Identification of a Discrete Planar Symmetric Shape from a Single Noisy View

Oleg V. Poliannikov
Department of Mathematics, Box 8205
North Carolina State University
Raleigh, NC 27695-8205

Hamid Krim
Department of ECE, Box 7914
North Carolina State University
Raleigh, NC 27695-7914

Abstract

In this paper, we propose a method for identifying a discrete planar symmetric shape from an arbitrary viewpoint. Our algorithm is based on a newly proposed notion of a view’s skeleton. We show that this concept yields projective invariants which facilitate the identification procedure. It is, furthermore, shown that the proposed method may be extended to the case of noisy data to yield an optimal estimate of a shape in question. Substantiating examples are provided.

1 Introduction

Automatic understanding of objects from their images is a fundamental problem in computer vision [1]. In the vast majority of practical situations, we do not have direct access to an object of interest and frequently have to rely on its images which are a result of a variety of potential devices. Given that an object lives in a three-dimensional space, and its image is typically two-dimensional, it is clear that the problem of extracting all possible information about the object from its image is fundamentally ill-posed.
Mathematical formulations of the problem of identifying an object vary considerably with the exact and prevailing practical setting along with the subsequent natural assumptions, which follow [2,3,4,5,6]. A lot of research has been devoted to studying objects based on two and more images [7,8,9]. It is clear that the more data is available about an object, the more information one may potentially extract from it. Often times, however, a single image of an object is all that is available, and fully exploiting the extracted information becomes an important problem.

This paper is devoted to the problem of identifying a symmetric planar discrete shape from a single observed image. By a discrete planar shape, we will understand a collection of discrete planar points. While seemingly unnatural, and in light of the continuity of real-life objects along with their images, a discrete shape may be argued as being a good model in many practical cases. For purposes such as recognition, a shape defined by a finite number of so-called landmarks has been proposed [10,11,12,13]. It is implicitly understood that a continuous shape may be reconstructed from its discrete counterpart or that the latter carries most of the required information.

It has been shown that if no additional assumptions are imposed on a discrete object, the knowledge gleaned from the image is insufficient, i.e., different objects may generate the same image which underscores a relative difficulty in telling them apart [14]. A common approach to recognition has been therefore to impose certain constraints on a shape space that would allow the construction of invariants able to capture the geometric information about the object in question.

Our interest lies in studying symmetric shapes, which are quite common in many applications. Zabrodsky et al [15] has proposed a method to detect symmetry in an object by means of introducing a concept of a symmetric difference - mismatch between an object and its closest symmetric approximation. An ability to detect symmetry in an automated fashion enables on to use this feature in an automatic recognition procedure. Because for the purposes of our paper, one requires to detect symmetry in an arbitrary projective view of an object, we outline a possible algorithm to do that in Appendix A.
This paper is organized as follows. In the first part of this paper, we show that virtually all geometric information about any shape satisfying the above conditions may be conveniently encoded by what we refer to as the skeleton of a shape and consequently by its features [16]. A skeleton of a shape has a very simple geometric structure, which bears all information about the shape of interest. This shape may be computed from an image of any object, affording one the ability to analyze the object regardless of its available image.

The second part of the paper seeks a more practical approach by adapting methods developed in the first part to noisy shapes. Discretizing a continuous shape may be performed by either sampling algorithm or manually. In either case the accuracy of landmarks is a very crucial issue. Noisy samples of a shape will typically invalidate all properties of discrete shapes, making the recovery of relevant information about the object, impossible.

2 Elements of Projective Geometry

2.1 Basic Definitions

We will consider a standard projective space \( \mathbb{P}^n \), which is an extension of the Euclidean space \( \mathbb{R}^n \) with added “infinite points”. For any point \( x \in \mathbb{P}^n \) we will use its projective, \((x_1, \ldots, x_n, x_{n+1})^T\), and Euclidean, \((\frac{x_1}{x_{n+1}}, \ldots, \frac{x_n}{x_{n+1}})\), coordinate representations interchangeably. If \( x_{n+1} = 0 \), the point \( x \) is called a point “at infinity” in the direction \((x_1, \ldots, x_n)\). For example, in the case of a 2−D projective space, the point \((1, 0, 0)\) corresponds to the horizontal direction, and \((0, 1, 0)\) corresponds to the vertical direction on the plane.

2.2 Imaging in 3−D Projective Space

It is customary to model 3−D real objects as collections of points lying in a 3−D projective space \( \mathbb{P}^3 \). The image of a given object in \( \mathbb{P}^3 \) is itself an object lying in the 2−D projective space \( \mathbb{P}^2 \).
A pin-hole camera is the most widely used model. A pinhole camera is defined by a retinal plane $R$, together with an optical center $C$. The image $\mathbf{x}$ of a given point $\mathbf{X}$ through this camera is obtained, as shown in Figure 2, by intersecting the line connecting the points $C$ and $\mathbf{X}$ with the plane $R$. Mathematically, $\mathbf{x} = \mathbf{P}\mathbf{X}$, where $\mathbf{X}$ is the world point, $\mathbf{P}$ is the camera matrix, and $\mathbf{x}$ is the image of $\mathbf{X}$ through the camera $\mathbf{P}$, where

$$
\mathbf{P} = \begin{pmatrix}
    f_x & 0 & c_x & 0 \\
    0 & f_y & c_y & 0 \\
    0 & 0 & 1 & 0
\end{pmatrix}.
$$

(1)

Here $f_x, f_y, c_x, c_y$ are the so-called intrinsic parameters of the camera.

