AFFINE ALIGNMENT OF COMPOUND OBJECTS: A DIRECT APPROACH

Csaba Domokos, Zoltan Kato

Image Processing and Computer Graphics Dept., University of Szeged H-6701 Szeged, PO. Box 652., Hungary, Fax: +36 62 546-397

ABSTRACT

A direct approach for parametric estimation of 2D affine deformations between compound shapes is proposed. It provides the result as a least-square solution of a linear system of equations. The basic idea is to fit Gaussian densities over the objects yielding covariant functions, which preserves the effect of the unknown transformation. Based on these functions, linear equations are constructed by integrating nonlinear functions over appropriate domains. The main advantages are: linear complexity, easy implementation, works without any time consuming optimization or established correspondences. Comparative tests show that it outperforms state-of-the-art methods both in terms of precision, robustness and complexity.

Index Terms— Compound shape, Affine registration, Gaussian distribution, Mahalanobis distance

1. INTRODUCTION

Image registration is an important step for various problems in computer vision, object matching and medical image processing. The basic problem is to find an unknown transformation between two images of the same scene [1]. Classical approaches fall into two main categories: *point-based methods* try to determine point correspondences between the images using interest point coordinates (*e.g.* corners, lines crossing, control points), then a system of equations is constructed using these point pairs. However, finding reliable correspondences is a hard problem in itself, which is further challenged by the lack of radiometric information in the binary case. Therefore most of the methods assume that the true distortion is close to identity. Finding an optimal aligning transformation usually requires an iterative procedure. Other type of approaches, called *area-based methods*, use a similarity measure (usually based on cross-correlation or mutual information) to estimating the transformation parameters provided by an optimization procedure. In this case, it is also important to restrict the admissible transformations as much as possible in order to speed-up the optimization procedure.

Recently, affine registration of binary shapes received more attention. The binary registration algorithm proposed in [2] constructs a system of equations by basically looking at the images at 3 different scales. Although the resulting system is linear, the solution is inherently less precise as in each equation they can only use part of the available information. In [3] a novel image normalization method is proposed with respect to unknown affine transformations based on affine image moments. The transformation is decomposed into basic transformations, which are subsequently eliminated one by one. In our previous work [4], we have proposed a method for estimating the parameters of an affine transformation by constructing covariant functions and then tracing back the original estimation problem to the solution of a linear system of equations. Unfortunately, the method requires an almost perfect segmentation in order to achieve satisfactory alignment.

In this paper, we show how this requirement can be relaxed in the case of compound shapes. An elegant, fast and robust solution is proposed, where linear equations are constructed by integrating nonlinear functions over corresponding domains derived from compound shapes. The performance of the proposed method has been tested on a large synthetic dataset as well as on real images.

2. ESTIMATION OF AFFINE DEFORMATIONS

We are interested in the general problem of finding an unknown affine transformation (denoted by $(\mathbf{A},\mathbf{t}) \in (\mathbb{R}^{2\times 2} \times \mathbb{R}^{2\times 1})$) aligning a *template* shape and its distorted *observation*, such that the following relation is satisfied:

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

where $\mathbf{x} = [x_1, x_2]^T$, $\mathbf{y} = [y_1, y_2]^T \in \mathbb{R}^2$ denote the coordinates of the *template* and *observation* points, respectively. If we can observe some image features (*e.g.* gray-levels of the pixels [5]) that are invariant under the transformation $(\mathbf{A},\mathbf{t})$, then an additional relation can be stated:

$$g(\mathbf{x}) = h(\mathbf{A}\mathbf{x}+\mathbf{t}) \quad \Leftrightarrow \quad g(\mathbf{A}^{-1}(\mathbf{y}-\mathbf{t})) = h(\mathbf{y}), \quad (2)$$

where $g, h : \mathbb{R}^2 \to \mathbb{R}$ are *covariant functions* under the transformation $(\mathbf{A},\mathbf{t})$, defined on those observed features. Furthermore, the above relations are still valid when an $\omega : \mathbb{R} \to \mathbb{R}$ function is acting on both sides of the equation [5]. Indeed, for a properly chosen ω

$$\omega(g(\mathbf{x})) = \omega(h(\mathbf{A}\mathbf{x}+\mathbf{t})) \Leftrightarrow \omega\big(g(\mathbf{A}^{-1}(\mathbf{y}-\mathbf{t}))\big) = \omega(h(\mathbf{y})).$$

This research was partially supported by the Hungarian Scientific Research Fund (OTKA) – K75637

The basic idea of [4, 5] is to use nonlinear ω functions. This way, we can generate as many linearly independent equations as needed. Note that these equations doesn't contain new information, they simply impose new linearly independent constraints allowing for a unique solution.

