonlinear transformation with the following constraint: the objects are distinct, *i.e.* either disconnected or separated by segmentation (*e.g.* an articulated object), each of them being subject to a *different* affine deformation. Such problems arise in various application domains like object matching or robot vision [2]. Classical *feature based* approaches consist in identifying point correspondences between the images using corners, lines crossing, control points, and then the solution of a system of equations, constructed using these point pairs, provides the parameters of the unknown transformation. On the other hand, *area-based* approaches make use of intensity correlation between image patches, where, in general, the transformation parameters are provided by an iterative optimization procedure. These methods are relying on rich discriminative and invariant radiometric information. Unfortunately, these requirements are difficult to satisfy in real applications due to low discriminative features (*e.g.* industrial parts with highly homogeneous surface intensities) or illumination variations. On the other hand, segmentation of these images is usually available in many applications thus shape registration is a valid alternative too. A closely related problem, mainly occuring in medical imaging [6, 1], is the approximation of a nonlinear deformation by piecewise affine transformations. In [6], the distortion is modeled as a locally affine but globally smooth transformation, which accounts for local and global variations in image intensities. The algorithm is built on a differential multiscale framework and incorporates the expectation maximization algorithm. In [1], a novel general framework (called *log-Euclidean polyaffine*) is proposed which provides a way of fusing local rigid or affine deformations into a global invertible transformation. However, these approaches have limited applicability on binary images as they are based on an optimization procedure, where the cost function is defined upon image intensity. In [5], an elegant linear solution is presented to solve a similar problem, where the segmentation of the *observation* into the distinct objects is avoided. This approach, however, makes use of intensity information hence its extension to binary images is far from trivial. Another possibility is to use affine moment invariants [7] to find the correspondences between objects. However, this would require to compute higher order (≥ 4) moments, which are numerically unstable.

Herein, we propose a linear solution based on constructing relations between the images [3], where the only available geometric information is used. The main contribution is the construction of these relations between shapes without establishing correspondences of any kind. The unknown transformation parameters are then obtained as the *least-squares* solution of an overdetermined linear system of equations.

2. Problem statement

Let us denote the *template* and *observation* points by $\mathbf{x} = [x_1, x_2]$ and $\mathbf{y} = [y_1, y_2] \in \mathbb{R}^2$, respectively. Furthermore, let $n \geq 2$ denote the number of shapes

*Partially supported by the grants CNK80370 of NIH & OTKA; the European Union and the European Regional Development Fund within the project TÁMOP-4.2.1/B-09/1/KONV-2010-0005.

$$\mathbf{C}:=\begin{bmatrix} \frac{\int_{\mathcal{D}_1} x_1\omega_1\left(P_1(\mathbf{x})\right)}{\sqrt{|\Sigma_1|}} & \frac{\int_{\mathcal{D}_1} x_2\omega_1\left(P_1(\mathbf{x})\right)}{\sqrt{|\Sigma_1|}} & \frac{\int_{\mathcal{D}_1} \omega_1\left(P_1(\mathbf{x})\right)}{\sqrt{|\Sigma_1|}} & \cdots & \frac{\int_{\mathcal{D}_n} x_1\omega_1\left(P_n(\mathbf{x})\right)}{\sqrt{|\Sigma_n|}} & \frac{\int_{\mathcal{D}_n} x_2\omega_1\left(P_n(\mathbf{x})\right)}{\sqrt{|\Sigma_n|}} & \frac{\int_{\mathcal{D}_n} \omega_1\left(P_n(\mathbf{x})\right)}{\sqrt{|\Sigma_n|}} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \frac{\int_{\mathcal{D}_1} x_1\omega_m\left(P_1(\mathbf{x})\right)}{\sqrt{|\Sigma_1|}} & \frac{\int_{\mathcal{D}_1} x_2\omega_m\left(P_1(\mathbf{x})\right)}{\sqrt{|\Sigma_1|}} & \frac{\int_{\mathcal{D}_1} \omega_m\left(P_1(\mathbf{x})\right)}{\sqrt{|\Sigma_1|}} & \cdots & \frac{\int_{\mathcal{D}_n} x_1\omega_m\left(P_n(\mathbf{x})\right)}{\sqrt{|\Sigma_n|}} & \frac{\int_{\mathcal{D}_n} x_2\omega_m\left(P_n(\mathbf{x})\right)}{\sqrt{|\Sigma_n|}} & \frac{\int_{\mathcal{D}_n} \omega_m\left(P_n(\mathbf{x})\right)}{\sqrt{|\Sigma_n|}} \end{bmatrix} \quad (1)$$