Introduce the matrix $\mathbf{K}$ as follows

$$
\mathbf{K} = \begin{pmatrix}
    f_x & 0 & c_x \\
    0 & f_y & c_y \\
    0 & 0 & 1
\end{pmatrix}.
$$

(2)

Then

$$
\mathbf{P} = \mathbf{K}(\mathbf{I}_{3\times3} | \mathbf{0}_{3\times1}),
$$

(3)

where $\mathbf{I}_{3\times3}$ is a $3 \times 3$ identity matrix. The matrix $\mathbf{K}$ is the so-called camera calibration matrix and its entries $f_x, f_y, c_x, c_y$ form the intrinsic parameters of the camera [7]. Intrinsic parameters of each type
of camera are typically estimated in laboratory conditions. For applications where we have access to the camera or its physical parameters, we may safely assume the latter known.

2.2.1 Extrinsic Parameters of Camera

Given that an arbitrary world coordinate system may be rotated and translated with respect to the camera frame, one introduces $R$ - a $3 - D$ rotation matrix, and $C$ - the camera center’s coordinate vector in the world coordinate frame.

Then we can write \[7\]:

$$x = KR(I_{3 	imes 3} | -C) X. \tag{4}$$

The rotation matrix $R$ is a function of three parameters $\phi, \psi$ and $\alpha$. The first two determine the Euclidean angles of the rotation axis, and the latter is the rotation angle (see Figure 4).

The rotation parameters $\phi, \psi, \alpha$ and the vector $C$ are the extrinsic parameters of a camera.
Figure 3: Arbitrary coordinate frame.

Figure 4: Rotation parameters.
2.3 Views of Planar Shapes and Projective Transformations of Plane

Consider a planar point in a 3−D projective space $\mathbf{X}$, such that $\mathbf{X} = (X, Y, 0, 1)$. Let $\mathbf{P}$ be an arbitrary camera matrix, and $\mathbf{x}$ be the image of $\mathbf{X}$ through that camera, i.e. $\mathbf{x} = \mathbf{P}\mathbf{X}$. This may be written as

$$\mathbf{x} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} & p_{14} \\ p_{21} & p_{22} & p_{24} \\ p_{31} & p_{32} & p_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}. \quad (5)$$

**Definition 1** Denote the frontal view of $\mathbf{X}$ by $\mathbf{x}^0 = (X, Y, 1)^T$. A point $\mathbf{x}$ obtained by $\mathbf{x} = \mathbf{H}\mathbf{x}^0$ is the projective transformation of $\mathbf{x}^0$.

It is clear that an arbitrary view of a planar shape may always be viewed as a projective transformation of its frontal view via an arbitrary matrix $\mathbf{H}$.
2.3.1 Projective Transformations of Plane

Definition 2 A mapping \( \mathcal{H} : \mathbb{P}^2 \to \mathbb{P}^2 \) is called a projective transformation if it is of the form:

\[ \mathcal{H}(x) = Hx, \text{ where } H \text{ is a non-degenerate } 3 \times 3 \text{ matrix.} \]

Projective transformations on the plane preserve the collinearity of points [7], (see Figure 6), - a property that will later be useful.

3 Identification of Shape from a Single View

Suppose we have a discrete symmetric planar object in a 3 − D projective space, which, recall, is a collection of points \( \{X_i\}_{i=1}^{2N} \), such that \( X_i \in \{X \equiv (X, Y, Z, 1) \in \mathbb{P}^3 \mid Z = 0\} \), \( i = 1, \ldots, 2N \), with \( X_i \) symmetric to \( X_{i+N} \) about the axis \( \{Y = 0\} \), i.e. \( X_i = -X_{i+N}, \ i = 1, \ldots, N \). Assume further that an object \( X \) is observed through a projective camera \( P \) yielding a vector \( x \in \mathbb{P}^2 \) given by \( x = PX \).

Different camera matrices \( P \) generate different views of the object of interest. For two different images, \( x_1 \) and \( x_2 \) of an object \( X \) viewed by two different cameras \( P_1 \) and \( P_2 \), one may establish, by way of a projective transformation, the relation \( x_2 = Hx_1 \). In particular, all other views of the same object are projective transformations of the front view \( x^0 \).
3.1 Skeleton of Symmetric Planar Shape

Let \( \{x_i\}_{i=1}^{2N} \) be a fixed image of an object \( \{X\}_{i=1}^{2N} \), that is \( x_i = Hx_0^i \), \( i = 1, \ldots, 2N \) for some fixed projective transformation (matrix) \( H \), where \( \{x_0^i\}_{i=1}^{2N} \) is a frontal view of the object. Our immediate goal is now to prove the following

**Proposition 1** Virtually all projective geometry of the views \( \{x_i\}_{i=1}^{2N} \) can be conveniently captured by only \( 2N - 3 \) real parameters. (Note that \( \{x_i\}_{i=1}^{2N} \) is in general determined by \( 4N \) real numbers).

