4.4 An example of landmark-based registration

Figure 4.4 displays results for the landmark-based registration of reference and template images which have already been shown in Fig. 4.1. As is apparent from this figure, the registration ranges from an interpolation of the landmarks ($\alpha = 0$) to almost affine linear registration for large values of α. Note that the registration is governed completely by the landmarks.

Finally, Fig. 4.5 illustrates that landmark-based registration does not always result in a meaningful registration. The landmarks are chosen such that one expects a bending of the rectangular bar displayed in the template image. Note that the landmarks are chosen in a meaningful ordering. However, although the transformation is smooth, it fails to be diffeomorphic. This can be seen in the bottom right picture, where the sign of the Jacobian of the transformation, i.e., $\operatorname{sign}(\det \nabla\varphi)$, is shown.

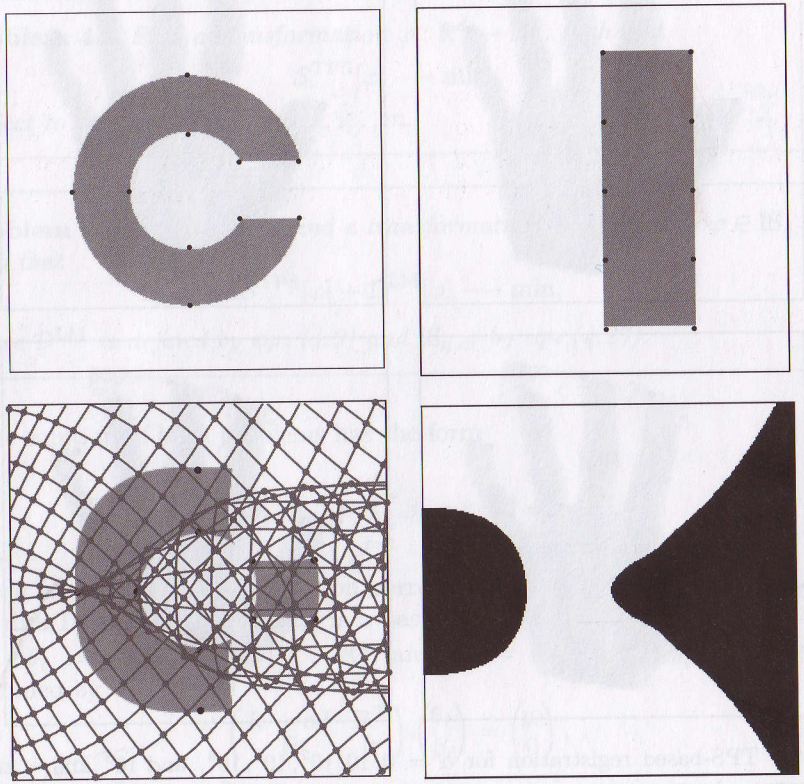

FIG. 4.5 TOP LEFT: reference image with landmarks (dots), TOP RIGHT: template image with landmarks (dots), BOTTOM LEFT: template image after TPS registration with landmarks and deformed grid, BOTTOM RIGHT: sign of determinant of Jacobian of φ (white: $\det \nabla\varphi \geq 0$, black: $\det \nabla\varphi < 0$).

5

PRINCIPAL AXES-BASED REGISTRATION

The registration technique introduced in Chapter 4 has a major drawback. That is, the registration process is governed by the location and correspondence of the landmarks.

Although there are many sophisticated ideas for automatically locating the landmarks (see, e.g., Rohr (2001) and references therein) the process is still not fully automated. Moreover, the location of landmarks might be very complicated, since even experts are not always able to characterize mathematically, for example, anatomical landmarks in medical images.

In this chapter we follow the idea of registering image features, but now the registration is based on features which can be deduced from the images automatically. To this end, we consider an image as a density function or mass distribution. From this distribution we derive the so-called *principal axes*. The registration is also called *principal axes transformation* (PAT). For the relevant literature, we refer to Maurer & Fitzpatrick (1993) and references therein, and particularly Alpert et al (1990).

