



A nonparametric approach to 3D shape analysis from digital camera images – I

V. Patrangenaru^{a,*}, X. Liu^a, S. Sugathadasa^b

^a Florida State University, United States

^b Texas Tech University, United States

ARTICLE INFO

Article history:

Received 25 August 2008

Available online 9 March 2009

To the Memory of W.P. Dayawansa

AMS subject classification:

primary 62H11

secondary 62H10

62H35

Keywords:

Pinhole camera images

High level image analysis

3D reconstruction

Projective shape

Extrinsic means

Asymptotic distributions on manifolds

Nonparametric bootstrap

Confidence regions

ABSTRACT

This article for the first time develops a nonparametric methodology for the analysis of projective shapes of configurations of landmarks on real 3D objects from their regular camera pictures. A fundamental result in computer vision, emulating the principle of human vision in space, claims that, generically, a finite 3D configuration of points can be retrieved from corresponding configurations in a pair of camera images, up to a projective transformation. Consequently, the projective shape of a 3D configuration can be retrieved from two of its planar views, and a projective shape analysis can be pursued from a sample of images. Projective shapes are here regarded as points on projective shape manifolds. Using large sample and nonparametric bootstrap methodology for extrinsic means on manifolds, one gives confidence regions and tests for the mean projective shape of a 3D configuration from its 2D camera images. Two examples are given: an example of testing for accuracy of a simple manufactured object using mean projective shape analysis, and a face identification example. Both examples are data driven based on landmark registration in digital images.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Until now, statistical analysis of similarity shape from images has been restricted to a small amount of data, since similarity shape appearance is relative to the camera position with respect to the scene pictured. In this paper, for the first time, we study the shape of a 3D configuration from its 2D images in pictures of this configuration, without requiring any restriction for the camera positioning vs. the scene pictured. Our nonparametric methodology is manifold based, and uses standard reconstruction methods in computer vision. In the absence of occlusions, a set of point correspondences in two views can be used to retrieve the 3D configuration of points. A key result due to Faugeras [1] and to Hartley et al. [2] states that two such reconstructions differ by a projective transformation in three dimensions. Sugathadasa [3] noticed that actually the object which is recovered without ambiguity is the projective shape of the configuration. This casts a new light on the role of projective shape in the identification of a spatial configuration.

Projective shape is a natural approach to shape analysis from digital images, since the vast majority of libraries of images are acquired via a central projection from the scene pictured to the black box recording plane. Hartley and Zisserman [4, p. 1] note that “this often renders classical shape analysis of a spatial scene impossible, since similarity is not preserved when a camera is moving.” Advances in statistical analysis of projective shape have been slowed down due to the

* Corresponding address: Department of Statistics, Florida State University, Statistics, 32306 Tallahassee, FL, United States.

E-mail address: vic@stat.fsu.edu (V. Patrangenaru).

overemphasized importance of similarity shape in image analysis, while ignoring the basic principle of image acquisition. Progress was also hampered by the lack of a geometric model for the space of projective shapes, and ultimately by insufficient dialogue between researchers in geometry, computer vision and statistical shape analysis.

For the reasons presented above, projective shapes have been studied only recently, and except for one concrete 3D example due to Sugathadasa [3], to be found in [5], the literature including [6–14,31,33] was bound to linear or planar projective shape analyzes.

Our main goal here is to derive a natural concept of 3D projective shape that can be extracted from data recorded from regular camera images. Our statistical methodology for estimation of a mean 3D projective shape is nonparametric, based on Efron's bootstrap [30]. In this paper, a 3D projective shape is regarded as a random object on a projective shape space. Since typically samples of images are small, in order to estimate the mean projective shape we use nonparametric bootstrap for the studentized sample mean projective shape on a manifold, as shown in [15]. This bootstrap distribution was essentially derived in [12].

A summary by sections follows. Section 2 is devoted to a recollection of basic geometry facts needed further in the paper, such as projective invariants, projective frames, and projective coordinates from [9,12]. We then introduce the essential matrix (respectively, the fundamental matrix) associated with a pair of a camera views of a 3D scene that is needed in the reconstruction of that scene from both 2D calibrated and noncalibrated camera images. For reconstruction of a configuration of points in space from its views in a pair of images we refer to computational algorithms in [16]. The section concludes with the Faugeras–Hartley–Chang–Gupta projective ambiguity theorem for the scene reconstructed from two noncalibrated camera views, which is the key point for our projective shape analysis of spatial configurations.

In Section 3 we introduce projective shapes of configurations of k points in \mathbb{R}^m or in $\mathbb{R}P^m$, and the multivariate axial geometric model for the projective shape space, which is our choice for a statistical study of projective shapes. The Faugeras–Hartley–Chang–Gupta theorem is reformulated in Theorem 3.1 as a result on projective shapes: the 3D reconstruction \mathcal{R} of a spatial configuration \mathcal{C} from two ideal camera views of \mathcal{C} , and \mathcal{C} itself have the same projective shape. This opens the statistical shape analysis door to computer vision and pattern recognition of 3D scenes, including face analysis.

Since projective shape spaces are identified via projective frames with products of axial spaces, in Section 4 we approach multivariate axial distributions via a quadratic equivariant embedding of a product of $k - m - 2$ copies of $\mathbb{R}P^m$ in products of spaces of symmetric matrices. A theorem on the asymptotic distributions of extrinsic sample means of multivariate axes, stated without proof and with a minor typo in [12], is given here (Theorem 4.1) and its proof is provided in Appendix A. The asymptotic and nonparametric bootstrap distribution results are used to derive confidence regions for extrinsic mean projective shapes. If a random projective shape has a nondegenerated extrinsic covariance matrix one may studentize the extrinsic sample mean to generate asymptotically chi-square distributions that are useful for large sample confidence regions for mean projective shapes in Corollary 4.2 or nonparametric bootstrap confidence regions if the sample is small in Corollary 4.4. If the extrinsic covariance matrix is degenerated, and the axial marginals have nondegenerated extrinsic covariance matrices, one gives a Bonferroni type of argument for axial marginals to derive confidence regions for the mean projective shape in Corollary 4.5.

Section 5 is dedicated to concrete applications. In a first example, applicable in manufacturing, we test if a simple 3D polyhedral object, built from three cubes of different sizes, matches the designed blueprint for this object. The 2D coordinates of these vertices were recorded in [3] and for convenience are listed in Appendix B. This leads to a one sample test for the mean projective shape of a configuration of 19 vertices that are visible in a sample of 16 random images of the object. In a second example, we consider a data set that was previously used in [12] to emphasize the role of 2D projective shape in face identification. In this example we consider eight facial landmarks that are not coplanar, to verify that one quarter views of a face can be also used to identify a person. The standard practice for shape based identification of a face was to use only frontal or lateral images. Although the sample sizes are very small, the computations of nonparametric bootstrap confidence regions of the mean projective shapes of the eight point configurations for frontal and for the one quarter views are very encouraging for inclusion of arbitrary views of a face in the analysis.

2. Basic projective geometry for ideal pinhole camera image acquisition

Image acquisition from the 3D world to the 2D camera film is based on a central projection principle; therefore projective geometry governs the physics of ideal pinhole cameras. A point in the outer space, and its central projection via the camera pinhole, determine a unique line in space; therefore an image point captured on the camera film can be regarded as a line going through the pinhole, leading to the definition of the real projective plane $\mathbb{R}P^2$ as the space of all lines going through the origin of \mathbb{R}^3 . Projective geometry also provides a logical justification for the mental reconstruction of a spatial scene from binocular retinal images, playing a central role in vision. In this section we review some of the basics of projective geometry that are useful in understanding of image formation and 3D scene retrieval from a pair of ideal pinhole camera images.

2.1. Basics of projective geometry

Consider a real vector space V , and let 0_V be the zero of this vector space. Two vectors $x, y \in V \setminus \{0_V\}$ are equivalent if they differ by a scalar multiple. The equivalence class of $x \in V \setminus \{0_V\}$ is labeled $[x]$, and the set of all such equivalence classes is the projective space $P(V)$ associated with V , $P(V) = \{[x], x \in V \setminus \{0_V\}\}$. The real projective space in m dimensions, $\mathbb{R}P^m$, is $P(\mathbb{R}^{m+1})$. Another notation for a projective point $p = [x] \in \mathbb{R}P^m$, the equivalence class of $x = (x^1, \dots, x^{m+1}) \in \mathbb{R}^{m+1}$,

is $p = [x^1 : x^2 : \dots : x^{m+1}]$ featuring the *homogeneous coordinates* (x^1, \dots, x^{m+1}) of p , which are determined up to a multiplicative constant. As a general notation rule, coordinates of a point in \mathbb{R}^{m+1} are separated by commas within parentheses and homogeneous coordinates of the corresponding projective point are separated by the column symbols “:” within squared brackets. That is, $[x^1 : \dots : x^{m+1}] = [(x^1, \dots, x^{m+1})]$. A projective point p admits also a *spherical representation*, when thought of as a pair of antipodal points on the m dimensional unit sphere, $p = \{z, -z\}$, $z = (z^1, z^2, \dots, z^{m+1})$, $(z^1)^2 + \dots + (z^{m+1})^2 = 1$. A d -dimensional *projective subspace* of $\mathbb{R}P^m$ is a projective space $P(V)$, where V is a $(d + 1)$ -dimensional vector subspace of \mathbb{R}^{m+1} . An $(m - 1)$ -dimensional projective subspace of $\mathbb{R}P^m$ is also called *hyperplane*. The *linear span* of a subset D of $\mathbb{R}P^m$ is the smallest projective subspace of $\mathbb{R}P^m$ containing D . We say that k points in $\mathbb{R}P^m$ are in *general position* if the linear span of their set is $\mathbb{R}P^m$. If k points in $\mathbb{R}P^m$ are in general position, then $k \geq m + 2$.

The numerical space \mathbb{R}^m can be *embedded* in $\mathbb{R}P^m$, preserving collinearity. An example of such an *affine embedding* is

$$h((u^1, \dots, u^m)) = [u^1 : \dots : u^m : 1] = [\tilde{u}], \tag{2.1}$$

where $\tilde{u} = (u^1, \dots, u^m, 1)$. The complement of the range of the embedding h in (2.1) is the set $\mathbb{R}P^{m-1}$ of points $[x^1 : \dots : x^m : 0] \in \mathbb{R}P^m$.