To first introduce the so called skeleton\(^1\) of a shape (see Figure 7), consider two pairs of adjacent symmetric points \( x_i, x_{i+N}, x_{i+1}, x_{i+1+N} \). Using \( l_{x_1,x_2} \) to denote a line connecting two points \( x_1, x_2 \), we may define two points: \( n_i = l_{x_i,x_{i+1+N}} \cap l_{x_{i+1},x_{i+N}} \), and \( p_i = l_{x_i,x_{i+1}} \cap l_{x_{i+N},x_{i+1+N}} \).

The symmetry condition guarantees that the points \( n_i, p_i \) are on the imaged axis of symmetry, i.e. \( n_i, p_i \in l_{n_i,p_i} \equiv l_{sym} \equiv \{Hx \mid x \in l_{sym}\} \), where \( l_{sym} \) is the symmetry axis of the frontal view. Let \( m_i = l_{x_i,x_{i+N}} \cap l_{n_i,p_i} \), where \( m_i \) is on the image of the symmetry axis \( l_{sym} \), and the index \( i \) varies from 1 to \( 2N \), we get \( 2N - 1 \) collinear points \( \{m_1, n_1, m_2, n_2, \ldots, m_{N-1}, n_{N-1}, m_N\} \in l_{sym} \). This relabeled collection as \( \{s_1, \ldots, s_{2N-1}\} \) is henceforth referred to as the “skeleton” of the shape \( \{x_i\}_{i=1}^{2N} \).

---

\(^1\)”Skeleton” is a somewhat unfortunate choice of a term, as other objects are also known by this name. For the lack of a better alternative, we will nonetheless stick with this terminology.
Figure 8: Example of skeleton of a perspective view of a jar.

Our interest in the shape of an object view aims at understanding the geometry of an image of the original symmetrical shape. We will show that almost all such information is included in the skeleton. Studying planar shapes via their skeleton greatly simplifies the analysis of problems, such as object classification and recognition.

3.1.1 A Skeleton’s Signature

As just mentioned, a skeleton is a collection of points on a line, which is the projective image of a symmetry axis of a shape’s frontal view (see Figure 8). The geometric characteristics of a skeleton are the number of points and their relative positions. The inter-distances between all points are, however, an inefficient geometric measure as these vary with different views.

If we are given four points $s_1, \ldots, s_4$ on a line, we may instead define their cross-ratio as

$$\sigma(s_1, s_2, s_3, s_4) = \frac{\|s_1s_3\| \|s_2s_4\|}{\|s_1s_4\| \|s_2s_3\|},$$

where $\|s_is_j\|$ denotes the Euclidean length of the vector connecting the points $s_i$ and $s_j$.

The fundamental property of a cross-ratio of 4 collinear points is its invariance under any projective
transformation (18). This will be particularly useful in our analysis of symmetric planar shapes whose
skeletons have been identified. The choice of 4-tuples from a skeleton set of points is not unique and
the latter, does, however, encapsulate all the information about a skeleton.

**Definition 3** Let \( s_1, \ldots, s_M \), where \( M \equiv 2N - 1 \geq 4 \), be \( N \) collinear points (a skeleton of a projective
image of a symmetric planar shape). We will define its signature as a vector of cross-ratios \( \sigma \equiv \{ \sigma_1, \ldots, \sigma_{M-3} \} \), where \( \sigma_i = \sigma(s_i, s_{i+1}, s_{i+2}, s_{i+3}) \), \( i = 1, \ldots, M - 3 \).

Since all individual cross-ratios are projective invariants and independent of a choice of a projective
camera, the joint vector \( \sigma \) naturally inherits this property. As a signature, it hence represents a valid
quantitative description of the geometry of a skeleton (see Figure 9).

**Proposition 2** A signature defines the skeleton up to a projective transformation.

**Proof.** We prove this fact by constructing all possible skeletons from a given \( M - 3 \) dimensional
signature vector \( \sigma \). Recall that a signature remains invariant to a rotation of the point bearing the line,
which may just as well be chosen to coincide with the ordinate axis. By fixing the first three points
\( s_1, s_2, s_3 \) at arbitrary locations, the remaining points may immediately be derived from the following
recursion:

\[
s_{i+3} = \frac{s_{i+2} - s_i}{(1 - \sigma_i)} \frac{s_{i+2} - s_i}{s_{i+1} - s_i} + s_i, \quad i = 1, \ldots, M - 3.
\]
It is hence clear, that the space of skeletons engendered by a given signature enjoys three (3) degrees of freedom. This confirms a signature definition of a skeleton up to a projective transform, hence its encapsulation of all related geometric information.

3.2 Reconstructing View of Shape from Skeleton

Recall that a projective image of an object of a symmetric planar shape, has an associated unique skeleton and consequently a signature. This signature is independent of the choice of a projective camera and is thus truly a projective invariant.

In this light, and for a given signature \( \{\sigma_1, \ldots, \sigma_{M-3}\} \), our goal is to reconstruct a view of an object up to a projective transformation. To that end, we proceed by first constructing a corresponding skeleton. Fix \( s_1 = (0, 1) \) and \( s_2 = (0, 2) \), and then using the result in the previous subsection, we can find \( s_3, \ldots, s_M \), such that the entire skeleton \( \{s_i\}_{i=1}^M \) matches the signature \( \{\sigma_j\}_{j=1}^{M-3} \).