on the *template* and *observation*, related by n affine transformations. The i^{th} transformation, denoted by $(\mathbf{A}_i, \mathbf{t}_i) \in \mathbb{R}^{2\times 2} \times \mathbb{R}^{2\times 1}$, aligns the i^{th} *template object* with the corresponding object on the *observation*. Would these object-correspondences be known, a simple pairwise alignment could be easily recovered by any standard binary registration method like [4]. Establishing the correspondence between the objects is not trivial, however. Therefore we are interested in a direct solution without identifying corresponding object-pairs. The labeling of objects on the input images are given by the functions $\ell, \ell' : \mathbb{R}^2 \to \{0, 1, \ldots, n\}$.

Let us now consider the i^{th} object, where $\mathcal{D}_i = \{\mathbf{x}|\ell(\mathbf{x}) = i\}$ and $\mathcal{D}'_i = \{\mathbf{y}|\ell'(\mathbf{y}) = i\}$ denote the points of the i^{th} *template object* and its distorted *observation*, respectively. They are related as

$$\mathbf{A}_i\mathbf{x} + \mathbf{t}_i = \mathbf{y} \Leftrightarrow \mathbf{x} = \mathbf{A}_i^{-1}(\mathbf{y} - \mathbf{t}_i). \quad (2)$$

Furthermore, if we could observe image features that are invariant under the transformation $(\mathbf{A}_i, \mathbf{t}_i)$ (*e.g.* gray-levels [5]) then additional relations can be stated by defining a *covariant function* pair $P_i, S_i : \mathbb{R}^2 \to \mathbb{R}$ satisfying

$$P_i(\mathbf{x}) = P_i\big(\mathbf{A}_i^{-1}(\mathbf{y} - \mathbf{t}_i)\big) = S_i(\mathbf{y}). \quad (3)$$

Following [3], such functions can be constructed in the binary case using the Mahalanobis-distance (*i.e.* the first and second order statistics of the coordinates): $P_i(\mathbf{x}) = (\mathbf{x}-\mu_i)^T\Sigma_i^{-1}(\mathbf{x}-\mu_i)$ and $S_i(\mathbf{y}) = (\mathbf{y}-\mu'_i)^T\Sigma_i^{'-1}(\mathbf{y}-\mu'_i)$, where μ_i, μ'_i and Σ_i, Σ'_i are the means and covariance matrices of the i^{th} object and its distorted version, respectively. Notice that Eq. (3) remains valid, when an arbitrary $\omega : \mathbb{R} \to \mathbb{R}$ function acts on the both sides of the equation [3]:

$$\omega\big(P_i(\mathbf{x})\big) = \omega\big(S_i(\mathbf{y})\big). \quad (4)$$

Hence, applying linearly independent ω functions to the *covariant functions*, we can generate as many relations as needed. Finally, multiplying Eq. (2) and Eq. (4) and then integrating out individual point correspondences over the domain $\mathcal{D}_i$, we get [3]

$$\int_{\mathcal{D}_i} (\mathbf{A}_i\mathbf{x} + \mathbf{t}_i)\omega\big(P_i(\mathbf{x})\big)d\mathbf{x} = \frac{1}{|\mathbf{A}_i|}\int_{\mathcal{D}'_i} \mathbf{y}\omega\big(S_i(\mathbf{y})\big)d\mathbf{y}, \quad (5)$$

where the integral transformation $\mathbf{x} = \mathbf{A}_i^{-1}(\mathbf{y} - \mathbf{t}_i)$, $d\mathbf{x} = |\mathbf{A}_i|^{-1}d\mathbf{y}$ has been applied. Furthermore, as shown in [3], $\Sigma'_i = \mathbf{A}_i\Sigma_i\mathbf{A}_i^T$ yielding the formula $|\mathbf{A}_i| = \sqrt{|\Sigma'_i|}/\sqrt{|\Sigma_i|}$.