Conversely, the *inhomogeneous (affine) coordinates* (u^1, \dots, u^m) of a point $p = [x^1 : x^2 : \dots : x^{m+1}] \in \mathbb{R}P^m \setminus \mathbb{R}P^{m-1}$ are given by

$$u^j = \frac{x^j}{x^{m+1}}, \quad \forall j = 1, \dots, m. \tag{2.2}$$

Consider a matrix $B \in M(m + 1, m' + 1; \mathbb{R})$ and the linear subspace $K = \{x \in \mathbb{R}^{m'+1}, Bx = 0\}$ of $\mathbb{R}^{m'}$. The *projective map* $\beta : \mathbb{R}P^{m'} \setminus P(K) \rightarrow \mathbb{R}P^m$ associated with B is defined by $\beta([x]) = [Bx]$. In particular, a *projective transformation* β of $\mathbb{R}P^m$ is the projective map associated with a nonsingular matrix $B \in GL(m + 1, \mathbb{R})$ and its action on $\mathbb{R}P^m$:

$$\beta([x^1 : \dots : x^{m+1}]) = [B(x^1, \dots, x^{m+1})^T]. \tag{2.3}$$

If u (respectively, v) are the affine coordinates (inverse of the affine embedding (2.1)) of $[x]$ (respectively, of $\beta([x])$), then the equation of the projective transformation (2.3) in affine coordinates is given by

$$v^j = \frac{a_{m+1}^j + \sum_{i=1}^m a_i^j u^i}{a_{m+1}^{m+1} + \sum_{i=1}^m a_i^{m+1} u^i}, \quad \forall j = 1, \dots, m \tag{2.4}$$

where $B = ((a_i^j)_{i,j=1,\dots,m+1})$, and $\det(B) \neq 0$. An *affine transformation* of \mathbb{R}^m , $v = Au + b$, $A \in GL(m, \mathbb{R})$, $b \in \mathbb{R}^m$, is a particular case of projective transformation α , associated with the matrix $B \in GL(m + 1, \mathbb{R})$, where

$$B = \begin{pmatrix} A & b \\ 0_m & 1 \end{pmatrix}. \tag{2.5}$$

A *projective frame* in an m -dimensional projective space (or *projective basis* in the computer vision literature, see e.g. [17]) is an ordered set of $m + 2$ projective points in general position. An example of projective frame in $\mathbb{R}P^m$ is the *standard projective frame* $([e_1], \dots, [e_{m+1}], [e_1 + \dots + e_{m+1}])$.

In projective shape analysis it is preferable to employ coordinates invariant with respect to the group $PGL(m)$ of projective transformations. A projective transformation takes a projective frame to a projective frame, and its action on $\mathbb{R}P^m$ is determined by its action on a projective frame; therefore if we define the *projective coordinate(s)* of a point $p \in \mathbb{R}P^m$ w.r.t. a projective frame $\pi = (p_1, \dots, p_{m+2})$ as being given by

$$p^\pi = \beta^{-1}(p), \tag{2.6}$$

where $\beta \in PGL(m)$ is a projective transformation taking the standard projective frame to π , these coordinates automatically have the invariance property.

Remark 2.1. Assume u, u_1, \dots, u_{m+2} are points in \mathbb{R}^m , such that $\pi = ([\tilde{u}_1], \dots, [\tilde{u}_{m+2}])$ is a projective frame. If we consider the $(m + 1) \times (m + 1)$ matrix $U_m = [\tilde{u}_1^T, \dots, \tilde{u}_{m+1}^T]$, the projective coordinates of $p = [\tilde{u}]$ w.r.t. π are given by

$$p^\pi = [y^1(u) : \dots : y^{m+1}(u)], \tag{2.7}$$

where

$$v(u) = U_m^{-1} \tilde{u}^T \tag{2.8}$$

and

$$y^j(u) = \frac{v^j(u)}{v^j(u_{m+2})}, \quad \forall j = 1, \dots, m + 1. \tag{2.9}$$

Let z^j be given by

$$z^j = \frac{y^j}{\|y^j\|}. \quad (2.10)$$

Note that, in our notation, the superscripts are reserved for the components of a point whereas the subscripts are for the labels of points. The projective coordinate(s) of x are given by the point $[z^1(x) : \dots : z^{m+1}(x)] \in \mathbb{R}P^m$.

2.2. Image acquisition in ideal digital cameras

Ideal pinhole camera image acquisition can be thought of in terms of a central projection β from $\mathbb{R}P^3 \setminus \mathbb{R}P^2$ to $\mathbb{R}P^2$, whose representation in conveniently selected affine coordinates $(x, y, z) \in \mathbb{R}^3$, $(u, v) \in \mathbb{R}^2$ is given by

$$\begin{aligned} u &= -f \frac{x}{z} \\ v &= -f \frac{y}{z}, \end{aligned} \quad (2.11)$$

where f is the focal length, the distance from the image sensor or film to the pinhole or principal plane of the lens $\mathbb{R}P^2$, the complement of the domain of β . In homogeneous coordinates $[x : y : z : w]$, $[u : v : t]$ the perspective projective map β can be represented by the matrix $B \in M(3, 4; \mathbb{R})$ given by

$$B = \begin{pmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (2.12)$$

In addition to the projective map (2.12), image formation in digital cameras assumes a composition with matrices accounting for camera internal calibration parameters, such as the pixel aspect ratio, skew parameter, origin of image coordinates in the principal plane (principal point) and for a change of coordinates between two camera positions involving a roto-translation $(R, t) \in SO(3) \times \mathbb{R}^3$. The projective map of pinhole camera image acquisition $\tilde{\pi}$, in homogeneous coordinates, is associated with the matrix

$$\tilde{B} = C_{\text{int}} B E = \begin{pmatrix} k_u & k_c & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} R & t \\ 0_3^T & 1 \end{pmatrix} = N E, \quad (2.13)$$

where k_u and k_v are scale factors of the image plane in units of the focal length f , $\theta = \cot^{-1} k_c$ is the skew, and (u_0, v_0) is the principal point. The matrix N contains the internal parameters and the perspective map (2.12), while E contains the external parameters. The matrix \tilde{B} can be decomposed into a 3×3 matrix P and a 3×1 vector p :

$$\tilde{B} = \begin{pmatrix} P & p \end{pmatrix} \quad (2.14)$$

so that

$$P = AR \quad \text{and} \quad p = At \quad [34]. \quad (2.15)$$

2.3. Essential and fundamental matrices

We consider now a pair of cameras viewing a point $[u] \in \mathbb{R}P^3$. This point projects onto the two image planes to $m_1 = [u_1] \in \mathbb{R}P^2$ respectively to $m_2 = [u_2] \in \mathbb{R}P^2$. Since we are working in homogeneous coordinates, $[u]$ is represented by a 4×1 column vector, and m_1, m_2 are each represented by 3×1 column vectors. If we assume the camera's internal parameters are known (the camera is calibrated), then m_1, m_2 are given each with respect to its camera's coordinate frame; therefore $C_{\text{int}} = I_3$.

Definition 2.1. The epipolar constraint refers to the fact that the vector from the first camera's optical center to the first imaged point, the vector from the second optical center to the second imaged point, and the vector from one optical center to the other are all coplanar.

If we use only one coordinate system, say the coordinate system of the second camera, the vector from the first camera's optical center to the first imaged point is $t + Ru_1$, the vector from the second optical center to the second imaged point is u_2 , and the vector from one optical center to the other is t . Here the change of coordinates between the Euclidean frames of the two cameras is given by a roto-translation $(R, t) \in SO(3) \times \mathbb{R}^3$. The epipolar constraint can be expressed via a zero exterior product $u_2 \wedge (t + Ru_1) \wedge t = 0$, which is equivalent to

$$u_2^T (t \times (Ru_1)) = 0. \quad (2.16)$$

By defining t_\times as the matrix associated with the linear operator $y \rightarrow t \times y$ we can rewrite equation (2.16) as follows:

$$u_2^T (t_\times (Ru_1)) = u_2^T E u_1 = 0, \quad (2.17)$$

where $E = t \times R$ is the so-called *essential matrix*. If the cameras are uncalibrated, then the matrices A_1, A_2 from (2.15) containing the internal parameters of the two cameras are needed to transform the camera bound Euclidean coordinates into pixel coordinates:

$$\begin{aligned} v_1 &= A_1 u_1 \\ v_2 &= A_2 u_2. \end{aligned}$$

This yields the following equation:

$$(A_2^{-1} v_2)^T (T \times R A_1^{-1} v_1) = v_2^T A_2^{-1} (T \times R A_1^{-1} v_1) = 0, \tag{2.18}$$

and we obtain

$$v_2^T F v_1 = 0, \tag{2.19}$$

where $F = (A_2^{-1})^T E A_1^{-1}$ is the *fundamental matrix*. The fundamental matrix depends only on the relative position of the two cameras, and on their internal parameters. It has rank two, depending on seven real constants.

2.4. Reconstruction of a 3D scene from two of its 2D images

If we select conveniently the coordinates for the first camera, incorporating the internal parameters, we may assume that the matrix associated with β_1 in (2.13) is $B_1 = (I|0)$ and the fundamental matrix factors as $F = [t] \times R$, with $B_2 = (R|t)$ corresponding to β_2 , being a realization of the fundamental (or essential) matrix F (here R is nonsingular, and it does not necessarily represent the matrix of a rotation). Let $[u_1], [u_2] \in \mathbb{R}P^2$ be a pair of matched points in the two images. We seek a point $[u] \in \mathbb{R}P^3$ such that $[u_i] = \tilde{\beta}_i[u], i = 1, 2$. From the relation $u_2^T F u_1 = u_2^T t \times R u_1 = u_2^T (t \times R u_1) = 0$, it follows that $u_2, R u_1, t$ are linearly dependent. We may assume that $R u_1 = b u_2 - a t$, and since the position vector u_1 is defined up to a scalar multiple, we may assume that $R u_1 = u_2 - a t$ and define the corresponding landmark position $[u] \in \mathbb{R}P^3$ by $u = (u_1^T, a)^T$. Now $B_1 u = (I|0)u = u_1$, and $B_2 u = (R|t)u = R u_1 + a t = u_2$; therefore if β_1, β_2 are the projections associated with B_1, B_2 , it follows that $\beta_a([u]) = [u]_a, a = 1, 2$ and $[u]$ is a desired solution to the reconstruction problem. As shown, $[u]$ is determined by the two camera projection matrices B_1 and B_2 . If we choose a different pair of camera matrices $B_1 H$ and $B_2 H$ realizing the same fundamental matrix F , then in order to preserve the same pair of matched image points, the point $[u]$ must be replaced by $[H^{-1}u]$.

Problem 2.1. The problem of the reconstruction of a configuration of points in three dimensions from two ideal noncalibrated camera images with unknown camera parameters is equivalent to the following: given two camera images $\mathbb{R}P_1^2, \mathbb{R}P_2^2$ of unknown relative position and internal camera parameters and two matching sets of labeled points $\{p_{a,1}, \dots, p_{a,k}\} \subset \mathbb{R}P_a^2, a = 1, 2$, find all the sets of points in space p_1, \dots, p_k such that there exist two positions of the planes $\mathbb{R}P_1^2, \mathbb{R}P_2^2$ and internal parameters of the two cameras $c_a, a = 1, 2$ with the property that the c_a -image of p_j is $p_{a,j}, \forall a = 1, 2, j = 1, \dots, k$.