To reconstruct a shape from a skeleton, we recall that for a frontal view of the shape \( \{x_0^i\}_{i=1}^{2N} \), the lines connecting the symmetric points are parallel (in the Euclidean geometry), yielding their intersection at a vanishing point \( x_0^{x} = (1, 0, 0)^T \) in projective geometry. As was mentioned earlier, we have \( x_i = H x_0^i \). The point \( x_0^i \) is then mapped to a new “vanishing point” \( x_v \). Denote by \( l_v^i \) the line that connects the points \( s_{2i-1} \) and \( x_v \). It follows that the points \( x_i \) and \( x_{i+N} \) belong to the line \( l_v^i \). Let us fix \( x_1 \in l_v^1 \).

Noting that \( x_{1+N}, s_1, x_1, x_v \) are by construction 4 collinear points, we reconstruct the point \( x_{1+N} \). Their cross-ratio being a projective invariant, equals that of the corresponding points in the frontal view. The cross-ratio of any points of the form \( -x, 0, x, +\infty \) is 2 as follows immediately from the definition of the cross-ratio. Given three collinear points of the original view, and a known cross-ratio, there is sufficient information to uniquely reconstruct the fourth point.

As illustrated in Figure 10 a shape reconstruction is achieved by making use of the given skeleton together with a choice of two additional points: the vanishing point \( x_v \) and the exact location of the
Figure 10: Reconstruction of a view from a skeleton.

point $x_1$. The vanishing point is a property of a camera, and its variation yields different views of the same shape.

Different choices of $x_1$ yield views of different shapes. Note that the exact location of $x_1$ controls the width of the underlying shape, i.e. all different shapes produced by different choices of $x_1$ are copies of the same shape stretched in the horizontal direction, as illustrated in Figure 11.

Figure 11: Different choice of $x_1$ affects only the width of the underlying shape.
4 Identification of Shape from a Noisy Image

It was argued that a skeleton and its signature may play an important role in identifying a planar symmetric shape from an arbitrary viewpoint. More precisely, the following has been established: (1) To any projective view of a planar symmetric shape, we can associate a skeleton. (2) A skeleton is a collection of points lying on the same line, which is a projective image of the symmetry axis. (3) Cross-ratios are effective at capturing the projective geometry of skeleta and serve as a basis for the notion of signature of skeleton. (4) A signature characterizing a shape of interest from an arbitrary view may be used for object identification applications.

Several practical issues arise and somewhat hinder a systematic application of the proposed approach. Shapes are in theory continuous (at least comprise a large number of points), and their lower dimensional representation entails sampling, or a manual selection of landmarks. This invariably yields errors in the locations of these landmarks, which in turn requires further adjustments of the afore-proposed technique.

For a given noisy shape, it is indeed possible to apply a similar geometric construction, which would yield the shape’s “noisy skeleton points”. These will clearly not be collinear, and skeletons corresponding to different views will not be related by means of any projective transformation.

These two arguments warrant an additional effort towards the adapting of the previously proposed concepts to a noisy environment. Our goal will henceforth be to study noisy views of discrete planar symmetric shapes and to develop techniques to address their identification problems.

By a noisy view of a symmetric planar shape, we will understand a collection of Euclidean points \( \{x_i\}_{i=1}^{2N} \), such that \( x_i = (x_i, y_i) \), \( i = 1, \ldots, 2N \), where

\[
(x_i, y_i) = (x_i^0, y_i^0) + (\Delta x_i, \Delta y_i), \quad i = 1, \ldots, 2N,
\]

(8)

with \( \Delta x_i, \Delta y_j \) independent and identically distributed Gaussian random variables \( \mathcal{N}(0, \epsilon^2) \). The latter is a standard and well-accepted model for noise. A precise value for \( \epsilon \) can be either estimated experi-
mentally or set to be the maximum possible noise energy. The points \( \{(x_i^0, y_i^0)\}_{i=1}^{2N} \) in turn are a noise free observation of a shape through some arbitrary projective camera \( P \).

### 4.1 Skeleton of Noisy Shape

Note that the procedure of constructing a skeleton may be applied in presence / absence of noise. For a noisy view, the points \( \{(x_i, y_i)\}_{i=1}^{2N} \), as mentioned earlier, are not guaranteed to be collinear. As a result, each individual point of a “noisy skeleton” \( \{s_i\}_{i=1}^{2N-1} \) will be a perturbation of the corresponding point of the skeleton \( \{s_0^i\}_{i=1}^{2N-1} \) of the underlying noise free view \( \{(x_i^0, y_i^0)\}_{i=1}^{2N} \).

Our goal is to study the distribution of a noisy skeleton and to subsequently develop a technique for estimating the “clean skeleton”. Let a “clean view” \( \{x_0^i\}_{i=1}^{2N} \) of a symmetric planar shape be a result of a projective camera \( P \) with known parameters. Its corresponding “noisy view” is according to \( \{s_i\}_{i=1}^{2N-1} \). Denote the clean skeleton by \( \{s_0^i\} \) and the noisy skeleton by \( \{s_i\} \). Given the statistics of the perturbations, the vector \((s_1, \ldots, s_{2N-1})^T\) may be experimentally shown to be Gaussian \( \mathcal{N}(s^0, \Sigma) \),

\[ \text{In this section the superscript 0 does not necessarily refer to the frontal view.} \]
Figure 13: Realizations of noisy skeletons. For a fixed view, IID Gaussian noise in the data yields correlated Gaussian noise in the skeleton points.

where $E_s_i = s_i^0$, and the covariance matrix $\Sigma$ depend on the projection camera matrix $P$ as well as on the shape itself.