Definition 5.1 *Let $d \in \mathbb{N}$ and $B : \mathbb{R}^d \to \mathbb{R}$ be an image; cf., Definition 3.1. We define the* expectation value *of a function f with respect to B by*

$$\mathbb{E}_B[f] := \frac{\int_{\mathbb{R}^d} f(x)\, B(x) dx}{\int_{\mathbb{R}^d} B(x) dx}.$$

For $u : \mathbb{R}^d \to \mathbb{R}^{m\times n}$, we set $\mathbb{E}_B[u] := (\mathbb{E}_B[u_{j,k}])_{\substack{j=1,\ldots,m\\ k=1,\ldots,n}} \in \mathbb{R}^{m\times n}$.

The center *of an image is defined by $c_B := \mathbb{E}_B[x] \in \mathbb{R}^d$ and the* covariance *by $\operatorname{Cov}_B := \mathbb{E}_B\left[(x - c_B)(x - c_B)^\top\right] \in \mathbb{R}^{d\times d}$.*

The center and an eigendecomposition of the covariance matrix are used as features for the image. The resulting registration technique is named principal axes transformation (PAT); see, e.g., Alpert et al (1990).

Since the covariance matrix is real, symmetric, and positive semi-definite, it permits an orthogonal eigendecomposition

$$\operatorname{Cov}_B = D_B \Sigma_B^2 D_B^\top, \tag{5.1}$$

where $D_B \in \mathbb{R}^{d\times d}$ is unitary, i.e., $D_B^\top D_B = I_d$, and the matrix

$$\Sigma_B = \operatorname{diag}(\sigma_{B,1}, \dots, \sigma_{B,d}) \in \mathbb{R}^{d\times d}$$

is diagonal; cf., e.g., Golub & van Loan (1989). For normalization purposes, we arrange the columns of D_B such that for the *standard deviations* we have $\sigma_{B,1} \geq \cdots \geq \sigma_{B,d} \geq 0$. If the eigenvalues of Cov_B are simple, the decomposition is essentially unique, up to a sign in the columns of D_B.

5.1 Geometrical interpretation

Suppose we are looking for an optimal representation of the mass distribution B by a straight line $G = p + \mathbb{R}y$ (the *principal axis*), where $\|y\|_{\mathbb{R}^d} = 1$. The distance of an arbitrary point $x \in \mathbb{R}^d$ to this line is given by

$$d_{p,y}(x) := \|x - p - \langle x - p, y\rangle_{\mathbb{R}^d}\, y\|_{\mathbb{R}^d}$$

and the optimal line can be derived from the minimization of

$$D(p, y) \longrightarrow \min,$$

where

$$D(p,y) = \mathbb{E}_B\left[(d_{p,y}(x))^2\right] = \mathbb{E}_B\left[\langle x-p, x-p\rangle_{\mathbb{R}^d} - \langle x-p, y\rangle^2_{\mathbb{R}^d}\right].$$

Necessary conditions for a minimizer (p^*, y^*) are

$$\begin{aligned} 0 &= \nabla_p D(p^*, y^*) \\ &= \mathbb{E}_B\left[-2(x-p^*) + 2\langle x - p^*, y^*\rangle_{\mathbb{R}^d}\, y^*\right] \\ &= -2(c_B - p^*) + 2\langle c_B - p^*, y^*\rangle_{\mathbb{R}^d}\, y^* \end{aligned}$$

and

$$\begin{aligned} 0 &= \nabla_y D(p^*, y^*) \\ &= \mathbb{E}_B\left[-2\langle x - p^*, y^*\rangle_{\mathbb{R}^d}\,(x - p^*)\right] \\ &= -2\langle c_B - p^*, y^*\rangle_{\mathbb{R}^d}\,(c_B - p^*). \end{aligned}$$

These equations show that $p^* = c_B$ is optimal. Thus,

$$D(p^*, y) = \mathbb{E}_B\left[(x - c_B)^\top (x - c_B)\right] - y^\top \operatorname{Cov}_B y,$$

and the minimum is attained for y^* being any eigenvector of length one belonging to the eigenspace of the largest eigenvalue of the covariance matrix Cov_B. In other words, a line through the center in the direction of an eigenvector corresponding to the largest eigenvalue, a principal axis, is an optimal representation of the image.