The above discussion proves the following theorem [1,2]:

Theorem 2.2. *The reconstruction problem for two noncalibrated camera images has a solution in terms of the realization of the fundamental matrix $F = t \times R$. Any two solutions can be obtained from each other by a projective transformation in $\mathbb{R}P^3$.*

Remark 2.2. Note that although the configurations in correspondence are finite, their size is arbitrarily large, and the assumption of finite matching labeled pairs can be replaced by an assumption of parameterized sets in correspondence; therefore in the absence of occlusions, a 3D configuration can be reconstructed from 2D images, and this reconstruction is unique up a projective transformation.

2.5. Estimation of the fundamental matrix

Since equation (2.19) is homogeneous as a linear equation in F and F has rank two, this matrix depends on seven independent parameters; therefore, in principle, F can be recovered from the corresponding configurations of seven points. Due to the fact that the nature of digital imaging data is inherently discrete and other errors in landmark registration, F can be estimated using configurations of eight or more points $p_{a,i}, a = 1, 2, i = 1, \dots, k, k \geq 8$, whose stacked homogeneous coordinates are the $k \times 3$ matrices $y_a, a = 1, 2$. The linear system for F is

$$y_2^T F y_1 = 0. \tag{2.20}$$

This can be written as

$$f^T Y = 0, \tag{2.21}$$

where f is a vectorized form of F . If k is large the linear homogeneous system is overdetermined and the optimal estimated solution \hat{f} can be obtained using a simple least squares algorithm by minimizing $\|Y^T f\|^2$ subject to $\|f\| = 1$ (see [4, p. 593]).

3. Projective shape and 3D reconstruction

Definition 3.1. Two configurations of points in \mathbb{R}^m have the same *projective shape* if they differ by a projective transformation of \mathbb{R}^m .

Unlike similarities or affine transformations, projective transformations of \mathbb{R}^m do not have a group structure under composition (the domain of definition of the composition of two such maps is smaller than the maximal domain of a projective transformation in \mathbb{R}^m). To avoid this complication, rather than considering the projective shapes of configurations in \mathbb{R}^m , we consider projective shapes of configurations in $\mathbb{R}P^m$. A *projective shape* of a k -ad (configuration of k landmarks or labeled points) is the orbit of that k -ad under projective transformations with respect to the diagonal action

$$\alpha_k(p_1, \dots, p_k) = (\alpha(p_1), \dots, \alpha(p_k)). \quad (3.1)$$

Since the action (2.3) of $\beta \in PGL(m)$ on $[x] \in \mathbb{R}P^m$, when expressed in inhomogeneous coordinates (2.2) reduces to (2.4), if two configurations Γ_1, Γ_2 of points in \mathbb{R}^m have the same projective shape, then $h(\Gamma_1), h(\Gamma_2)$ have the same projective shape in $\mathbb{R}P^m$ (h is the affine embedding given by (2.1)).

Patrangenaru [18,9] considered the set $G(k, m)$ of k -ads (p_1, \dots, p_k) , $k > m + 2$ for which $\pi = (p_1, \dots, p_{m+2})$ is a projective frame. $PGL(m)$ acts simply transitively on $G(k, m)$ and the projective shape space $P\Sigma_m^k$ is the quotient $G(k, m)/PGL(m)$. Using the projective coordinates $(p_{m+3}^T, \dots, p_k^T)$ given by (2.6) one can show that $P\Sigma_m^k$ is a manifold diffeomorphic with $(\mathbb{R}P^m)^{k-m-2}$. The projective frame representation is an alternative to the projective invariants based representation, used earlier for projective shape analysis by Goodall and Mardia [8]. The projective frame representation has two useful features: firstly, the projective shape space has a manifold structure, thus allowing one to use the asymptotic theory for means on manifolds in [19,15], and secondly, it can be extended to infinite-dimensional projective shape spaces, such as projective shapes of curves, as shown in [14]. This approach has also the advantage of being inductive in the sense that each new landmark of a configuration adds an extra marginal axial coordinate, thus allowing one to detect its overall contribution to the variability of the configuration as well as correlation to the other landmarks. The effect of change of projective coordinates, due to projective frame selection, can be understood via a group of projective transformations, but is beyond the scope of this paper.

Remark 3.1. An approach to projective shape that is invariant with respect to the group of permutations on landmark indices, initiated by Kent and Mardia [13,31] has not yet been used in practical applications beyond dimension 1, possibly due to the fact that it requires nonlinear approximations to the matrix solution of the equation in A

$$A = \frac{m}{k} \sum_{i=1}^k \frac{x_i x_i^T}{x_i^T A^{-1} x_i} \quad (3.2)$$

in terms of a k -ad of points in $\mathbb{R}P^m$ given in their spherical representation, the key step in the Kent and Mardia description of a projective shape.

We return to the reconstruction of a spatial configuration. Having in view Definition 3.1 of a projective shape of a configuration, Theorem 2.2 can be stated as follows:

Theorem 3.1. A spatial \mathcal{R} reconstruction of a 3D configuration \mathcal{C} can be obtained in the absence of occlusions from two of its ideal camera views. Any such 3D reconstruction \mathcal{R} of \mathcal{C} has the same projective shape as \mathcal{C} .

Remark 3.2. Since the output in a reconstruction algorithm is a projective shape, and multiplying by an imposed internal camera parameters matrix keeps the projective shape of the reconstruction unchanged, one may use the essential matrix estimate using the eight point algorithm in [16, p. 121], for a conveniently selected internal parameters matrix. Refined eight point algorithms for the estimate \hat{F} of the fundamental matrix that can be found in [16, p. 188, p. 395] could also be used, given the projective equivalence of any two 3D reconstructions.

4. Nonparametric estimation and testing for the projective shape of a 3D configuration

In general, if $f : M_1 \rightarrow M_2$ is a differentiable function defined from the manifold M_1 to the manifold M_2 and $x \in M_1$, the differential of the function f at x is labeled $D_x f$. Assume $J : M \rightarrow \mathbb{R}^N$ is an embedding of the d -dimensional complete manifold M [20]. Bhattacharya and Patrangenaru [19] defined the extrinsic mean μ_J of a J -nonfocal random object (r.o.) Y on M by

$$\mu_J =: J^{-1}(P_J(\mu)), \quad (4.1)$$

where $\mu = E(J(Y))$ is the mean vector of $J(Y)$ and $P_J : \mathcal{F}^c \rightarrow J(M)$ is the ortho-projection on $J(M)$ defined on the complement of the set \mathcal{F} of focal points of $J(M)$. The extrinsic covariance matrix of Y with respect to a local frame field $y \rightarrow (f_1(y), \dots, f_d(y))$ for which $(D_y J(f_1(y)), \dots, D_y J(f_d(y)))$ are orthonormal vectors in \mathbb{R}^N was defined in [15]. If Σ is the covariance matrix of $J(Y)$ regarded as a random vector on \mathbb{R}^N , then P_J is differentiable at μ . In order to evaluate the differential $D_\mu P_J$ one considers a special orthonormal frame field to ease the computations. A local ortho-

frame field $(e_1(p), e_2(p), \dots, e_N(p))$ defined on an open neighborhood $U \subseteq \mathbb{R}^N$ of $P_j(M)$ is adapted to the embedding J if $\forall y \in J^{-1}(U), e_r(J(y)) = D_y J(f_r(y)), r = 1, \dots, d$. Let e_1, e_2, \dots, e_N be the canonical basis of \mathbb{R}^N and assume $(e_1(p), e_2(p), \dots, e_N(p))$ is an adapted frame field around $P_j(\mu) = J(\mu_j)$. Then Σ_E given by

$$\Sigma_E = \left[\sum_{a=1}^d D_\mu P_j(e_b) \cdot e_a(P_j(\mu)) e_a(P_j(\mu)) \right]_{b=1, \dots, N} \Sigma \left[\sum_{a=1}^d D_\mu P_j(e_b) \cdot e_a(P_j(\mu)) e_a(P_j(\mu)) \right]_{b=1, \dots, N}^T \tag{4.2}$$

is the extrinsic covariance matrix of Y with respect to $(f_1(\mu_j), \dots, f_d(\mu_j))$. Since $P \Sigma_m^k$ is homeomorphic to $M = (\mathbb{R}P^m)^q, q = k - m - 2$ and since $\mathbb{R}P^m$, as a particular case of a Grassmann manifold is equivariantly embedded in the space $S(m + 1)$ of $(m + 1) \times (m + 1)$ symmetric matrices [21] via $j : \mathbb{R}P^m \rightarrow S(m + 1)$,

$$j([x]) = xx^T \tag{4.3}$$

Mardia and Patranganaru [12] considered the resulting equivariant embedding

$$J = j_k : P \Sigma_m^k = (\mathbb{R}P^m)^q \rightarrow (S(m + 1))^q$$

defined by

$$j_k([x_1], \dots, [x_q]) = (j([x_1]), \dots, j([x_q])), \tag{4.4}$$

where $x_s \in \mathbb{R}^{m+1}, x_s^T x_s = 1, \forall s = 1, \dots, q$.

Remark 4.1. The embedding j_k in (4.4) yields the fastest known computational algorithms in projective shape analysis. Basic axial statistics related to Watson’s method of moments such as the sample mean axis [22] and extrinsic sample covariance matrix [23] can be expressed in terms of $j_{m+3} = j$.

A random projective shape Y of a k -ad in $\mathbb{R}P^m$ is given in axial representation by the multivariate random axes

$$(Y^1, \dots, Y^q), Y^s = [X^s], (X^s)^T X^s = 1, \quad \forall s = 1, \dots, q, q = k - m - 2. \tag{4.5}$$

From [19] or [12] it follows that, in this representation, the extrinsic mean projective shape of (Y^1, \dots, Y^q) exists if $\forall s = 1, \dots, q$, the largest eigenvalue of $E(X^s(X^s)^T)$ is simple. In this case μ_{jk} is given by

$$\mu_{jk} = ([\gamma_1(m + 1)], \dots, [\gamma_q(m + 1)]) \tag{4.6}$$

where $\lambda_s(a)$ and $\gamma_s(a), a = 1, \dots, m + 1$ are the eigenvalues in increasing order and the corresponding unit eigenvector of $E(X^s(X^s)^T)$.