The complete analytical proof of these properties remains an open problem because of non-trivial nested systems of linear equations tying together the shape points and the points of a skeleton, and their validity is experimentally demonstrated (see Figure 13).
4.1.1 Covariance Matrix of Skeleton Points

As just noted, we have experimentally established that to a given noisy view of a symmetric planar shape, correspond skeleton points with a Gaussian distribution, with means coinciding with noise-free skeleton-points. Due to intractability of the covariance matrix \( \Sigma \) derivation, we propose a numerical evaluation, which uses a function that is independent of a particular choice of shape and hence can be a priori tabulated.

Suppose, that \( s_i = (s_i, t_i) \), \( Es_i = s_i^0 \), \( Et_i = t_i^0 \), \( i = 1, \ldots, 2N-1 \) are skeleton points. Define the vector \( \tilde{s} = (s_1, \ldots s_{2N-1}, t_1, \ldots, t_{2N-1})^T \). The covariance matrix of interest then is \( \Sigma = (\Sigma_{ij})_{i,j=1}^{4N-2} \), \( \Sigma_{ij} = cov(\tilde{s}_i, \tilde{s}_j) \). An experimental determination of \( \Sigma \) would not only require a massive amount of data, but would be required for each view of each shape. To avoid such a computational burden, we propose the following alternative. Viewing each skeleton point as a local feature of a shape, we exploit the fact that not all skeleton points are correlated. Specifically, as illustrated in Figure 14, we note

\[
E[\Delta s_i \Delta s_{i+j}] = E[\Delta s_i \Delta t_{i+j}] = E[\Delta s_{i+j} \Delta t_i] = E[\Delta t_i \Delta t_{i+j}] = 0,
\]  

(9)

where \( j > 4, i = 2k - 1, \) and

\[
E[\Delta s_i \Delta s_{i+j}] = E[\Delta s_i \Delta t_{i+j}] = E[\Delta s_{i+j} \Delta t_i] = E[\Delta t_i \Delta t_{i+j}] = 0,
\]  

(10)

where \( j > 3, i = 2k, \) with \( \Delta s_i = s_i - s_i^0 \), \( \Delta t_i = t_i - t_i^0 \). This is easily seen by observing that a correlation of skeleton points results when a common set of data points is used in their construction. In particular, the values of \( \Sigma_{1,i}, \Sigma_{2,i}, \Sigma_{2N,i}, \Sigma_{2N+1,i} \) may be computed if the locations of data points \( x_1, x_2, x_3 \) are known, \( x_{1+N}, x_{2+N}, x_{3+N} \). Likewise, to compute the covariance coefficients corresponding to any two adjacent skeleton points \( s_{2k-1}, s_{2k} \) requires 6 data points whose noise impacts their distribution.

Define the function \( \Lambda(x_1, x_2, x_3, x_4, x_5, x_6) \), as

\[
\Lambda(x_1, x_2, x_3, x_4, x_5, x_6) = \begin{pmatrix}
\lambda_{1,1} & \lambda_{1,2} & \lambda_{1,3} & \lambda_{1,4} \\
\lambda_{2,1} & \lambda_{2,2} & \lambda_{2,3} & \lambda_{2,4} \\
\lambda_{1,2} & \lambda_{2,2} & \lambda_{2,3} & \lambda_{2,4} \\
\end{pmatrix},
\]  

(11)
where the function value is a matrix that consists of correlation coefficients of the skeleton points which are a result of the given data points. This function may then be \emph{a priori} tabulated on a grid, thus allowing us to inductively compute the full covariance matrix of the skeleton of an arbitrary view of a shape.

To help reduce the memory resources required by the computation of \( \Sigma \), we observe that we only need to store components resulting from sets of points (6) defined up to rotation, scale and translation, as illustrated in Figure 15. A translation of a shape clearly leaves the covariance matrix invariant, scaling of the shape results in corresponding scaling of \( \Lambda \), and finally a rotation

\[
\begin{pmatrix}
    s_i^2 \\
    t_i^2
\end{pmatrix}
= 
\begin{pmatrix}
    \cos \phi & \sin \phi \\
    -\sin \phi & \cos \phi
\end{pmatrix}
\cdot
\begin{pmatrix}
    s_i^1 \\
    t_i^1
\end{pmatrix},
\]

(12)
yields.