Now we know relations between the i^{th} object-pair, but we do not know the correspondence between the objects.

3. Direct solution

Based on Eq. (5), we can construct as many equations as needed by making use of a set of linearly independent functions $\{\omega_j\}_{j=1}^m$. However, since object pairs are not available, we have to sum all equations and solve the problem simultaneously, estimating all parameters in one system of equations. Thus Eq. (5) becomes

$$\sum_{i=1}^{n} \frac{1}{\sqrt{|\Sigma_i|}}\int_{\mathcal{D}_i} (\mathbf{A}_i\mathbf{x} + \mathbf{t}_i)\omega_j\big(P_i(\mathbf{x})\big)d\mathbf{x} = \sum_{i=1}^{n} \frac{1}{\sqrt{|\Sigma'_i|}}\int_{\mathcal{D}'_i} \mathbf{y}\omega_j\big(S_i(\mathbf{y})\big)d\mathbf{y} \quad (6)$$

For each ω_j, we have two equations: one for the first and another for the second point coordinate. Moreover, to solve for n affine transformations having $6n$ unknowns, we need at least $6n$ equations, hence $m \geq 3n$.

The terms $\int_{\mathcal{D}_i} (\mathbf{A}_{ik}\mathbf{x} + \mathbf{t}_{ik})\omega_j\big(P_i(\mathbf{x})\big)d\mathbf{x}$ on the left hand side of Eq. (6) can be rewritten for $k = 1, 2$ as

$$\sum_{l=1}^{2} a_{ikl}\int_{\mathcal{D}_i} x_l\omega_j\big(P_i(\mathbf{x})\big)d\mathbf{x} + t_{ik}\int_{\mathcal{D}_i} \omega_j\big(P_i(\mathbf{x})\big)d\mathbf{x},$$

where $\mathbf{A}_{ik}$ and $\mathbf{t}_{ik}$ denote the k^{th} row of $\mathbf{A}_i = (a_{ikl})_{2\times 2}$ and k^{th} component of $\mathbf{t}_i$, respectively. Hence Eq. (6) can be expressed in matrix form using the definition of the coefficient matrix $\mathbf{C}$ from Eq. (1)

$$\mathbf{C}\begin{bmatrix} a_{1k1} \\ a_{1k2} \\ t_{1k} \\ \vdots \\ a_{nk1} \\ a_{nk2} \\ t_{nk} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} \frac{\int_{\mathcal{D}'_i} y_k\omega_1\left(S_i(\mathbf{y})\right)}{\sqrt{|\Sigma'_i|}} \\ \vdots \\ \sum_{i=1}^{n} \frac{\int_{\mathcal{D}'_i} y_k\omega_m\left(S_i(\mathbf{y})\right)}{\sqrt{|\Sigma'_i|}} \end{bmatrix}, \quad k = 1, 2. \quad (7)$$

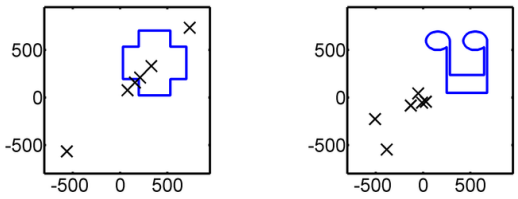

Figure 1. The contour of the objects (blue) and the points $\sqrt{|\mathbf{\Sigma}|^{-1}}\int \mathbf{x}\omega\big(P(\mathbf{x})\big)d\mathbf{x}$ in the case of a symmetric and non-symmetric objects.

Figure 2. Registration results on a synthetic image ($\delta = 5.59\%$). The *difference image* shows the overlapping areas in gray while misregistered areas in dark.