2.1. Construction of Covariant Functions

The key challenge is to construct appropriate covariant functions satisfying Eq. (2). Since we do not have any radiometric information, these functions must be defined using the only available geometric information. The solution proposed in our previous work [4] relies on fitting Gaussian densities over the *template* and *observation* shapes. We thus consider the points of the *template* as a sample from a bivariate normally distributed random variable $X \sim N(\mu, \Sigma)$ with

$$p(\mathbf{x}) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\Big(-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\Big) \quad (3)$$

probability density function (PDF). For any affine transformation $Y = \mathbf{A}X + \mathbf{t}$ is also normally distributed with $Y \sim N(\mu', \Sigma') = N(\mathbf{A}\mu + \mathbf{t}, \mathbf{A}\Sigma\mathbf{A}^T)$ and $s(\mathbf{y})$ PDF. It is then easy to see that $p(\mathbf{x})$ and $s(\mathbf{y})$ are covariant [4]:

$$p(\mathbf{x}) = |\mathbf{A}|s(\mathbf{y}), \quad (4)$$

where $|\mathbf{A}| = \sqrt{|\Sigma'|/|\Sigma|}$ is the Jacobian of the transformation. Without loss of generality, we can assume that $|\mathbf{A}| > 0$, since $\mathbf{A}$ is clearly non-singular, and a negative determinant would mean flipping of coordinates which is usually excluded by physical constraints in real applications.

It is well known, that the normalizing constant $1/(2\pi\sqrt{|\Sigma|})$ in Eq. (3) ensures that the integral of the PDF evaluates to 1. It is also the maximum value of the density function, which is inversely proportional to the area of the shape. This dependence on the shape size may cause numerical instabilities hence we rewrite Eq. (5) in the following form:

$$\begin{aligned} \mathcal{P}(\mathbf{x}) \equiv 2\pi\sqrt{|\Sigma|}p(\mathbf{x}) &= 2\pi\sqrt{|\Sigma'|}s(\mathbf{y}) \equiv \mathcal{S}(\mathbf{y}), \text{with} \quad (5)\\ \mathcal{P}(\mathbf{x}) &= \exp\Big(-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\Big). \end{aligned}$$

2.2. Compound Objects

Image analysis often deals with the matching of objects composed of several parts, yielding a group of disjoint shapes when segmented. The topology of such compound shapes will not change under the action of the affine group. Thus, assuming the *template* object consists of $m \geq 2$ disjoint shapes, each component has exactly one corresponding shape on the *observation*, *i.e.* there exist a bijective mapping between the *template* and *observation* components under the transformation ($\mathbf{A}$,$\mathbf{t}$). As a consequence, we can construct covariant functions $\mathcal{P}_i(\mathbf{x})$, $\mathcal{S}_i(\mathbf{y})$ for each pair of shapes satisfying Eq. (5). Furthermore, the overall shape (*i.e.* the whole foreground region) also gives rise to a pair of covariant functions $\mathcal{P}_0(\mathbf{x})$, $\mathcal{S}_0(\mathbf{y})$. Thus we have $m+1$ relations. If the correspondence between components would be known then we could simply construct a system of $m+1$ equations and solve for the unknowns. As such a matching is usually not known, we will sum these relations yielding

$$P(\mathbf{x}) \equiv \sum_{i=0}^{m}\mathcal{P}_i(\mathbf{x}) = \sum_{i=0}^{m}\mathcal{S}_i(\mathbf{y}) \equiv S(\mathbf{y}). \quad (6)$$

Note that these sums are mixtures of unnormalized Gaussian densities which can also be interpreted as a consistent coloring of the *template* and *observation* respectively (see Fig. 1). By consistent coloring, we mean that these functions preserve the effect of the unknown transformation. Furthermore, these functions can be constructed exactly and uniquely from the object points alone without any knowledge about the aligning transformation. As a result, we can transform the original binary images into graylevel ones, where corresponding pixels have exactly the same gray value.