If $Y_r, r = 1, \dots, n$ are i.i.d.r.o.’s (independent identically distributed random objects) from a population of projective shapes (in its multi-axial representation), for which the mean shape μ_{jk} exists, from a general consistency theorem for extrinsic means on manifolds in [19] it follows that the extrinsic sample mean $[\bar{Y}]_{jk,n}$ is a strongly consistent estimator of μ_{jk} . In the multivariate axial representation

$$Y_r = ([X_r^1], \dots, [X_r^q]), (X_r^s)^T X_r^s = 1; \quad s = 1, \dots, q. \tag{4.7}$$

Let J_s be the random symmetric matrix given by

$$J_s = n^{-1} \sum_{r=1}^n X_r^s (X_r^s)^T, \quad s = 1, \dots, q, \tag{4.8}$$

and let $d_s(a)$ and $g_s(a)$ be the eigenvalues in increasing order and the corresponding unit eigenvector of $J_s, a = 1, \dots, m + 1$. Then the sample mean projective shape in its multi-axial representation is given by

$$\bar{Y}_{jk,n} = ([g_1(m + 1)], \dots, [g_q(m + 1)]). \tag{4.9}$$

Remark 4.2. Some of the results in this section can be found without a proof in [12]. Their proofs are given in [32] and here in Appendix A.

If a is a positive integer, $\overline{1, a}$ is the set of indices from 1 to a . To determine the extrinsic covariance matrix (4.2) of (4.5), we note that the vectors

$$f_{(s,a)} = (0, \dots, 0, \gamma_s(a), 0, \dots, 0), \tag{4.10}$$

with the only nonzero term in position $s, s \in \overline{1, q}, a \in \overline{1, m}$ yielding a basis in the tangent space at the extrinsic mean $T_{\mu_{jk}}(\mathbb{R}P^m)^q$, that is orthonormal with respect to the scalar product induced by the embedding j_k . The vectors $e_{(s,a)}, \forall s \in \overline{1, q}, \forall a \in \overline{1, m}$, defined as follows:

$$e_{(s,a)} =: D_{\mu_{jk}} j_k(f_{(s,a)}), \tag{4.11}$$

form an orthobasis of $T_{j_k(\mu_{j_k})}(\mathbb{R}P^m)^q$. We complete this orthobasis to an orthobasis of q -tuples of matrices $(e_i)_{i \in \mathcal{I}}$ for $(S(m+1))^q$, that is indexed by the set \mathcal{I} , the first indices of which are the pairs (s, a) , $s = 1, \dots, q$; $a = 1, \dots, m$ in their lexicographic order. Let E_a^b be the $(m+1) \times (m+1)$ matrix with all entries zero, except for an entry 1 in the position (a, b) . The standard basis of $S(m+1)$ is given by $e_a^b = E_a^b + E_b^a$, $1 \leq a \leq b \leq m+1$. For each $s = 1, \dots, q$, the vector $({}_s e_a^b) = (0_{m+1}, \dots, 0_{m+1}, e_a^b, 0_{m+1}, \dots, 0_{m+1})$ has all the components zero matrices $0_{m+1} \in S(m+1)$, except for the s -th component, which is the matrix e_a^b of the standard basis of $S(m+1, \mathbb{R})$; the vectors ${}_s e_a^b$, $s = 1, \dots, q$, $1 \leq a \leq b \leq m+1$ listed in the lexicographic order of their indices (s, a, b) give a basis of $(S(m+1))^q$.

Let Σ be the covariance matrix of $j_k(Y^1, \dots, Y^q)$ regarded as a random vector in $(S(m+1))^q$, with respect to this standard basis, and let $P =: P_{j_k} : (S(m+1))^q \rightarrow j_k((\mathbb{R}P^m)^q)$ be the projection on $j_k((\mathbb{R}P^m)^q)$. From (4.2) it follows that the extrinsic covariance matrix of (Y^1, \dots, Y^q) with respect to the basis (4.10) of $T_{\mu_{j_k}}(\mathbb{R}P^m)^q$ is given by

$$\Sigma_E = [e_{(s,a)}(P(\mu)) \cdot D_{\mu}P(r e_a^b)]_{(s=1,\dots,q),(a=1,\dots,m)} \cdot \Sigma \cdot [e_{(s,a)}(P(\mu)) \cdot D_{\mu}P(r e_a^b)]_{(s=1,\dots,q),(a=1,\dots,m)}^T \tag{4.12}$$

Assume Y_1, \dots, Y_n are i.i.d.r.o.'s from a j_k -nonfocal probability measure on $(\mathbb{R}P^m)^q$ and μ_{j_k} in (4.6) is the extrinsic mean of Y_1 . We arrange the pairs of indices (s, a) , $s = 1, \dots, q$; $a = 1, \dots, m$, in their lexicographic order, and define the $(mq) \times (mq)$ symmetric matrix G_n , with the entries

$$G_{n(s,a),(t,b)} = n^{-1}(d_s(m+1) - d_s(a))^{-1}(d_t(m+1) - d_t(b))^{-1} \times \sum_{r=1}^n (g_s(a)^T X_r^s)(g_t(b)^T X_r^t)(g_s(m+1)^T X_r^s)(g_t(m+1)^T X_r^t). \tag{4.13}$$

Lemma 4.1. G_n is the extrinsic sample covariance matrix estimator of Σ_E .

From [15] it follows that G_n is a strongly consistent estimator of the population extrinsic covariance matrix in (4.12). In preparation for an asymptotic distribution of $\bar{Y}_{j_k,n}$ we set

$$D_s = (g_s(1) \cdots g_s(m)) \in \mathcal{M}(m+1, m; \mathbb{R}), s = 1, \dots, q. \tag{4.14}$$

If $\mu = ([\gamma_1], \dots, [\gamma_q])$, where $\gamma_s \in \mathbb{R}^{m+1}$, $\gamma_s^T \gamma_s = 1$, for $s = 1, \dots, q$, we define a Hotelling's T^2 -type statistic

$$T(\bar{Y}_{j_k,n}; \mu) = n(\gamma_1^T D_1, \dots, \gamma_q^T D_q) G_n^{-1} (\gamma_1^T D_1, \dots, \gamma_q^T D_q)^T. \tag{4.15}$$

Theorem 4.1. Assume $(Y_r)_{r=1,\dots,n}$ are i.i.d.r.o.'s on $(\mathbb{R}P^m)^q$, and Y_1 is j_k -nonfocal, with $\Sigma_E > 0$. Let $\lambda_s(a)$ and $\gamma_s(a)$ be the eigenvalues in increasing order and corresponding unit eigenvectors of $E[X_1^a (X_1^a)^T]$. If $\lambda_s(1) > 0$, for $s = 1, \dots, q$, then $T(\bar{Y}_{j_k,n}; \mu_{j_k})$ converges weakly to χ_{mq}^2 .

If Y_1 is a j_k -nonfocal population on $(\mathbb{R}P^m)^q$, since $(\mathbb{R}P^m)^q$ is compact, it follows that $j_k(Y_1)$ has finite moments of sufficiently high order. According to Bhattacharya and Ghosh [24], this, along with an assumption of a nonzero absolutely continuous component, suffices to ensure an Edgeworth expansion up to order $O(n^{-2})$ of the pivotal statistic $T(\bar{Y}_{j_k,n}; \mu_{j_k})$, and implicitly the bootstrap approximation of this statistic.

Corollary 4.1. Let $Y_r = ([X_r^1], \dots, [X_r^q])$, $X_{st}^T X_{st} = 1$, $s = 1, \dots, q$, $r = 1, \dots, n$, be i.i.d.r.o.'s from a j_k -nonfocal distribution on $(\mathbb{R}P^m)^q$ which has a nonzero absolutely continuous component, and with $\Sigma_E > 0$. For a random resample with repetition (Y_1^*, \dots, Y_n^*) from (Y_1, \dots, Y_n) , consider the eigenvalues of $\frac{1}{n} \sum_{r=1}^n X_{rs}^* X_{rs}^{*T}$ in increasing order and corresponding unit eigenvectors $d_s^*(a)$ and $g_s^*(a)$, $a = 1, \dots, m+1$. Let G_n^* be the matrix obtained from G_n , by substituting all the entries with $*$ -entries. Then the bootstrap distribution function of the statistic

$$T(\bar{Y}_{j_k}^*; \bar{Y}_{j_k}) = n(g_1(m+1)^T D_1^*, \dots, g_q(m+1)^T D_q^*) G_n^{*-1} (g_1(m+1)^T D_1^*, \dots, g_q(m+1)^T D_q^*)^T \tag{4.16}$$

approximates the true distribution of $T(\bar{Y}_{j_k}; \mu_{j_k})$ given by (4.15), with an error of order $O_p(n^{-2})$.

Remark 4.3. The above corollary is from [12]. The condition $\Sigma_E > 0$ is missing there as well as in their Theorem 4.1.

Theorem 4.1 and Corollary 4.1 are useful in estimation and testing for mean projective shapes. From Theorem 4.1 we derive a large sample confidence region for μ_{j_k} .

Corollary 4.2. Assume $(Y_r)_{r=1,\dots,n}$ are i.i.d.r.o.'s from a j_k -nonfocal probability distribution on $(\mathbb{R}P^m)^q$, and $\Sigma_E > 0$. An asymptotic $(1 - \alpha)$ -confidence region for $\mu_{j_k} = [v]$ is given by $R_\alpha(\mathbf{Y}) = \{[v] : T(\bar{Y}_{j_k,n}; [v]) \leq \chi_{mq,\alpha}^2\}$, where $T(\bar{Y}_{j_k}, [v])$ is given in (4.15). If the probability measure of Y_1 has a nonzero-absolutely continuous component w.r.t. the volume measure on $(\mathbb{R}P^m)^q$, then the coverage error of $R_\alpha(\mathbf{Y})$ is of order $O_p(n^{-1})$.

For small samples the coverage error could be quite large, and the bootstrap analogue in Corollary 4.1 is preferable. Consider for example the one sample testing problem for mean projective shapes:

$$H_0 : \mu_{jk} = \mu_0 \text{ vs. } H_1 : \mu_{jk} \neq \mu_0. \tag{4.17}$$

Corollary 4.3. *The large sample p-value for the testing problem (4.17) is $p = \Pr(T > T(\bar{Y}_{jk,n}; \mu_0))$, where $T(\bar{Y}_{jk,n}; \mu)$ is given by (4.15).*

In the small sample case, problem (4.17) can be answered based on Corollary 4.1 to obtain the following $100(1 - \alpha)\%$ bootstrap confidence region for μ_{jk} :

Corollary 4.4. *Under the hypotheses of Corollary 4.1, the corresponding $100(1 - \alpha)\%$ confidence region for μ_{jk} is*

$$C_{n,\alpha}^* := j_k^{-1}(U_{n,\alpha}^*) \tag{4.18}$$

with $U_{n,\alpha}^*$ given by

$$U_{n,\alpha}^* = \{\mu \in j_k((\mathbb{R}P^m)^q) : T(\bar{Y}_{jk,n}; \mu) \leq c_{1-\alpha}^*\}, \tag{4.19}$$

where $c_{1-\alpha}^*$ is the upper $100(1 - \alpha)\%$ point of the values of $T(\bar{Y}_{jk}^*; \bar{Y}_{jk})$ given by (4.16). The region given by (4.18) and (4.19) has coverage error $O_p(n^{-2})$.