5.2 Stochastic interpretation

Suppose we are looking for a Gaussian density

$$g_{\Sigma,\mu}(x) := (2\pi)^{-d/2} \det(\Sigma)^{-1} \exp\left(-\tfrac{1}{2}(x-\mu)^\top (\Sigma\Sigma^\top)^{-1}(x-\mu)\right)$$

which fits the density given by the image B optimally in the sense that the so-called log-*likelihood* is maximized,

$$\mathbb{E}_B\left[\log g_{\Sigma,\mu}\right] \xrightarrow{\Sigma,\mu} \max.$$

Elementary computations give

$$\begin{aligned} &\mathbb{E}_B\left[\log g_{\Sigma,\mu}\right] \\ &= -\frac{d}{2}\log(2\pi) - \log(\det\Sigma) - \mathbb{E}_B\left[\frac{1}{2}(x-\mu)^\top(\Sigma\Sigma^\top)^{-1}(x-\mu)\right]. \end{aligned}$$

Differentiation with respect to μ shows that $\mu = c_B = \mathbb{E}_B[x]$ is optimal. Now,

$$\begin{aligned} &2\mathbb{E}_B\left[\log g_{\Sigma,c_B}\right] + d\log(2\pi) \\ &= 2\log(\det(\Sigma^{-1})) - \mathbb{E}_B\left[\operatorname{trace}\left(\Sigma^{-1}(x-c_B)(x-c_B)^\top\Sigma^{-\top}\right)\right] \\ &= -\log(\det\operatorname{Cov}_B) + \log(\det(\Sigma^{-1}\operatorname{Cov}_B\Sigma^{-\top})) - \operatorname{trace}(\Sigma^{-1}\operatorname{Cov}_B\Sigma^{-\top}). \end{aligned}$$

Since $A := \Sigma^{-1}\operatorname{Cov}_B\Sigma^{-\top}$ is real, symmetric, and positive semi-definite, it permits an orthogonal eigendecomposition. Hence, the maximization of the log-likelihood is equivalent to the maximization of

$$\log(\det(A)) - \operatorname{trace}(A) = \sum_{j=1}^{d}(\log\lambda_j - \lambda_j),$$

where the eigenvalues of A are denoted by $\lambda_1, \dots, \lambda_d$. This finally shows that the log-likelihood attains its maximum if and only if $\mu = c_B$ and $\lambda_j = 1$ for $j = 1, \dots, d$. Hence, $I_d = A = \Sigma^{-1}\operatorname{Cov}_B\Sigma^{-\top}$ or $\operatorname{Cov}_B = \Sigma\Sigma^\top$.

Summarizing, the best possible description of the image B in the class of Gaussian densities is given by

$$g_B := (2\pi)^{-d/2}(\det\operatorname{Cov}_B)^{-1/2}\exp\left(-\tfrac{1}{2}(x-c_B)^\top\operatorname{Cov}_B^{-1}(x-c_B)\right).$$

Features of the reference density g_B can be used for a description of the image B.

Note that with $m := \int_{\mathbb{R}^d} B(x)dx$, we have

$$\begin{aligned} m\mathbb{E}_B[\log f] &= \int_{\mathbb{R}^d} \log f(x) B(x)\, dx \\ &= \int_{\mathbb{R}^d} \left[\log \frac{f(x)}{B(x)} + \log B(x)\right] B(x)\, dx \\ &= \int_{\mathbb{R}^d} \log \frac{f(x)}{B(x)} B(x) dx + \int_{\mathbb{R}^d} \log B(x) B(x)\, dx, \end{aligned}$$

where the first term is the non-symmetric so-called Kullback–Leibler distance between f and B (cf., Kullback & Leibler (1951)) and the second term is the negative entropy of the image B. Thus, the previous approach may also be interpreted as a minimization of the Kullback–Leibler distance too.