We thus have Eq. (1) and Eq. (6) as the only relations between the two objects. Since nothing is known about the correspondence between the *template* and *observation*, we will multiply these equations and integrate out individual point matches [5, 4]:

$$\int_{\mathbb{R}^2}\mathbf{x}\sum_{i=0}^{m}\mathcal{P}_i(\mathbf{x})d\mathbf{x} = |\mathbf{A}|^{-1}\int_{\mathbb{R}^2}\mathbf{A}^{-1}(\mathbf{y}-\mathbf{t})\sum_{i=0}^{m}\mathcal{S}_i(\mathbf{y})d\mathbf{y}, \quad (7)$$

where we have used the integral transformation $\mathbf{x} = \mathbf{A}^{-1}(\mathbf{y}-\mathbf{t})$, $d\mathbf{x} = |\mathbf{A}|^{-1}d\mathbf{y}$.

2.3. Selecting the Integration Domains

One obvious approach to restrict the integration in Eq. (7) to a finite domain is to use the segmented shape itself as the domain [5, 4]. However, a clear disadvantage of this approach is that any segmentation error will inherently result in erroneous integrals causing misalignment. Furthermore, even if the segmentation is perfect, the precision of these domains is always compromised by discretization error. Herein, we will take another approach to select appropriate domains $\mathcal{D}$ and $\mathcal{D}'$ satisfying the following properties: **1)** they are related by the unknown transformation $\mathbf{A}\mathcal{D} + \mathbf{t} = \mathcal{D}'$ and **2)** the integrands are rich enough (*i.e.* has a characteristic pattern) within the selected regions. Since the equidensity contours of a bivariate Gaussian PDF are ellipsoids centered at the mean, it is natural to chose one of the ellipses of the density fitted to the overall shape as the domain $\mathcal{D}$. To satisfy 2), we may choose an ellipse according to the *two sigma rule*, which guarantees that about 95% of values are within the enclosed ellipsoid. To satisfy 1), we have to select a pair of corresponding ellipses from the *template* and *observation*. This can be obtained by selecting points whose Mahalanobis distance are less than r^2 from μ: $\mathcal{D} = \{\mathbf{x} \in \mathbb{R}^2 | (\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu) \leq r^2\}$ and

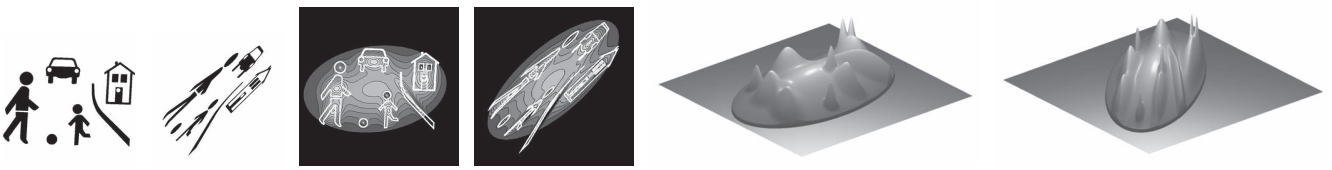

Fig. 1. Gaussian PDFs fitted over a binary shape yields consistent coloring. **First pair**: *template* and *observation* compound object. **Second pair**: Isovalue lines show $P(\mathbf{x})$ and $S(\mathbf{y})$ over the selected elliptic domains. The white curves show the objects' contour. **Last pair**: 3D plot of $P(\mathbf{x})$ and $S(\mathbf{y})$.

$\mathcal{D}' = \{\mathbf{y} \in \mathbb{R}^2 | (\mathbf{y}-\mu')^T \Sigma'^{-1} (\mathbf{y}-\mu') \leq r^2\}$ (see Fig. 1). These domains are analytical and less dependent on segmentation errors, thus the integration in Eq. (7) can be restricted to these finite domains. Experiments show that good results can be achieved by settings ranging from $r = 1$ to $r = 3$.