If Σ_E is singular and all the marginal axial distributions have positive definite extrinsic covariance matrices, one may use simultaneous confidence ellipsoids to estimate μ_{jk} . Assume $(Y_r)_{r=1,\dots,n}$ are i.i.d.r.o.'s from a j_k -nonfocal probability distribution on $(\mathbb{R}P^m)^q$. For each $s = 1, \dots, q$ let Σ_s be the extrinsic covariance matrix of Y_1^s , and let $\bar{Y}_{j,n}^s$ and $G_{s,n}$ be the extrinsic sample mean and the extrinsic sample covariance matrix of the s -th marginal axial and assume the probability measure of Y_1^s has a nonzero-absolutely continuous component w.r.t. the volume measure on $\mathbb{R}P^m$. For $s = 1, \dots, q$ and for $[\gamma_s] \in \mathbb{R}P^m$, $\gamma_s^T \gamma_s = 1$, we consider the statistics:

$$T_s = T_s(\bar{Y}_{j,n}^s, [\gamma_s]) = n\gamma^T D_s C_{s,n}^{-1} D_s^T \gamma \tag{4.20}$$

and the corresponding bootstrap distributions:

$$T_s^* = T_s(\bar{Y}_j^{s*}, \bar{Y}_{j,n}^s) = n g_s(m+1)^T D_s^* G_{s,n}^{-1} D_s^{*T} g_s(m+1). \tag{4.21}$$

Since by Theorem 4.1 T_s has asymptotically a χ_m^2 distribution, we obtain the following:

Corollary 4.5. *For $s = 1, \dots, q$ let $c_{s,1-\beta}^*$ be the upper $100(1 - \beta)\%$ point of the values of T_s^* given by (4.21). We set*

$$C_{s,n,\beta}^* := j_k^{-1}(U_{s,n,\beta}^*) \tag{4.22}$$

where

$$U_{s,n,\beta}^* = \{\mu \in \mathbb{R}P^m : T_s(\bar{Y}_{j,n}^s; \mu) \leq c_{s,1-\beta}^*\}. \tag{4.23}$$

Then

$$R_{n,\alpha}^* = \bigcap_{s=1}^q C_{s,n,\frac{\alpha}{q}}^* \tag{4.24}$$

with $C_{s,n,\beta}^*$, $U_{s,n,\beta}^*$ given by (4.22)–(4.23) is a region of approximately at least $100(1 - \alpha)\%$ confidence for μ_{jk} . The coverage error is of order $O_p(n^{-2})$.

Remark 4.4. If Σ_E is singular one may also use a method for constructing nonpivotal bootstrap confidence regions for μ_{jk} using Corollary 5.1 of [19].

5. Applications

5.1. Estimation of the mean 3D projective shape of a polyhedral surface from its images

Theorem 2.2 lays down the geometric principle of binocular vision that triggers 3D perception in combination with certain neurological mechanisms. The projective ambiguity of the reconstructed scene stated in that theorem was first recognized by Faugeras [1] and by Hartley et al. [2]. In this paper it is assumed that cameras are *noncalibrated*, meaning that images of the same scene are available and nothing is known about the cameras' internal parameters that recorded these images, or about the cameras' relative positions. Since initially in computer vision the reconstruction algorithms were designed for calibrated cameras [25], the projective ambiguity of the reconstructed scene for noncalibrated cameras was perceived as a lack of information, and research was directed mostly towards camera calibration.

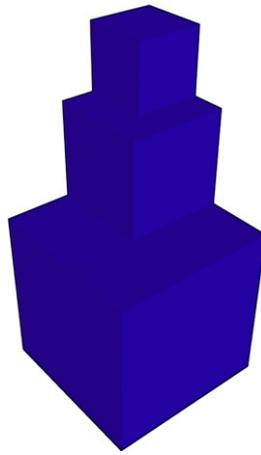


Fig. 1. 3D blueprint of a 3D object.

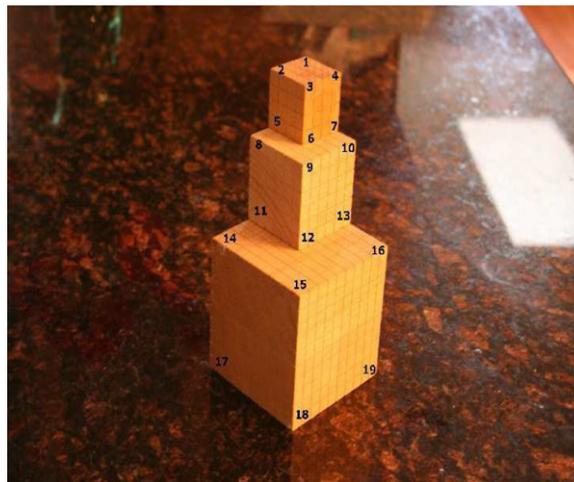


Fig. 2. Nineteen landmarks displayed that are used for 3D scene reconstruction, and statistical analysis.

Remark 5.1. Theorem 3.1 sheds a new light in pattern recognition of a 3D scene imaged by pairs of noncalibrated cameras, since the projective shape of a reconstructed configuration does not depend on the reconstruction scheme, so the projective shape already provides useful information about the imaged 3D scene, an important point in scene identification that was previously ignored in the literature.

Due to landmark registration, camera distortion or rounding errors in the reconstruction algorithms, 3D projective shapes from pairs of images can be regarded as random objects on the projective shape space $P\Sigma_3^k$. In our first example, we consider a 3D polyhedral object manufactured from three cubes based on a blueprint displayed in Fig. 1. The object is manufactured from three cubes that sit on the top of each other, whose sides from top to bottom are four, six and ten units. To see if the projective shape of the object matches the original blueprint, one takes a number of random pictures of the object, to recover its 3D projective shape. We assume the faces of the object are flat and consequently its visible surface is determined by the visible corners. Fig. 2 displays a digital image of the object with visible corners, taken as landmarks, numbered from 1 to 19. Sixteen randomly selected pictures of the object, that show all the selected landmarks, were paired into eight pairs of images. Recording of landmark coordinates of camera image pairs was done using the Matlab commands *imread* and *cpselect*. Sixteen images of the object are displayed in Fig. 3. The 2D coordinates of the landmarks selected are and listed in Table 4 in Appendix B.

Using a reconstruction algorithm from [16, p. 121] for each pair of corresponding planar configurations, we obtain a 3D reconstructed configuration. Homogeneous coordinates of the reconstructed configurations are given in Appendix B. The sample of 3D configurations of points that are joined to resemble the original object is displayed in Fig. 4:

We selected landmarks 8, 12, 17, 18, and 19 in this order to form a projective frame, and with respect to this frame, the projective coordinates of the other landmarks in their original labeling order yield a sample of points in $(\mathbb{R}P^3)^{14}$ ($19 - 3 - 2 = 14$), each point representing a projective shape in axial representation. The extrinsic mean of the eight

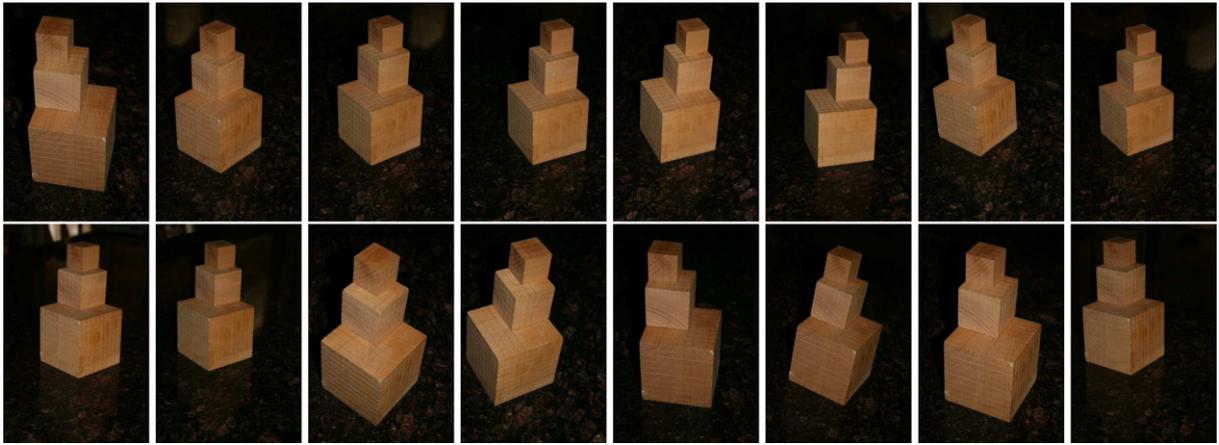


Fig. 3. Sixteen digital camera views of a 3D object resembling the blueprint.

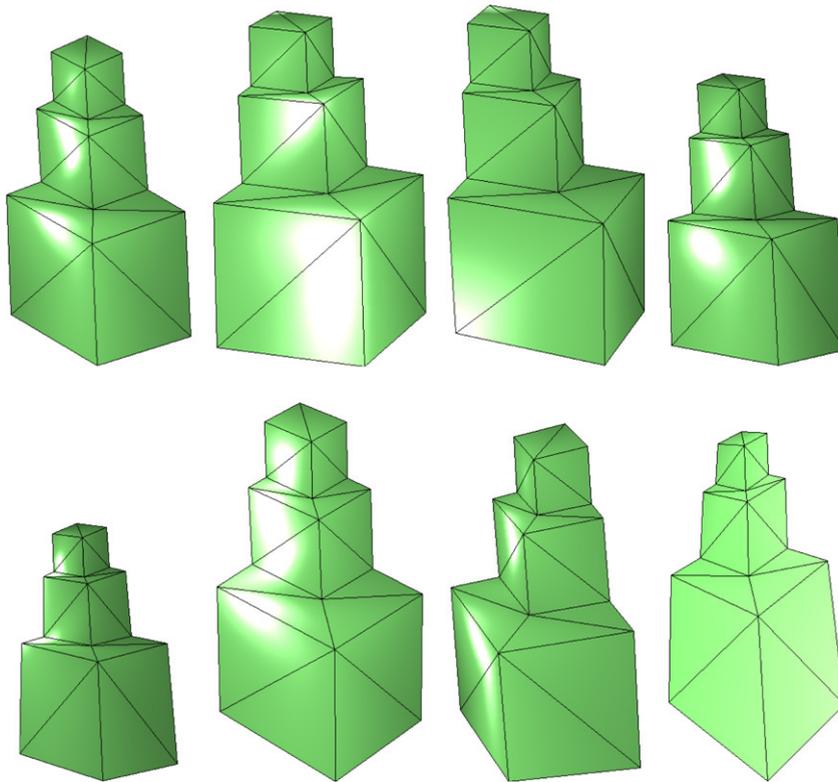


Fig. 4. The eight reconstructed 3D projective shapes.

Table 1
Extrinsic sample mean.