The solution of the above system provides all the unknown parameters of the overall deformation. If $m > 3n$ then the system is over-determined and a least squares solution is obtained. It is important to note, that the parameters of the Mahalanobis-distances μ_i, Σ_i and μ_i', Σ_i' can directly be computed from the input images and the coefficient matrix $\mathbf{C}$ needs to be computed only once! The size of $\mathbf{C}$ depends on the number of the objects. The time complexity of the proposed algorithm is thus $\mathcal{O}\big(n^2(M+N)\big)$, where M, N is the number of pixels of the *template* and *observation*. Since in practice $n \ll M + N$ the complexity is almost linear.

Let us now examine the structure of coefficient matrix $\mathbf{C}$: Each object generates 3 columns and each ω adds one row to $\mathbf{C}$ yielding a $m \times 3n$ matrix. The set of nonlinear $\{\omega\}_{j=1}^{m}$ functions can be arbitrary as long as they generate linearly independent equations. The more objects are on the images the method requires more ω functions. Intuitively, each ω generates a consistent coloring of the objects [3] and then Eq. (5) matches the center of mass of the objects with density given by ω. Obviously, different ωs generate different coloring, hence the center of mass of the "colored" object will also be different. Consequently, 3 numbers in the j^{th} row of $\mathbf{C}$, belonging to the j^{th} object, relate the center of mass coordinates for the given object after applying ω_j.

Since in practice we only have a limited precision digital image, the elements of the matrix $\mathbf{C}$ can only be *approximated* by a discrete sum over the foreground pixels introducing an inherent, although negligible error into our computation. However, these errors may accumulate in a larger system of equations causing less accurate registrations for an increasing number of objects. Nevertheless, such coarse registrations can be easily and efficiently refined by applying the affine registration method of [4].

The least-squares solution of the system Eq. (7) provides the best algebraic solution. When the system has a unique solution then this is also the geometrically correct solution. What happens if there are more geometrically correct solutions (*e.g.* there are two identical objects in the image or one of the objects is symmetric)? Eq. (7) becomes underdetermined and the best algebraic solution may be geometrically invalid. When we have two identical objects, then the unique solution is lost because the same three columns will appear twice in $\mathbf{C}$, which provide constraints just for one object. As for *symmetric* objects, since the applied *covariant functions* are inherently symmetric (they are based on the Mahalanobis-distance), the center of mass generated by various ω functions will be located on a single line (see Fig. 1) yielding only two constraints for the 3 unknowns.

4. Experimental results

The proposed method has been quantitatively evaluated on a large synthetic dataset containing 1500 images. The *templates* have 2,3 and 4 objects randomly chosen from a set of 14 different shapes. The *observations* were generated by applying a random transformation to each object composed of translation and skewing with parameters chosen from $[-100, 100]$ and $[-0.2, 0.2]$, respectively, along both axes; scaling with factors from $[0.5, 2]$; and and arbitrary rotation. A typical example can be seen in Fig. 2.

For the evaluation of registration results, we defined two kind of error measures: The first one (denoted by ϵ) measures the average distance between the true $(\mathbf{A}_i, \mathbf{t}_i)$ and the estimated $(\tilde{\mathbf{A}}_i, \tilde{\mathbf{t}}_i)$ transformation for all object. The second one is the absolute difference (denoted by δ) between the *observation* and the *registered* image:

$$\epsilon = \sum_{\substack{\mathbf{p}\in F_i \\ 1\le i\le n}} \frac{\|(\mathbf{A}_i - \tilde{\mathbf{A}}_i)\mathbf{p} + \mathbf{t}_i - \tilde{\mathbf{t}}_i\|}{|F|}, \delta = \frac{|\tilde{F} \triangle F'|}{|\tilde{F}| + |F'|} \cdot 100\%,$$

where $\triangle$ means the symmetric difference, while F, F' and $\tilde{F}$ denote the set of pixels of the *template*, *observation*, and the *registered* shape. Note that ϵ can only

#	Proposed method			After refining		
	δ (%)	ϵ (px)	time (s)	δ (%)	ϵ (px)	time (s)
2	2.2	3.67	4.65	0.04	0.06	7
3	6.45	11.01	6.07	0.05	0.06	9.07
4	14	24.29	7.18	0.05	0.06	10.78
Σ	6.55	13.8	6.07	0.05	0.06	9.08

Table 1. Median of error measures and runtime on images containing 2,3 and 4 objects.

be used when the true transformations are also known, while δ can always be computed.