\[
E \left[ \Delta s_i^2 \Delta s_j^2 \right] = 
\cos^2 \phi \ E \left[ \Delta s_i^1 \Delta s_j^1 \right] + 
\sin \phi \cos \phi \ (E \left[ \Delta s_i^1 \Delta t_j^1 \right] + 
E \left[ \Delta s_j^1 \Delta t_i^1 \right]) + 
\sin^2 \phi \ E \left[ \Delta t_i^1 \Delta t_j^1 \right]
\]

\[
E \left[ \Delta s_i^2 \Delta t_j^2 \right] = 
\cos^2 \phi \ E \left[ \Delta s_i^1 \Delta t_j^1 \right] + 
\sin \phi \cos \phi \ (E \left[ \Delta t_i^1 \Delta t_j^1 \right] - 
E \left[ \Delta s_i^1 \Delta s_j^1 \right]) - 
\sin^2 \phi \ E \left[ \Delta t_i^1 \Delta s_j^1 \right]
\]
Figure 15: Correlation coefficients of skeleton points are functions of only 4 shape points.

\[ E[\Delta t_i^2 \Delta t_j^2] = \cos^2 \phi \ E[\Delta t_i^1 \Delta t_j^1] - \sin \phi \cos \phi \ (E[\Delta s_i^1 \Delta t_j^1]) + \sin^2 \phi \ E[\Delta s_i^1 \Delta s_j^1]. \]

It follows then that the covariance matrix \( \Sigma \) is completely defined by a function of only 4 data points \( \tilde{\Lambda}(x_2, x_3, x_5, x_6) = \Lambda(p', x_2, x_3, p'', x_5, x_6) \), where \( p' = (-1, 1)^T \), \( p'' = (1, 1) \).

4.2 Detection of Shape by Best Linear Fit

4.2.1 Detection of Shape from Known View

As just noted, a noisy skeleton is a collection of correlated 2 – \( D \) Gaussian random variables, whose means are collinear. To determine such a line entails finding the best linear fit of points in the plane in presence of correlated Gaussian noise. This is also equivalent to the problem of non-linear fitting in the presence of independent and identically distributed Gaussian noise.

For a known camera matrix \( P \), the problem is to associate to a noisy skeleton a specific shape. While strictly speaking the problem of reconstructing the shape from a known view is an easy exercise, we present a framework which will be used in the subsequent subsection for the case of unknown camera
parameters. Toward that end, we formulate the following hypothesis testing problem

\[ H_0 : (s_i, t_i) = (s_i^0, t_i^0) + (\Delta s_i^0, \Delta t_i^0), \quad i = 1, \ldots, 2N - 1, \]

\[ H_1 : (s_i, t_i) = (s_i^1, t_i^1) + (\Delta s_i^1, \Delta t_i^1), \quad i = 1, \ldots, 2N - 1. \]

\[ (13) \]

\[ s_i = (s_i, t_i), \quad i = 1, \ldots, 2N - 1 \]

are the observed skeleton points, and \( s_i^j = (s_i^j, t_i^j), \quad j = 0, 1 \)

are the points of the clean skeleton corresponding to each shape. The noise vector \( (\ldots, \Delta s_i^j, \ldots, \Delta t_i^j, \ldots) \)

is then Gaussian with zero mean and some known covariance matrix \( \Sigma^j \). Both the means and the

covariance matrix depend on the underlying shape, and the covariance matrices \( \Sigma^j \) in turn, depend on

the projective camera matrix \( P \).

Let us introduce the following notation

\[ \bar{s} = \begin{pmatrix} s_1 \\ \vdots \\ s_{2N-1} \\ t_1 \\ \vdots \\ t_{2N-1} \end{pmatrix}, \quad \bar{s}^j = \begin{pmatrix} s^j_1 \\ \vdots \\ s^j_{2N-1} \\ t^j_1 \\ \vdots \\ t^j_{2N-1} \end{pmatrix}, \quad j = 0, 1. \]

\[ (14) \]

The hypothesis problem becomes

\[ H_0 : \quad \mathbb{E}[\bar{s}] = \bar{s}^0, \quad \mathbb{E} \left[ (\bar{s} - \bar{s}^0) (\bar{s} - \bar{s}^0)^T \right] = \Sigma^0, \]

\[ H_1 : \quad \mathbb{E}[\bar{s}] = \bar{s}^1, \quad \mathbb{E} \left[ (\bar{s} - \bar{s}^1) (\bar{s} - \bar{s}^1)^T \right] = \Sigma^1. \]

\[ (15) \]

Let \( A^j, \quad j = 0, 1 \) be whitening matrices, that is such that

\[ (A^j)^T A^j = \Sigma^j, \quad A^j \sim (4N - 2) \times (4N - 2). \]

\[ (16) \]

Define the “whitened data” for each hypothesis

\[ \bar{s}^j_h = (A^j)^{-1} \bar{s}, \quad \bar{s}^j = \left( (A^j)^T \right)^{-1} \bar{s}^j. \]

\[ (17) \]
We then deduce that $\bar{s}_n^j$ is a Gaussian vector with $\mathbb{E} \left[ \bar{s}_n^j \right] = \bar{s}^j$ and
$\mathbb{E} \left[ \left( \bar{s}_n^j - \bar{s}^j \right) \left( \bar{s}_n^j - \bar{s}^j \right)^T \right] = I$. Upon defining the costs: $C^j \equiv C \left( \bar{s}^j \right) = \left\| \bar{s}_n^j - \bar{s}^j \right\|^2$, a decision rule may be designed as $C^0 >_{H_1} H_0 C^1$. This decision rule may be generalized to extend it to a library of $N$ shapes, so that the detector is of the form: $i_{opt} = \arg \min_{i=1,...,N} C^i$.