Proj. Sp. Copy		1	2	3	4	5	6	7	9	10	11	13	14	15	16
j_{19} Ext. Samp. Mean	x	0.21	0.29	-0.29	-0.25	0.54	-0.10	-0.07	-0.45	-0.28	0.85	-0.00	0.65	-0.43	-0.32
	y	0.75	0.67	0.46	0.63	0.66	0.57	0.51	0.10	0.40	0.41	0.24	0.10	-0.41	0.08
	z	-0.41	0.36	0.46	-0.50	0.08	0.10	-0.77	0.64	-0.79	0.02	-0.96	0.71	0.77	-0.93
	w	0.47	0.58	0.70	0.54	0.51	0.81	0.38	0.62	0.37	0.32	0.15	0.24	0.22	0.15

projective shapes was computed; a configuration having that projective shape is displayed in Fig. 5(a). The spherical coordinates of the landmarks are given in Table 1. Given the large number of covariates in the tangent space (42), we display only the “heat map” of the extrinsic covariance matrix (4.13), here shown in Fig. 6.

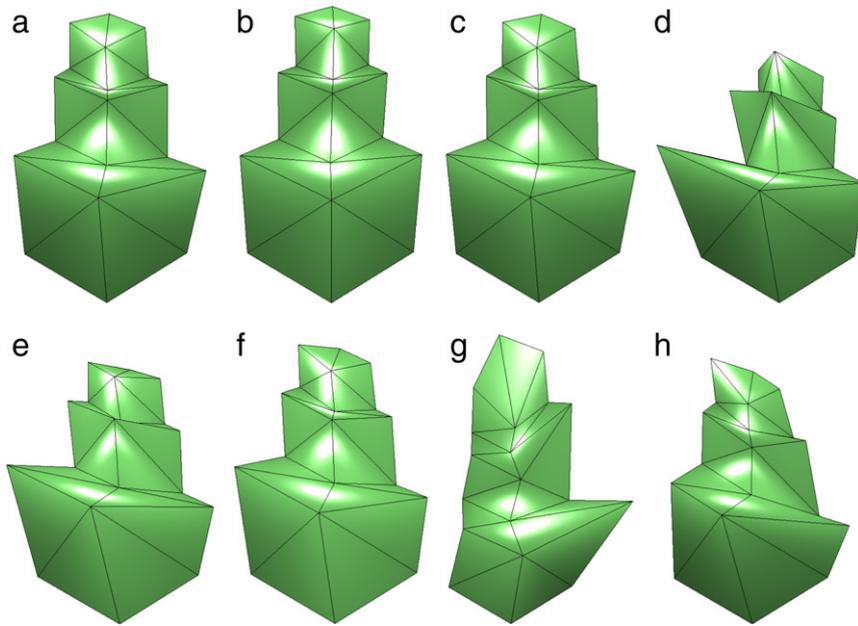


Fig. 5. (a) The estimated extrinsic mean shown as a 3D projective shape. (b)–(h) Randomly selected estimated extrinsic means based on bootstrap samples. In each case, a projective transformation is applied to each shape so that landmarks 13, 15, 17, 18, and 19 match that of reconstructed 3D image 1 as shown in Fig. 4 in \mathbb{R}^3 .

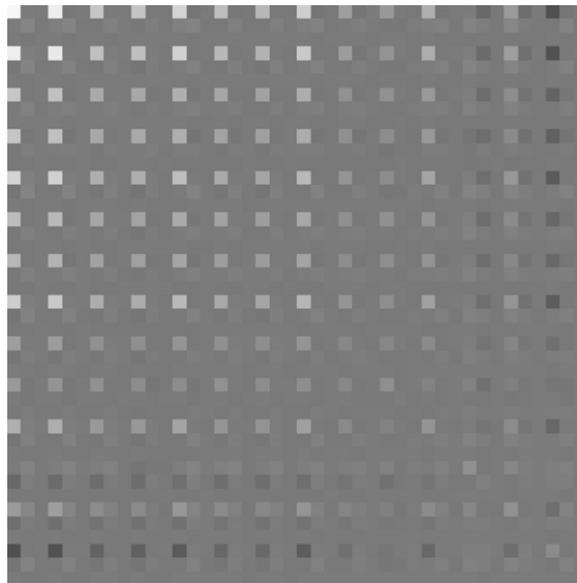


Fig. 6. Extrinsic sample covariance matrix shown as an image.

We formulate the original question as a hypothesis testing problem

$$H_0 : \mu_{j_{19}} = \mu_0 \text{ vs. } H_1 : \mu_{j_{19}} \neq \mu_0, \quad (5.1)$$

where μ_0 is the projective shape of the blueprint (Table 3), given in (5.1) and $\mu_{j_{19}}$ is the extrinsic mean projective shape of the random 3D configuration of 19 vertices on the polyhedral surface of the object as recovered from its pictures, corresponding to the embedding j_{19} (Table 2).

It suffices to check that, for a significant level α , μ_0 is in a $(1 - \alpha)100\%$ confidence region of $\mu_{j_{19}}$. Since the sample size (8) is too small, the extrinsic sample covariance matrix is a degenerate matrix, so one cannot use a pivotal bootstrap for the mean projective shape of the entire configuration of 19 landmarks as given in Corollary 4.4. Nevertheless we could find $(1 - \alpha)100\%$ confidence regions using a pivotal bootstrap, based on Corollary 4.5. There are $q = 14$ marginal axial distributions that correspond to the 14 existing landmarks beside those that are part of the projective frame. To achieve a

Table 2
Coordinates of the blueprint.

Landmark no.		1	2	3	4	5	6	7	8	9	10
Blue print	x	10.00	10.00	0.00	10.00	0.00	0.00	4.00	4.00	0.00	4.00
	y	0.00	10.00	10.00	10.00	6.00	0.00	0.00	4.00	4.00	0.00
	z	0.00	0.00	0.00	10.00	10.00	20.00	20.00	20.00	20.00	16.00
	w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Landmark no.		11	12	13	14	15	16	17	18	19	
Blue print	x	4.00	0.00	6.00	6.00	0.00	6.00	6.00	10.00	0.00	
	y	4.00	4.00	0.00	6.00	6.00	0.00	6.00	0.00	10.00	
	z	16.00	16.00	16.00	16.00	16.00	10.00	10.00	10.00	10.00	
	w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	

Table 3
Projective coordinates of the blueprint.

Landmark no.		1	2	3	4	5	6	7
Blue print	x	0.22	0.26	-0.31	-0.26	0.52	-0.09	-0.05
	y	0.77	0.66	0.46	0.66	0.67	0.58	0.58
	z	-0.33	0.39	0.46	-0.39	0.09	0.13	-0.68
	w	0.50	0.59	0.69	0.59	0.52	0.80	0.45
Landmark no.		9	10	11	13	14	15	16
Blue print	x	-0.42	-0.31	0.85	0.00	0.68	-0.47	-0.39
	y	0.11	0.46	0.42	0.31	0.11	-0.47	0.10
	z	0.64	-0.69	0.00	-0.92	0.68	0.70	-0.89
	w	0.64	0.46	0.32	0.23	0.25	0.26	0.22

reliable conclusion, we used 20 000 resamples from the original sample. For example, μ_0 is in the 95% confidence region for $\mu_{j_{19}}$, if for each $s = 1, \dots, 14$ the value of $T_s = T(\bar{Y}_{j,7}^s; \mu_{0,s})$ in (4.15), corresponding to the s -th marginal is between the 72nd ranked and the 19 928th ranked observation of the corresponding bootstrap distribution (values of degenerated $G_{s,7}^*$ have been omitted). The results are as follows:

- First marginal (Landmark 1): $T_1 = 3.0279647168E+00$ is between 6301 ($T^* = 3.0243210949E+00$) and 6302 ($T^* = 3.0294218108E+00$).
- Second marginal (Landmark 2): $T_2 = 2.6459766362E+00$ is between 3942 ($T^* = 2.6434475892E+00$) and 3943 ($T^* = 2.6920988816E+00$).
- Third marginal (Landmark 3): $T_3 = 1.5175491E-01$ is between 397 ($T^* = 1.4510789E-01$) and 398 ($T^* = 1.5271147E-01$).
- Fourth marginal (Landmark 4): $T_4 = 3.7407490E+00$ is between 7379 ($T^* = 3.7288447E+00$) and 7380 ($T^* = 3.7464786E+00$).
- Fifth marginal (Landmark 5): $T_5 = 2.6168385E+00$ is between 5355 ($T^* = 2.6166643704E+00$) and 5356 ($T^* = 2.6216985741E+00$).
- Sixth marginal (Landmark 6): $T_6 = 1.7898784E+00$ is between 3294 ($T^* = 1.7859106E+00$) and 3295 ($T^* = 1.7914946E+00$).
- Seventh marginal (Landmark 7): $T_7 = 3.9364703E+00$ is between 7194 ($T^* = 3.9191776E+00$) and 7195 ($T^* = 3.9388019E+00$).
- Eighth marginal (Landmark 9): $T_8 = 1.5700171E+00$ is between 4432 ($T^* = 1.5687626E+00$) and 4433 ($T^* = 1.5748148E+00$).
- Ninth marginal (Landmark 10): $T_9 = 5.0491394E+00$ is between 8507 ($T^* = 5.0407173E+00$) and 8508 ($T^* = 5.0521943E+00$).
- Tenth marginal (Landmark 11): $T_{10} = 1.3735517E+01$ is between 15155 ($T^* = 1.3706638E+01$) and 15156 ($T^* = 1.3750192E+01$).
- Eleventh marginal (Landmark 13): $T_{11} = 2.0352336E+00$ is between 4198 ($T^* = 2.0327469E+00$) and 4199 ($T^* = 2.0412641E+00$).
- Twelfth marginal (Landmark 14): $T_{12} = 3.9488573860E+00$ is between 6837 ($T^* = 3.9442688E+00$) and 6838 ($T^* = 3.9573099E+00$).
- Thirteenth marginal (Landmark 15): $T_{13} = 3.6973595941E+00$ is between 7857 ($T^* = 3.6946285E+00$) and 7858 ($T^* = 3.6986865E+00$).
- Fourteenth marginal (Landmark 16): $T_{14} = 2.8730067E+00$ is between 5065 ($T^* = 2.8723605E+00$) and 5066 ($T^* = 2.8770536E+00$).

These results show that we fail to reject H_0 for any reasonable level α , thus proving that the projective shape of the object is following the projective shape of the blueprint closely.

Simultaneous confidence intervals for affine coordinates of the extrinsic mean projective shape, using a nonpivotal bootstrap as in Remark 4.4, yield similar results, but are less reliable given the poorer coverage error of the nonparametric bootstrap. Such results can be provided by the authors on request and can be found in [5] for a similar data set with different cube relative sizes.