2.4. Linear Solution

Now we can adopt the idea from [5, 4] and generate a set of linear equations to solve for the unknown transformation $(\mathbf{A}, \mathbf{t})$. The basic idea is to apply nonlinear $\omega : \mathbb{R} \to \mathbb{R}$ functions to both sides of Eq. (6) yielding linearly independent equations:

$$\int_{\mathcal{D}} \mathbf{x}\omega\big(P(\mathbf{x})\big)d\mathbf{x} = |\mathbf{A}|^{-1} \int_{\mathcal{D}'} \mathbf{A}^{-1}(\mathbf{y}-\mathbf{t})\omega\big(S(\mathbf{y})\big)d\mathbf{y}.$$

This way, we can generate as many linearly independent equations as needed (at least 6 as we have 6 unknowns for $(\mathbf{A}, \mathbf{t})$). If q_{ki} denotes the elements of $\mathbf{A}^{-1}$ (for $i = 1, 2$) and $-\mathbf{A}^{-1}\mathbf{t}$ (for $i = 3$); and by adopting a set of functions $\{\omega_i\}_{i=1}^n$, the above linear system can be written in matrix form [5, 4]:

$$\begin{bmatrix} \int_{\mathcal{D}'} y_1\omega_1\big(S(\mathbf{y})\big) & \int_{\mathcal{D}'} y_2\omega_1\big(S(\mathbf{y})\big) & \int_{\mathcal{D}'} \omega_1\big(S(\mathbf{y})\big) \\ \vdots & \vdots & \vdots \\ \int_{\mathcal{D}'} y_1\omega_n\big(S(\mathbf{y})\big) & \int_{\mathcal{D}'} y_2\omega_n\big(S(\mathbf{y})\big) & \int_{\mathcal{D}'} \omega_n\big(S(\mathbf{y})\big) \end{bmatrix} \times$$

$$\begin{bmatrix} q_{k1} \\ q_{k2} \\ q_{k3} \end{bmatrix} = |\mathbf{A}| \begin{bmatrix} \int_{\mathcal{D}} x_k\omega_1\big(P(\mathbf{x})\big) \\ \vdots \\ \int_{\mathcal{D}} x_k\omega_n\big(P(\mathbf{x})\big) \end{bmatrix}, \quad k = 1, 2. \quad (8)$$

The solution of this linear system provides the parameters of the transformation. If $n > 3$ then the system is overdetermined and the solution is obtained as a least squares solution. The parameters of the probability densities $N(\mu, \Sigma)$ and $N(\mu', \Sigma')$ can directly be computed from the images and this is all what is needed for our method. Once these parameters are computed, we can construct covariant functions and select appropriate integration domains. A clear advantage of the proposed approach is that theoretically, the most accurate and less sensitive feature of the objects are their statistics, *i.e.* their sample means and covariances. Although we should compute the integrals over an elliptic domain, in practice, we can easily approximate these integrals by finite sums

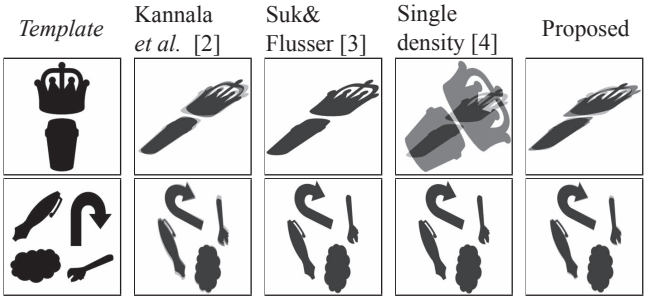

Fig. 2. Registration results: Overlapping areas are shown in dark while misregistered areas in gray.

over a grid with sufficient resolution. In our experiments an 1500×1500 grid gave satisfactory results. The set of nonlinear ω functions can be arbitrary as long as they generate linearly independent equations. In our experiments we found that the trigonometric family provides satisfactory results: $\{x, \sin x, \cos x, \sin 2x, \cos 2x\}$.

3. EXPERIMENTAL RESULTS

The proposed method has been quantitatively evaluated on a set of $\approx$ 1500 synthetically generated observations for 60 different compound shapes. The applied transformations were randomly composed of $0°, 10°, \ldots, 350°$ rotations; $0, 0.4, \ldots, 1.2$ shearings; $0.5, 0.7, \ldots, 1.9$ scalings, and $-20, 0, 20$ translations along both axes. The resulting images are of size $\approx 1400 \times 1400$, some typical examples 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 $(\mathbf{A}, \mathbf{t})$ and the estimated $(\tilde{\mathbf{A}}, \tilde{\mathbf{t}})$ transformation. The second one is the absolute difference (denoted by δ) between the *observation* and the *registered* image:

$$\epsilon = \frac{1}{|F|} \sum_{\mathbf{p} \in F} \|(\mathbf{A} - \tilde{\mathbf{A}})\mathbf{p}\|, \quad \text{and} \quad \delta = \frac{|\tilde{F} \triangle F'|}{|\tilde{F}| + |F'|} \cdot 100\%,$$

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.