4.2.2 Skeleton-Based Estimation of a Shape from Known View

In many applications, the set of possible shapes is either underdetermined or so large that matching each individual shape to an observation is prohibitive and hence impractical. The focus is hence redirected on estimating the shape from its noisy observed view. With an understanding of a shape as its frontal view, the problem statement is as follows.

Let $\{x_i\}$ be noisy data points defined according to Equation (8). Assume that the projective camera matrix $P$ is given, so that the camera rotation parameters $\phi, \psi, \alpha$ are known. The problem is then to estimate the frontal view $\{x_{f_i}\}$ of a shape that matches the data.

Let $\{s_i\}$ be a noisy skeleton computed form the data points $\{x_i\}$. We propose an optimization technique, whose output is the estimate of the skeleton of the true shape corresponding to the observed data.

Denote by $l_{\phi,\psi,\alpha}$ the image of the line $\{x = 0\}$ through the camera $P(\phi, \psi, \alpha)$. Note that the skeleton points of the true shape must lie on $l_{\phi,\psi,\alpha}$. Define the skeleton by the initial guess $\{s_{0_i}\}$, obtained by projecting the noisy skeleton points on that line. The matching function (cost) of $\{s_{0_i}\}, C \left( \tilde{s}^{0,f}; \phi, \psi, \alpha \right)$, is defined analogously to that of detection, accounting for the fact that a skeleton of a frontal view defines a shape only up to a horizontal scale.

Define by $\left\{ s_{0_i}^{0,f} = \left( s_{0_i}^{0,f}, t_{0_i}^{0,f} \right) \right\}$ the skeleton of the frontal view corresponding to $\{s_{0_i}\}$, and $\tilde{s}^{0,f} = \left( s_{0,f}^{0,f}, ..., s_{2N-1}^{0,f}, t_{0_i}^{0,f}, ..., t_{2N-1}^{0,f} \right)^T$. As shown in the previous sections, any choice of $x_{0_i}^0$ together with the skeleton $\{s_{0_i}^{0,f}\}$, yields a reconstruction of the frontal view $\{x_{0_i}^f\}$. Were this reconstruction a true shape, a skeleton of its rotated view would assume the distribution $\mathcal{N} \left( \{s_{0_i}^0\}, \Sigma_{\{s_{0_i}^0\},x_{1,\phi,\psi,\alpha}} \right)$, which in
Figure 16: Finding the optimal skeleton.

turn suggests a matching cost as following

\[ C_{x_1}(\tilde{s}, \phi, \psi, \alpha) = \| \tilde{s} - \tilde{s}^0 \|^2, \]  

(18)

where \( \tilde{s} \) and \( \tilde{s}^0 \) are “whitened” vectors \( \tilde{s} \) and \( \tilde{s}^0 \). Minimizing over all possible “base points” \( x_1 \), independently of a skeleton, we get

\[ C(\tilde{s}^0; \phi, \psi, \alpha) = \min_{x_1} C_{x_1}(\tilde{s}, \phi, \psi, \alpha). \]  

(19)

The optimal skeleton is determined by minimizing the above cost, i.e. by numerically solving the problem:

\[
\left\{ \begin{array}{l}
\{ s_{i}^{\text{opt}} \} = \arg \min_{\{ s_i \}} C(\tilde{s}; \phi, \psi, \alpha) \\
\text{subject to: } \{ s_i \} \in l_{\phi, \psi, \alpha}, \forall i = 1, \ldots, 2N - 1.
\end{array} \right.
\]  

(20)

In computer simulations, the minimizations above were performed in MATLAB by using \texttt{fminsearch} routine, which is based on simplex search method \cite{19}. The initial conditions for the minimization evolutions were chosen to correspond to the frontal view. Figures 17 and 18 demonstrate two examples of estimating a shape through its skeleton from its single noisy observation.

4.2.3 Identification of Shape from Unknown View

The core of the skeleton detection algorithm is a generalized best linear fit in the presence of correlated \( 2 - D \) Gaussian noise. It was pointed out that the correlation matrix of the vector representing the
Figure 17: Shape of a wineglass (right) is reconstructed from a noisy observation (left) with known camera parameters.

Figure 18: Shape of a jar (right) is reconstructed from a noisy observation (left) with known camera parameters.
skeleton of a noisy skeleton \((s_1, \ldots, s_{2N-1})\) depends not only on a shape but also on a projective camera matrix \(P\), providing the corresponding view of the shape, i.e. different camera matrices \(P\) yield different correlation matrices \(\Sigma\).

Recall that a camera matrix depends on two groups of parameters: extrinsic and intrinsic. The intrinsic parameters, which are assumed known herein, model the inner structure of the camera, whereas the extrinsic parameters govern the location of the camera in the 3−D space with respect to the shape of interest.

The extrinsic parameters of interest are rotation angles of the shape with respect to the camera along with its distance from the camera. The distance from the camera to the object in a practical situation when a clear view of the available object is relatively easy to estimate with modern range sensing equipment. Also, while strictly speaking by varying this distance we can change the projective geometry of a view, its effect in most situations is quite insignificant and may thus be neglected. Therefore we will assume that out of all camera parameters only the three rotation parameters \(\phi, \psi\) and \(\alpha\) are unknown.