5.2. Face identification example

Our second example is in the area of face recognition. Our example is based on a data set used in a live BBC program “Tomorrow’s World”. The example was introduced in [12], where six landmarks (ends of eyes plus ends of lips) have been recorded from 14 digital images of the same person (an actor posing in different disguises), in 14 pictures. The face appearance in these pictures may be neither frontal or lateral.

The data set made available to us, henceforth called actor data, has eight frontal images and seven one quarter images of the actor’s face. In [12] seven of the frontal pictures (respectively, seven one quarter pictures) were used, for which the coordinates of six anatomically landmarks that are approximately coplanar were recorded (four corners of the eyes canthus and two end points of the lips mouth edge points). Using the four eye-corner landmarks as the projective frame, the landmarks coordinates were converted into bivariate axial observations. An empirical test was performed in that paper showing evidence that the frontal and one quarter views of the group of landmarks could be of the same person.

In this paper we use two additional landmarks (“bridge of the nose” and “tip of the nose”). The eight landmarks considered are significantly not coplanar, as shown by Balan et al. [29], therefore a 3D projective shape analysis is more appropriate for this configuration. If one compares the nonparametric bootstrap distributions of the extrinsic sample mean 2D projective shape of a configuration of five points, in one quarter views versus frontal views of the actor, we notice that, even for close to coplanar configurations, these regions have only a small overlap. In Fig. 8 one may notice this effect when the fifth landmark called “bridge of the nose” is added to a configuration of four coplanar landmarks, showing the limitations of the 2D projective shape analysis of spatial scenes.

In Fig. 7, on the top row, we display the eight frontal pictures in the actor data, and six one quarter images of the same individual, in order to reconstruct 3D configurations from pairs of images.

Landmark	1	2	3	4	5	6	7	8
Image								
1	466, 403	469, 191	350, 501	608, 501	278, 191	397, 199	554, 200	665, 202
2	482, 385	483, 164	328, 552	602, 570	254, 148	40, 161	598, 167	730, 173
3	511, 358	533, 170	371, 500	600, 531	308, 136	428, 157	616, 178	742, 188
4	482, 334	513, 162	365, 462	589, 480	299, 140	412, 154	577, 169	701, 185
5	505, 260	511, 84	394, 422	628, 430	302, 83	421, 89	599, 83	719, 91
6	530, 328	539, 142	389, 450	605, 468	344, 133	457, 152	619, 155	728, 173
7	465, 264	478, 56	338, 457	620, 454	263, 98	385, 103	580, 113	700, 120
8	491, 382	503, 164	343, 494	593, 521	287, 143	430, 172	580, 181	727, 199
9	208, 238	300, 59	224, 422	421, 423	157, 94	239, 98	391, 106	497, 109
10	155, 295	240, 94	139, 495	389, 515	91, 94	193, 122	376, 128	494, 139
11	167, 303	244, 129	161, 449	365, 468	127, 113	196, 128	353, 136	475, 148
12	196, 280	256, 91	152, 447	401, 482	92, 101	203, 113	373, 117	514, 130
13	690, 484	683, 268	412, 569	620, 662	449, 212	568, 257	728, 310	842, 349
14	195, 315	240, 141	250, 477	439, 446	152, 173	224, 155	350, 140	454, 124

The matched configurations from 1 to 8 are for frontal views and from 9 to 14 for one quarter views. The 3D reconstruction was obtained for each pair of the images, with 1 and 2 being the first pair, 3 and 4 being the second pair, and so on. In total there are four reconstructed 3D facial landmark configurations using the front views and three reconstructed 3D faces using the side views. The 3D projective projective frames are given by landmarks 1 to 5.

Remark 5.2. Note that an asymptotic chi-square two sample test statistic for the hypothesis testing problem $\mu_{j_{19},1} = \mu_{j_{19},2}$ similar with the statistic T_{n_1,n_2} in formula (3.16) in [26] for the equality of two extrinsic means on $\mathbb{C}P^{k-2}$ may be derived also on $(\mathbb{R}P^m)^{k-m-2}$. However such a test statistic assumes that the sample extrinsic covariance matrices are nonsingular, while in our example the sample sizes $n_1 = 4, n_2 = 3$ are too small to insure this property for these 9×9 matrices.

Therefore we are giving only a heuristic computational justification for the equality of the extrinsic mean projective shapes of the reconstructed configurations from frontal images (respectively, from one quarter images), similar to the one used in [12]. We use a nonpivotal bootstrap as mentioned in Remark 4.4. The affine coordinates of the extrinsic sample mean 3D projective shapes of the configurations of eight landmarks retrieved from the side images falls inside seven out of nine 95% simultaneous bootstrap confidence intervals for the affine coordinates of the extrinsic mean 3D projective shapes of the corresponding configurations retrieved from the frontal images. The joint 95% confidence regions for the two means overlap.



Fig. 7. BBC data: 14 views of an actor's face.

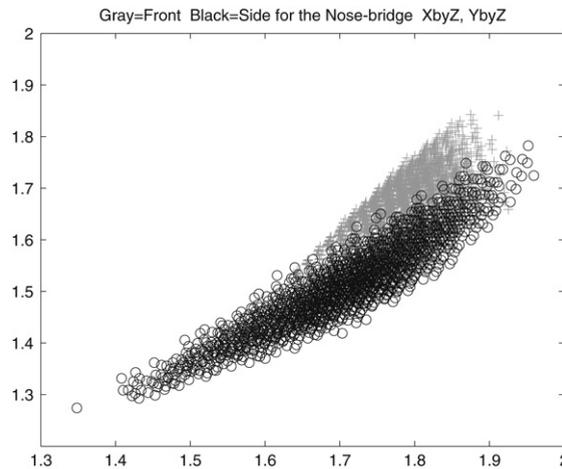


Fig. 8. Affine views of the bootstrap distribution of the extrinsic mean axial coordinate corresponding to the “bridge of nose” landmark. Frontal views = +. One quarter views = o.

Simultaneous 95% bootstrap confidence intervals for the affine coordinates of the extrinsic mean 3D projective shapes of spatial configurations of landmarks obtained from pairs of frontal images (circles) (respectively, from pairs of one quarter images (crosses)) are displayed in Fig. 9. On the horizontal axis we have the nine affine coordinates (three for each projective space marginal of the multi-axial representation). For each affine coordinate, the simultaneous confidence intervals are displayed on the vertical of that coordinate. Note that, according to Hall [27, p. 283], if n is the sample size, the number of atoms for the nonparametric bootstrap is $\binom{2n-1}{n}$. In our example, $n = 4$ for reconstructions from frontal images, and we have $\binom{7}{4} = 35$ bootstrap atoms, and $n = 3$ for reconstructions from frontal images, and for this group we have only $\binom{5}{3} = 10$ bootstrap atoms. The affine coordinates of bootstrap means in the two groups are displayed in Fig. 9 as small circles (respectively, as small crosses).

The intersection interval along each of the affine coordinates is given below:

The intersection interval for the 1st coordinate is 0.099812 0.642128 with length 0.542315.

The intersection interval for the 2nd coordinate is -0.160658 0.450349 with length 0.611007.

The intersection interval for the 3rd coordinate is 0.185366 0.292291 with length 0.106926.

The intersection interval for the 4th coordinate is -0.002991 0.657261 with length 0.660252.

The intersection interval for the 5th coordinate is -0.376397 0.983902 with length 1.360299.

The intersection interval for the 6th coordinate is -1.092520 -0.001708 with length 1.090812.

The intersection interval for the 7th coordinate is -0.002792 1.131982 with length 1.134774.

The intersection interval for the 8th coordinate is -0.769301 1.474715 with length 2.244016.

The intersection interval for the 9th coordinate is -1.992983 -0.008885 with length 1.984098.

Remark 5.3. No matter what model of 3D projective shape space is used, the dimension of that manifold is $3k - 15$. Therefore, for each additional landmark, ideally one should increase the sample size by at least six additional images. For this reason, Corollary 4.5, based on the multivariate axial model, offers a reasonable way of bypassing this high-dimensionality–small-sample problem, as seen in our application. This is a promising feature of our methodology for solving pattern recognition questions from digital images of a scene.

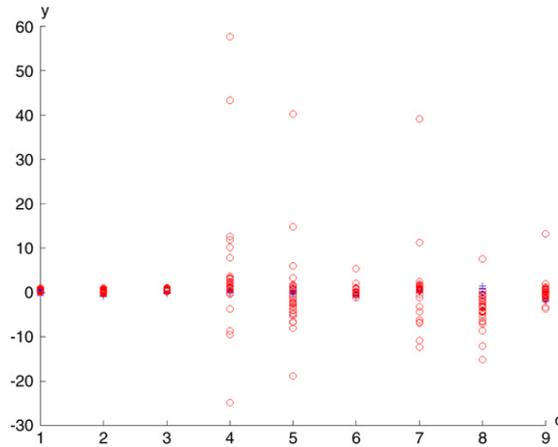


Fig. 9. 95% simultaneous bootstrap confidence intervals for affine coordinates of mean 3D projective shapes of eight facial landmarks in front views and one quarter views.

Acknowledgments

The authors wish to thank the National Security Agency, the National Science Foundation and the National Institutes of Health for their generous support. We are grateful to Rabi N. Bhattacharya, the Associate Editor and two anonymous Referees for constructive comments and suggestions that helped improve our manuscript significantly. We also thank Ying Wang and Michael Crane for helping with data registration, and some computations. The first author’s research was supported by National Science Foundation Grant DMS-0805977 and the National Security Agency Grant MSP-H98230-08-1-0058. The second author’s work was partially funded by NSF grants CCF-0514743 and DMS-0713012, and the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 RR021813. Information on the National Centers for Biomedical Computing can be obtained from <http://nihroadmap.nih.gov/bioinformatics>.