5.3 A robust generalization

The main advantage of this second interpretation given in Section 5.2 is that it can be generalized. Since the Gaussian distribution (referred to as *standard*) is not robust with respect to perturbations, one might wish to replace it for example by the Cauchy distribution (referred to as *robust*).

Kent & Tyler (1988) showed that the Gaussian density might be replaced by the density of the t-distribution on $\nu > 0$ degrees of freedom. The density of the t-distribution is given by

$$g_{\Sigma,\mu}(x) = C_{\nu,p} \det \Sigma^{-1} (1 + \nu^{-1} \|\Sigma^{-1}(x-\mu)\|^2)^{-(\nu+d)/2},$$

where $C_{\nu,p}$ is some suitable normalization constant which is not dependent on Σ or μ; see, e.g., Mardia et al (1979). In contrast to the Gaussian distribution, which can be viewed as a special t-distribution for $\nu \to \infty$, where the optimal parameter can be computed explicitly, the solution for a general t-distribution is only known in terms of a fixed-point type of equation

$$M = \mathbb{E}_1 \left[u_\nu(x^\top M^{-1} x)\, x x^\top\right] \tag{5.2}$$

where, with $\lambda > 0$, M is an augmented moment matrix,

$$M = \begin{pmatrix} \Sigma\Sigma^\top + \lambda^{-1}\mu\mu^\top & \lambda^{-1}\mu \\ \lambda^{-1}\mu^\top & \lambda^{-1} \end{pmatrix} \in \mathbb{R}^{(d+1)\times(d+1)},$$

and $u_\nu(t) = (\nu + d)/(\nu + t)$; see Kent & Tyler (1991).

Equation (5.2) can be solved numerically using a fixed-point iteration as shown in Kent & Tyler (1991).

5.4 A simple comparison with $d = 2$

The performance of the standard and robust approaches is illustrated for the spatial dimension $d = 2$. We take advantage of an eigendecomposition of the covariance matrix Cov_B,

$$\mathrm{Cov}_B = D(\rho_B)\Sigma_B^2 D(-\rho_B),$$

where for $d = 2$,

$$D(\rho_B) = \begin{pmatrix} \cos\rho_B & -\sin\rho_B \\ \sin\rho_B & \cos\rho_B \end{pmatrix}, \quad \Sigma_B = \begin{pmatrix} \sigma_{B,1} & 0 \\ 0 & \sigma_{B,2} \end{pmatrix};$$

cf., eqn (5.1). The direction of the principal axis is given by $(\cos\rho_B, \sin\rho_B)^\top$ and the direction of the secondary axis is given by $(-\sin\rho_B, \cos\rho_B)^\top$. Thus, the angle ρ_B can be used to describe both axes.

Figure 5.1 illustrates these quantities. The figure shows the principal (solid) and secondary axis (dashed), where the lengths of the axes indicate the standard deviations. The center $c_B = (c_{B,1}, c_{B,2})^\top$ is the intersection point of the two axes. These five quantities might by viewed as an image feature, i.e.,

$$\mathcal{F}_B := (c_{B,1}, c_{B,2}, \sigma_{B,1}, \sigma_{B,2}, \rho_B). \tag{5.3}$$

The four pictures in Fig. 5.1 illustrate this feature for different PAT approaches for an unperturbed image:

1. *standard approach* based on the Gaussian density (cf., Section 5.2),
2. *robust approach* based on the Cauchy density (cf., Section 5.3),
3. *standard approach* for the binary image, and
4. *robust approach* for the binary image.

In Fig. 5.2 we show analogous results for the artificially perturbed image $\hat{B}$. The numerical values for these examples are summarized in Table 5.1.