Consider a noisy view \(\{x_i\}\) of a symmetric planar shape, where we assume that the noise is Gaussian with zero-mean and a variance \(\delta^2\). The conditional distribution of its skeleton given the camera matrix, i.e. the rotation parameters, is as shown earlier, Gaussian

\[
p(s_1, \ldots, s_{2N-1}) \sim \mathcal{N}(s_0^1, \ldots, s_{2N-1}^0; \Sigma_s; \phi, \psi, \alpha, \delta^2).
\]  

(21)

We assume the rotation parameters restricted to some intervals:

\[
\phi \in (\phi_{\text{min}}, \phi_{\text{max}}), \ \psi \in (\psi_{\text{min}}, \psi_{\text{max}}), \ \alpha \in (\alpha_{\text{min}}, \alpha_{\text{max}}).
\]  

(22)

Then the shape estimation is obtained by optimizing the cost derived for a known view over all rotation angles from the admissible set defined by \(\{22\}\), i.e.

\[
\{s_i\} = \operatorname{arg\ min}_{\{s_i\}, \phi, \psi, \alpha} C(\hat{s}, \phi, \psi, \alpha);
\]  

(23)
The examples on Figures 19 and 20 demonstrate reasonable performance for realistic shapes with the landmarks manually chosen with visible errors.

5 Appendix A: Detecting Projective Symmetry of Discrete Shapes

Any algorithm that relies on a discrete representation of a shape is in need of a sampling algorithm, which would take a continuous shape and produce the desired landmarks. This in general is an extremely challenging problem, and in the case of extreme perspective transformations (cf. [20]) of the original shape and/or excessive noise (including loss of data points), it is a priori unsolvable. However, due to the impressive power of the theory of discrete shapes, this problem has been of high interest to a lot of researchers, and some significant results have been obtained (cf. [21] and the references therein).

While we hasten to emphasize that a complete development of the sampling of a continuous shape
is a separate problem, which is well beyond the scope of this paper, it is nevertheless important to address the issue of detecting a symmetry of a discrete shape based on its projective view.

First, note that in practice in a shape analysis context, the most natural way to place landmarks is where the contour exhibits high curvature (corner). This is a fairly standard technique, which has been widely used described in the literature in many variations. With a set of landmarks in hand, we may proceed to their ordering and establishing their symmetry.

Given a projective view of a generic discrete shape (Figure 11), we note that if \( x_1 \) is symmetric to \( x_4 \), \( x_2 \) is symmetric to \( x_5 \), and \( x_3 \) is symmetric to \( x_6 \), then the lines passing through symmetric points will all intersect at one point \( x_v \), which is often called the vanishing point.

5.1 Symmetry of a Frontal View

Suppose we have a frontal (but possibly noisy) view of a symmetric shape \( (x_v = (1, 0, 0)^T) \). The goal is to find the exact symmetry correspondence between the landmarks. This problem has been addressed in [15]. Here we briefly outline its solution.

1. For every pair of points, e.g. \( x_1 \) and \( x_4 \):

   (a) fold by reflecting the points about the symmetry axis, to obtain \( \tilde{x}_1 \) and \( \tilde{x}_4 \);

   (b) average \( x_1 \) and \( \tilde{x}_1 \) points to produce a single point \( \hat{x}_1 \);

   (c) reflect \( \hat{x}_1 \) back about the symmetry axis, thus obtaining \( \hat{x}_1, \hat{x}_4 \).

2. Define the disparity by \( C(\{x_i\}) = \sum_i \|x_i - \hat{x}_i\|_2 \);

3. Minimize \( C \) over all possible axes of symmetry.

   It has been proven that the above minimization problem can be solved explicitly by directly computing the optimal axis of symmetry, and thus avoiding a time consuming numerical search. This algorithm can be applied to all possible groupings of landmarks to find the one with the minimum dis-
parity. If the order of landmarks is known from the original continuous shape, sorting over all possible combinations is a problem of just linear complexity and therefore quick to solve.

5.2 Symmetry of a Projective View

Let \( \{x_i\} \) be a skewed view of a symmetric discrete shape. Then there exists a projective transformation defined by a \( 3 \times 3 \) matrix \( H \), such that \( \{\bar{x}_i\} \) is symmetric, where \( \bar{x}_i = Hx_i \). We can therefore define a symmetry disparity, \( C_p \), for a projective view as follows.

\[
C_p(\{x_i\}) = \min_H C(\{Hx_i\}).
\]

(24)

As in the previous case, the correct symmetric correspondence will in theory translate to the minimum value of \( C_p \). Because a matrix of a projective transformation is defined up to a multiplicative constant, this minimization problem is over 8 variables, which can be solved numerically using the direct search method, cf. [19].

Below we show the examples of skewed views of different shapes along with the symmetry axis found by the algorithm outlined above. Our experiments have shown that while the symmetric pairs are identified correctly, the exact location of the axis of symmetry is recovered with an error which is due to both the noise in the data points and the imperfections of the numerical solution of the optimization problem (24). It therefore is impossible to use the “optimal” projection \( H \) for the purpose of direct reconstruction of the original shape from the view.

References


Figure 21: Detecting symmetry of a generic airplane.

Figure 22: Detecting symmetry of a jar.