Method and runtime		0%	10%	50%	90%
Kannala *et al.* [2]	ϵ	2.7	3.65	7.37	26.44
29.79 sec.	δ	1.65	2.35	4.77	14.39
Suk&Flusser [3]	ϵ	0.43	19.25	109.85	258.34
3.91 sec.	δ	0.06	9.91	51.16	92.11
Single density [4]	ϵ	0.64	82.65	407.31	748.95
0.48 sec.	δ	0.31	35.11	84.01	100
Proposed	ϵ	0.58	2.64	7.7	23.66
4.65 sec.	δ	0.25	1.55	4.59	13.23

Table 1. Median of error measures and runtime.

In practice, segmentation never produces perfect shapes. Therefore we also evaluated the robustness of the proposed approach when $10\%, 20\%, \ldots, 90\%$ of the foreground pixels has been removed from the *observation* before registration. Evaluation results are summarized in Table 1 and some registration results are shown in Fig. 2. It is clear that our method provides good results up to as high as 50% removed pixels, and results for 90% are also acceptable. In general, our method will perform well as long as the first and second order statistics of shapes does not change dramatically. The proposed approach has also been compared to some recent binary registration approaches [2, 3], including our previous method using a single density [4]. The method of Kannala *et al.* [2] is clearly outperformed in both quality, robustness and computing time. On the other hand, the method of Suk and Flusser [3] achieves slightly better results. In terms of robustness, however, our method clearly dominates [3], which already performs poorly for 10% removed pixels and fails completely for 50%. Finally, our previous method [4] performs quite well when there are no segmentation errors, but it also fails for 50% missing pixels.

In Fig. 3, we present some registration results on real images of traffic signs. Automatic traffic sign recognition [6] is a hot topic, which usually requires the registration of a reference shape to an observed one. There are many methods to automatically detect and segment signs [6], but in our experiments we simply performed this task manually via thresholding. The results illustrate that the proposed method provides good results under real-life conditions.

4. CONCLUSION

In this paper we proposed a novel approach to estimate 2D affine transformations between compound shapes. The main novelty of our method is that, in contrast to our previous work [4], this approach selects consistent domains under the unknown affine transformation instead of the segmentation. Therefore it is more robust against segmentation errors. It makes use of all available information in the input images and provides the transformation parameters as a solution of a linear system of equations. It works without established point pairs or optimization step; furthermore it has a linear time complexity and is easy to implement.

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

5. REFERENCES

[1] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, October 2003.

[2] J. Kannala, E. Rahtu, J. Heikkilä, and M. Salo, "A new method for affine registration of images and point sets," in *Proc. of Scandinavian Conference on Image Analysis*, Joensuu, Finland, June 2005, vol. 3540 of *LNCS*, pp. 224–234, Springer.

[3] T. Suk and J. Flusser, "Affine normalization of symmetric objects," in *Proc. of Int. Conf. on Advanced Concepts for Intelligent Vision Systems*, Antwerp, Belgium, September 2005, vol. 3708 of *LNCS*, pp. 100–107, Springer.

[4] C. Domokos and Z. Kato, "Binary image registration using covariant Gaussian densities," in *Proc. of Int. Conf. on Image Analysis and Recognition*, Póvoa de Varzim, Portugal, June 2008, vol. 5112 of *LNCS*, pp. 455–464, Springer.

[5] R. Hagege and J. M. Francos, "Parametric estimation of multi-dimensional affine transformations: An exact linear solution," in *Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing*, Philadelphia, USA, March 2005, IEEE, vol. 2, pp. 861–864.

[6] C. F. Paulo and P. L. Correia, "Automatic detection and classification of traffic signs," in *Proc. of Int. Workshop on Image Analysis for Multimedia Interactive Services*, Santorini, Greece, June 2007, IEEE, pp. 11–11.