As is apparent from this example, the standard approach (based on a Gaussian density) is much more sensitive with respect to a perturbation of the data than the robust approach (based on a Cauchy density). The impact of the perturbation can be dampened using the binary image instead of the image itself as a reference density. However, ignoring the intensity might not be a good idea in many applications since typically the intensity is meaningful.

5.5 Principal axes under transformation

We are interested in the transformation properties of the feature (5.3) under an affine linear map. In other words, we are interested in the center and covariance matrix of the image $\hat{B} := B \circ \varphi$ where $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ is affine linear, $\varphi(x) = Ax + b$, $\det A > 0$.

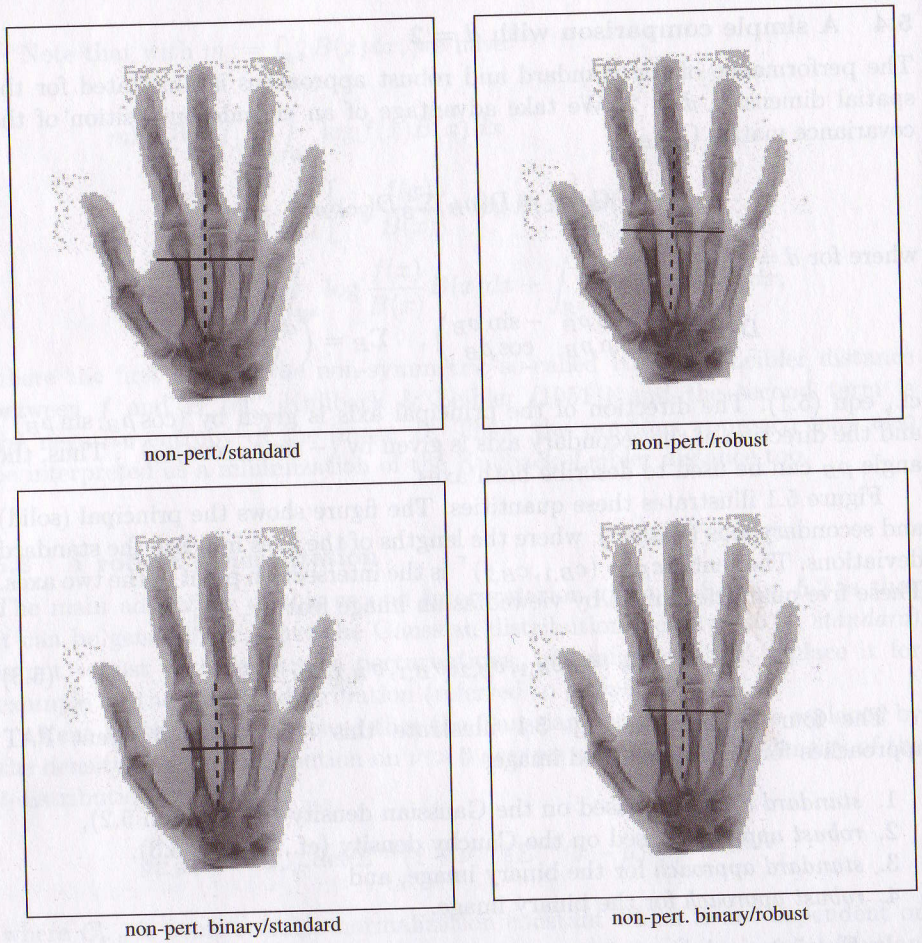

FIG. 5.1 Non-perturbed images of a human hand with principal axis (solid) and secondary axis (dashed); the intersection point indicates the location of the center c_B and the length of an axis illustrates the corresponding eigenvalue of Cov_B.