The proposed method was implemented in Matlab and ran under Linux with 3GHz CPU and 3GB memory. In our experiments, we found that the $\{\omega\}$ set consisting of the identity function together with the trigonometric family provides satisfactory results: $\{x, \sin x, \cos x, \sin 2x, \cos 2x, \ldots, \sin lx, \cos lx\}$, where $l \in \mathbb{N}$ and $m = 3n + 2$ is fixed, *i.e.* the system is over-determined, hence the solution is obtained as a least-squares solution. Table 1 shows the registration results. Increasing the number of objects results in a larger coefficient matrix $\mathbf{C}$ yielding less accurate alignments. However, as noted in Section 3, the coarse registration results can be be easily refined by applying the method of [4]. It can be seen in Table 1, that independently of the number of the objects, we got excellent registration results after refining, at the price of a slight increase in computing time.

4.1. Real Images

In Fig. 3, we present some registration results on real images. The main challenges were segmentation errors (the objects are segmented via simple thresholding), symmetry and modelling error due to slight projective distortions. The results illustrate that the proposed method provides good results under real-life conditions. We note that there are many application of this problem in robot vision systems [2], *e.g.* when a robot should navigate or choose an object of unknown position. Although the second image contains symmetric objects, which may cause numerical instability in theory, our method provides good solution, since the symmetry is broken by segmentation errors in practice. Finally, the last image pair shows that our method can successfully recover *articulated deformations*.

5. Conclusion

We have proposed a novel approach to estimate multiple affine deformations between binary images. The main contribution is a direct solution without correspondences. We constructed a linear system of equations where relations between the images have been established using covariant functions defined on the only available geometric information of the input shapes. Quantitative evaluation on a large syntehtic dataset confirms the efficiency and accuracy of the proposed algorithm. The main advantages are that it does not require any correspondences or time consuming optimization step; it is fast and easy to implement while being insensitive to the strength of defomrations.

Figure 3. The registered contour of the images in the first row are overlayed on the *observation*s.

References

[1] V. Arsigny, O. Commowick, N. Ayache, and X. Pennec. A fast and log-euclidean polyaffine framework for locally linear registration. *Journal of Mathematical Imaging and Vision*, 33(2):222–238, Feb. 2009.

[2] F. Bonin-Font, A. Ortiz, and G. Oliver. Visual navigation for mobile robots: A survey. *Journal of Intelligent and Robotic Systems*, 53(3):263–296, Nov. 2008.

[3] C. Domokos and Z. Kato. Binary image registration using covariant Gaussian densities. In *Proc. of Int. Conf. on Image Analysis and Recognition*, volume 5112 of *LNCS*, pages 455–464, Póvoa de Varzim, Portugal, June 2008.

[4] C. Domokos and Z. Kato. Parametric estimation of affine deformations of planar shapes. *Pattern Recognition*, 43(3):569–578, Mar. 2010.

[5] R. R. Hagege and J. M. Francos. Estimation of affine geometric transformations of several objects from global measurements. In *Proc. of IEEE Int. Workshop on Multimedia Signal Processing*, pages 1–5, Rio de Janeiro, Brazil, Oct. 2009.

[6] S. Periaswamy and H. Farid. Medical image registration with partial data. *Medical Image Analysis*, 10(3):452–464, June 2006.

[7] T. Suk and J. Flusser. Graph method for generating affine moment invariants. In *Proc. of Int. Conf. on Pattern Recognition*, volume 2, pages 192–195, Cambridge, England, UK, Aug. 2004.