Theorem 5.1 *Let $B : \mathbb{R}^d \to \mathbb{R}$ be an image with center c_B and covariance matrix Cov_B; cf., Definition 5.1. Moreover, let $\hat{B}(x) := B(Ax + b)$, where $A \in \mathbb{R}^{d\times d}$, $\det(A) > 0$, and $b \in \mathbb{R}^d$. Then*

$$c_B = Ac_{\hat{B}} + b \quad \textit{and} \quad \mathrm{Cov}_B = A\mathrm{Cov}_{\hat{B}}A^\top.$$

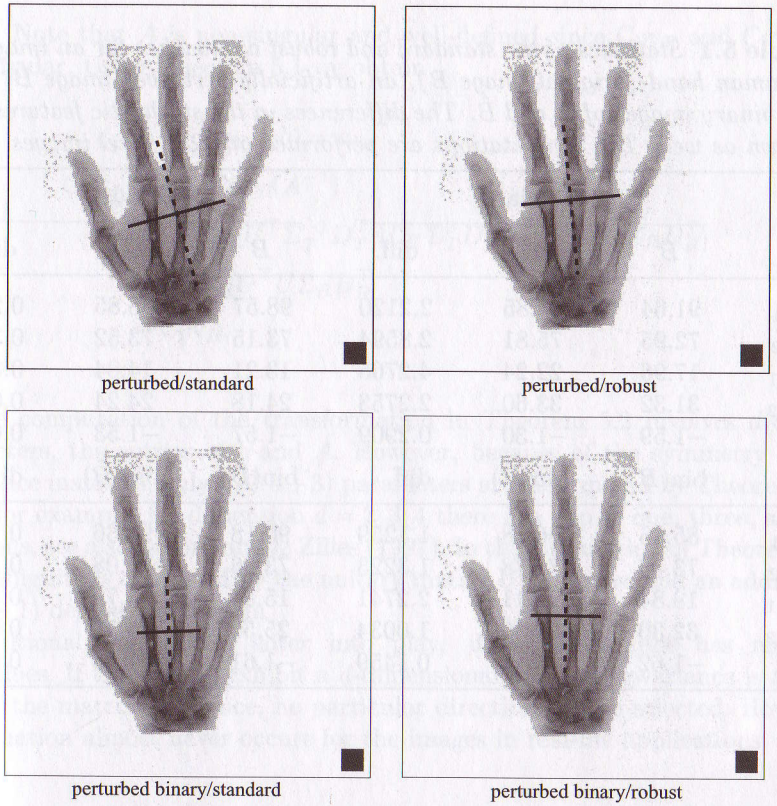

FIG. 5.2 Perturbed images of a human hand with principal axis (solid) and secondary axis (dashed); the intersection point indicates the location of the center c_B and the length of an axis illustrates the corresponding eigenvalue of Cov_B.

Proof From the transformation theorem (cf., e.g., Buck (1978, Theorem 6, §8.3)) with $y = \varphi(x) = Ax + b$ and $dy = \det A\, dx$, we have

$$\int_{\mathbb{R}^d} f(y)B(y)dy = \det(A) \int_{\mathbb{R}^d} f \circ \varphi(x)\hat{B}(x)dx$$

and hence,

$$\begin{aligned}
\mathbb{E}_B[f] &= \mathbb{E}_{\hat{B}}[f \circ \varphi], \\
c_B &= \mathbb{E}_B[x] = A\mathbb{E}_{\hat{B}}[x] + b = Ac_{\hat{B}} + b, \\
\mathrm{Cov}_B &= \mathbb{E}_B\left[(x - c_B)(x - c_B)^\top\right] \\
&= \mathbb{E}_{\hat{B}}\left[A(x - c_{\hat{B}})(x - c_{\hat{B}})^\top A^\top\right] = A\mathrm{Cov}_{\hat{B}}A^\top.
\end{aligned}$$

□

Table 5.1 *Statistics of the standard and robust approaches for an image of a human hand (original image B), an artificially perturbed image $\hat{B}$, and the binary images of B and $\hat{B}$. The differences in the stochastic features are shown as well. The computations are performed on 128^2 pixel images.*

	Standard			Robust		
	B	$\hat{B}$	diff.	B	$\hat{B}$	diff.
$c_{B,1}$	91.64	93.85	2.2120	98.57	98.85	0.2829
$c_{B,2}$	72.95	75.81	2.8594	73.15	73.52	0.3711
$\sigma_{B,1}$	17.96	22.24	4.2765	13.21	14.04	0.8371
$\sigma_{B,2}$	31.32	33.60	2.2753	24.18	24.24	0.0509
ρ_B	−1.59	−1.30	0.2909	−1.57	−1.53	0.0448
	bin(B)	bin($\hat{B}$)	diff.	bin(B)	bin($\hat{B}$)	diff.
$c_{B,1}$	85.65	86.78	1.1259	90.18	90.36	0.1847
$c_{B,2}$	73.46	74.78	1.3223	73.86	74.08	0.2180
$\sigma_{B,1}$	19.84	22.11	2.2741	15.31	15.70	0.3949
$\sigma_{B,2}$	32.30	33.30	1.0034	25.52	25.57	0.0554
ρ_B	−1.62	−1.47	0.1459	−1.61	−1.58	0.0266

Theorem 5.1 is a starting point for registration purposes. The idea is to carry out some normalization of the images. Exploiting two maps φ_R and φ_T, the reference image R and the template image T are mapped to $\widetilde{R}$ and $\widetilde{T}$, respectively, such that $c_{\widetilde{R}} = c_{\widetilde{T}} = 0$ and $\mathrm{Cov}_{\widetilde{R}} = \mathrm{Cov}_{\widetilde{T}} = I_d$. Hence, $\varphi = \varphi_T \varphi_R^{-1}$ gives a registration of T. This idea is the topic of the next theorem.

Theorem 5.2 *Let R and T be two images with centers c_R and c_T and non-singular covariance matrices Cov_R and Cov_T, respectively. Defining $\hat{T}(x) := T(Ax + b)$, where*

$$A := D_T \Sigma_T U \Sigma_R^{-1} D_R^\top \in \mathbb{R}^{d\times d}, \quad b := c_T - A c_R \in \mathbb{R}^d,$$

$U \in \mathbb{R}^{d\times d}$ is an arbitrary unitary matrix, and the eigendecompositions $\mathrm{Cov}_R = D_R \Sigma_R^2 D_R^\top$ and $\mathrm{Cov}_T = D_T \Sigma_T^2 D_T^\top$ are used.

Then we have

$$c_{\hat{T}} = c_R \quad \text{and} \quad \mathrm{Cov}_{\hat{T}} = \mathrm{Cov}_R.$$

Proof Note that A is non-singular and well-defined since Cov_R and Cov_T are non-singular. Using Theorem 5.1 we obtain

$$\begin{aligned} c_{\hat{T}} &= A^{-1}(c_T - b) = c_R, \\ \mathrm{Cov}_{\hat{T}} &= A^{-1}\mathrm{Cov}_T (A^{-1})^\top \\ &= [D_R \Sigma_R U^\top \Sigma_T^{-1} D_T^\top][D_T \Sigma_T^2 D_T^\top][D_T \Sigma_T^{-1} U \Sigma_R D_R^\top] \\ &= D_R \Sigma_R U^\top U \Sigma_R D_R^\top \\ &= \mathrm{Cov}_R. \end{aligned}$$

□

The computation of the transformation in Theorem 5.2 involves $d(d+1)$ parameters, the entries of b and A. However, because of the symmetry of the covariance matrices, only $d/2(d+3)$ parameters are determined by Theorem 5.1. Thus, for example, for dimension $d = 2, 3, 4$ there is a gap of one, three, and six equations; see also Schormann & Zilles (1997). In the formulation of Theorem 5.2, this ambiguity is expressed by the unitary matrix U. Here, we find an additional $d/2(d-1)$ degrees of freedom.

Additional ambiguities enter into play, if the covariance has multiple eigenvalues. If the images exhibit a d-dimensional ball, the covariance is a multiple of the matrix I_d. Hence, no particular direction can be selected. However, this situation almost never occurs for the images in real-life applications.

5.6 An example of principal axis-based registration

Figure 5.3 demonstrates that the PAT can be used for image registration successfully. However, Theorem 5.2 already indicates that there is ambiguity in the choice of the unitary matrix U. This ambiguity is illustrated in Fig. 5.4, where

FIG. 5.3 PAT registration images from hands using the robust approach. Reference image R (LEFT), template image T (MIDDLE), and transformed template $\hat{T}$ (RIGHT).

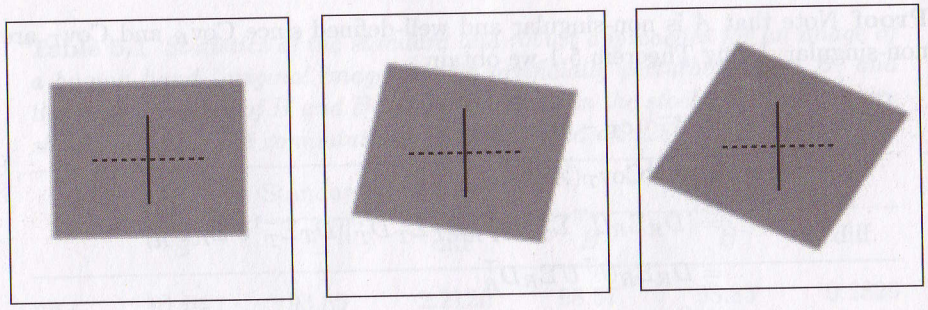

FIG. 5.4 Three different rectangles sharing the same stochastic features.

three different rectangles are depicted. Although these rectangles look quite different, they all share the same stochastic features. As a consequence, the PAT cannot distinguish between these different images.

6

OPTIMAL LINEAR REGISTRATION

In this chapter we investigate the question of how to find an optimal linear registration based on a distance measure $\mathcal{D}$. An analytical solution cannot be expected for the images from our application. Thus, we have to look for a numerical solution. Moreover, since the images under consideration are of high resolution, we focus on fast and efficient schemes, which typically exploit derivatives. Thus, an important property of the distance measure to be discussed is its differentiability.

The choice of an appropriate distance measure is a difficult task. Popular choices to be discussed in the subsequent sections are based on *intensity* (see, e.g., Brown (1992)), *correlation* (see, e.g., Collins & Evans (1997)), or *mutual information* (see, e.g., Viola (1995) or Collignon et al (1995)).

For some particular applications, modifications of these similarity measures have been investigated; see, e.g., Studholme et al (1996) or Roche et al (1999). In addition, the distance measure used in the registration may also be based on particular image features, e.g., edges or surfaces.

We start by introducing a set of feasible transformations, which here are supposed to be affine linear maps, i.e., $\varphi \in \Pi_1^d(\mathbb{R}^d)$; cf., Definition 3.6. A mathematical formulation of the registration problem then reads as follows.

Problem 6.1 *Find* $\varphi \in \Pi_1^d(\mathbb{R}^d)$ *such that* $\mathcal{D}[\varphi] = \min$.

The essential point here is that the set $\in \Pi_1^d(\mathbb{R}^d)$ can be parameterized. For a specific element φ of $\in \Pi_1^d(\mathbb{R}^d)$, we make use of the notation φ_a, where

$$\varphi_{a;\ell}(x) = a_{\ell,0} + \sum_{j=1}^{d} a_{\ell,j} x_j, \quad \ell = 1, \ldots, d.$$

The parameters $a_{\ell,j}$ are gathered together in a vector,

$$a = (a_{1,0}, \ldots, a_{1,d}, \ldots, \quad a_{d,0}, \ldots, a_{d,d})^\top \in \mathbb{R}^n, \quad n = d(d+1).$$

Moreover, we set

$$D(a) := \mathcal{D}[\varphi_a] \quad \text{and} \quad T_a := T \circ \varphi_a. \tag{6.1}$$

Thus, Problem 6.1 may be reformulated in terms of a parameterized finite-dimensional optimization problem.