Appendix A

Proof. The proof of Lemma 4.1 is based on the equivariance of the embedding j_k . As a preliminary step note that the group $SO(m + 1)$ acts as a group of isometries of $\mathbb{R}P^m$. If $R \in SO(m + 1)$ and $[x] \in \mathbb{R}P^m$ then the action $R([x]) = [Rx]$ is well defined. $SO(m + 1)$ acts by isometries also on $S_+(m + 1, \mathbb{R})$ via $R(A) = RAR^T$. Note that the map $j(x) = xx^T$ is equivariant since

$$j(R[x]) = j([Rx]) = (Rx)(Rx)^T = Rj([x])R^T = R(j([x])).$$

Therefore, for $q > 1$ the group $(SO(m + 1))^q$ acts as a group of isometries of $(\mathbb{R}P^m)^q$ and also

$$(R_1, \dots, R_q) \cdot (A_1, \dots, A_q) = (R_1A_1R_1^T, \dots, R_qA_qR_q^T), \quad R_j \in SO(m + 1), j = 1, \dots, q. \tag{A.1}$$

and the map j_k is equivariant with respect to this action since

$$j_k((R_1, \dots, R_q) \cdot ([x_1], \dots, [x_q])) = (R_1, \dots, R_q) \cdot j_k([x_1], \dots, [x_q]), \\ \forall (R_1, \dots, R_q) \in (SO(m + 1))^q, \forall ([x_1], \dots, [x_q]) \in (\mathbb{R}P^m)^q. \tag{A.2}$$

We set $M = j_k((\mathbb{R}P^m)^q)$. Let M_1^{m+1} be the set of all matrices of rank 1 in $S_+(m + 1)$. Note that M is the direct product of q copies of M_1^{m+1} . Recall that

$$P : (S_+(m + 1, \mathbb{R}))^q \rightarrow M \tag{A.3}$$

is the projection on M . If $Y_r = ([X_r^1], \dots, [X_r^q])$, $r = 1, \dots, n$, are i.i.d.r.o.’s from a probability distribution on $(\mathbb{R}P^m)^q$, we set

$$V_r = j_k(Y_r).$$

From the equivariance of j_k , w.l.o.g. (without loss of generality) we may assume that $\overline{j_k(Y)} = \tilde{D} = (\tilde{D}_1, \dots, \tilde{D}_q)$, where $\tilde{D}_s \in S_+(m + 1, \mathbb{R})$ is a diagonal matrix, $s = 1, \dots, q$. Therefore

$$\overline{Y}_{j_k, n} = ([g_1(m + 1)], \dots, [g_q(m + 1)])$$

and $\forall s = 1, \dots, q, \forall a = 1, \dots, m + 1$, with $g_s(a) = e_a$ are the eigenvectors of \tilde{D}_s .

Table 4
Coordinates of landmarks from camera images.

Landmark no.		1	2	3	4	5	6	7	8	9	10
Image 1	x	89.50	72.50	142.00	154.00	75.50	144.00	156.50	68.00	171.00	188.50
	y	29.50	66.00	73.00	37.50	125.50	134.50	96.00	146.50	159.50	99.50
Image 2	x	135.00	97.50	137.50	174.50	99.00	138.50	175.00	80.00	138.50	194.00
	y	39.00	54.50	70.50	53.50	105.00	122.00	104.00	116.00	140.50	113.00
Landmark no.		11	12	13	14	15	16	17	18	19	
Image 1	x	72.50	170.00	187.50	56.00	225.50	249.50	66.50	222.50	244.50	
	y	227.50	238.50	179.50	273.00	295.50	187.00	389.00	412.00	303.50	
Image 2	x	81.50	138.50	193.00	43.00	142.00	233.50	50.00	144.00	231.50	
	y	187.50	214.00	184.50	208.50	256.00	202.50	319.00	373.00	312.50	
Landmark no.		1	2	3	4	5	6	7	8	9	10
Image 3	x	177.50	133.50	165.50	209.50	133.00	164.50	207.50	112.50	157.00	222.50
	y	28.00	39.00	56.50	46.00	89.00	109.50	95.50	96.00	125.50	107.50
Image 4	x	223.00	176.50	201.00	248.50	175.50	199.50	246.50	152.00	187.00	258.50
	y	35.50	44.50	63.50	53.50	94.50	114.50	104.50	100.00	130.00	115.00
Landmark no.		11	12	13	14	15	16	17	18	19	
Image 3	x	110.50	154.00	219.00	66.00	140.00	251.00	69.50	139.50	246.00	
	y	168.00	199.50	179.00	183.00	240.00	200.50	293.00	356.50	312.00	
Image 4	x	150.50	184.50	254.50	104.50	159.50	281.00	106.00	158.50	275.50	
	y	169.50	202.50	186.00	181.50	240.00	209.50	289.00	354.50	321.00	
Landmark no.		1	2	3	4	5	6	7	8	9	10
Image 5	x	189.00	140.50	162.50	212.50	138.00	159.00	209.00	114.00	144.50	219.00
	y	33.50	41.00	65.00	57.00	89.00	113.00	106.00	94.50	132.00	119.50
Image 6	x	209.50	163.00	178.00	227.00	161.00	175.00	224.50	138.00	157.00	231.50
	y	63.50	67.50	83.00	80.50	115.50	132.50	129.00	119.00	145.00	138.50
Landmark no.		11	12	13	14	15	16	17	18	19	
Image 5	x	110.00	140.00	212.00	62.50	110.50	235.00	61.00	106.00	225.00	
	y	162.00	201.50	188.00	173.00	241.50	216.00	276.00	350.50	321.00	
Image 6	x	135.00	154.50	228.00	88.00	117.50	245.00	88.00	116.50	238.00	
	y	187.50	216.00	209.00	192.00	243.00	230.50	300.50	360.50	344.50	
Landmark no.		1	2	3	4	5	6	7	8	9	10
Image 7	x	110.00	75.00	112.50	146.50	80.00	117.00	151.00	63.00	118.00	169.00
	y	24.00	40.50	53.50	35.50	88.50	101.50	83.00	99.50	119.00	90.00
Image 8	x	161.50	121.50	147.00	187.50	122.00	147.00	187.00	102.50	139.50	199.50
	y	47.00	56.00	71.50	62.00	101.00	117.00	106.00	107.50	130.00	115.50
Landmark no.		11	12	13	14	15	16	17	18	19	
Image 7	x	69.00	122.50	173.50	34.00	125.00	211.50	47.00	136.50	217.50	
	y	167.00	187.50	158.50	188.50	227.00	173.00	292.00	336.00	277.00	
Image 8	x	103.00	139.00	199.00	63.00	124.00	225.50	68.00	126.50	224.50	
	y	171.50	195.50	180.00	184.00	229.00	198.00	283.00	335.00	299.50	
Landmark no.		1	2	3	4	5	6	7	8	9	10
Image 9	x	176.50	141.50	176.50	210.50	140.00	175.50	210.50	122.50	174.00	226.50
	y	38.00	44.00	52.50	47.00	94.00	103.00	95.50	99.00	112.50	101.50
Image 10	x	146.50	109.50	136.50	176.00	109.00	136.50	175.00	89.00	131.00	187.50
	y	35.00	41.00	51.50	46.00	88.00	100.50	93.00	93.50	110.50	99.00
Landmark no.		11	12	13	14	15	16	17	18	19	
Image 9	x	120.00	171.50	224.50	82.50	169.50	259.50	83.00	167.50	256.00	
	y	171.50	188.00	174.50	181.50	211.50	185.50	298.00	337.00	303.00	
Image 10	x	88.50	128.50	186.50	48.50	118.50	215.50	50.00	117.50	212.00	
	y	160.50	180.50	166.50	171.00	206.50	180.50	277.50	321.50	289.50	
Landmark no.		1	2	3	4	5	6	7	8	9	10
Image 11	x	146.50	102.50	159.50	201.00	99.50	153.50	194.00	77.50	159.00	221.00
	y	40.00	71.50	104.00	70.00	128.00	159.00	126.00	145.00	194.50	141.50
Image 12	x	167.00	111.50	143.50	200.50	105.50	136.00	191.00	78.50	123.50	208.00
	y	29.00	42.50	80.50	65.50	96.50	136.50	118.00	104.00	163.00	138.00
Landmark no.		11	12	13	14	15	16	17	18	19	
Image 11	x	75.00	150.00	210.00	30.00	161.50	260.50	32.50	150.00	240.50	
	y	218.00	266.00	213.00	256.00	343.50	244.50	356.00	441.00	342.00	

(continued on next page)

Table 4 (continued)

Image 12	x	71.00	113.00	193.00	16.00	85.50	225.50	14.00	76.00	204.00	
	y	175.50	232.50	206.50	193.00	298.00	248.50	292.00	396.50	345.00	
Landmark no.		1	2	3	4	5	6	7	8	9	10
Image 13	x	91.00	78.00	142.50	152.00	79.00	142.50	153.50	72.50	168.50	181.50
	y	35.50	63.00	70.00	41.50	122.00	127.50	99.50	138.50	148.00	103.00
Image 14	x	163.50	138.00	190.00	212.00	128.00	178.50	201.50	114.50	190.50	225.50
	y	47.00	62.50	82.00	66.00	115.50	135.00	118.00	126.00	156.00	128.50
Landmark no.		11	12	13	14	15	16	17	18	19	
Image 13	x	74.50	166.50	180.50	60.00	220.50	239.00	66.00	216.00	234.50	
	y	221.00	229.00	182.00	258.00	276.00	189.50	382.00	400.00	308.50	
Image 14	x	99.50	173.50	208.00	71.50	200.50	256.50	52.50	172.50	229.50	
	y	200.50	232.00	200.00	223.50	278.50	221.00	340.00	395.00	330.00	
Landmark no.		1	2	3	4	5	6	7	8	9	10
Image 15	x	135.00	105.50	167.50	193.00	105.50	166.00	190.50	91.00	180.50	219.00
	y	38.00	64.00	79.50	54.50	124.50	140.50	112.50	140.00	165.00	120.50
Image 16	x	104.00	72.50	112.50	142.00	74.00	114.50	143.50	58.50	119.00	163.50
	y	28.00	36.50	43.50	35.00	87.00	97.00	84.50	94.00	105.50	89.00
Landmark no.		11	12	13	14	15	16	17	18	19	
Image 15	x	89.00	176.00	212.50	57.50	209.00	270.50	60.00	201.00	260.00	
	y	223.50	249.00	203.00	258.00	304.50	219.00	382.50	432.50	338.50	
Image 16	x	60.50	120.00	164.00	27.00	131.50	203.00	33.50	133.00	205.00	
	y	167.50	181.00	160.50	181.50	207.00	170.50	298.50	329.50	284.50	

It is obvious that if \bar{V} is the sample mean of $V_r, r = 1, \dots, n$ then

$$j_k(\bar{Y}_{j_k, n}) = P(\bar{V}) = P(\tilde{D}). \tag{A.4}$$

Therefore w.l.o.g. we may assume that

$$g_s(a) = e_a, \quad \forall s = 1, \dots, q, \forall a = 1, \dots, m + 1, \tag{A.5}$$

and that $j_k(p) = P(\bar{V})$ is given with $p = ([e_{m+1}], \dots, [e_{m+1}])$. The tangent space $T_p(\mathbb{R}P^m)^q$ can be identified with $(\mathbb{R}^m)^q$, and with this identification $f_{(s,a)}$ in (4.10) is given by $f_{(s,a)} = (0, \dots, 0, e_a, 0, \dots, 0)$ has all vector components zero except for position s , which is the vector e_a of the standard basis of \mathbb{R}^m . We may then assume that $e_{(s,a)}(\tilde{D}) := D_{p,j_k}(\tilde{D})e_s^b$. From a straightforward computation which can be found in [15] it follows that $D_{\tilde{D}}P(e_s^{m+1}) = 0$, except for

$$D_{\tilde{D}}P(e_s^{m+1}) = \{d_s(m + 1) - d_s(a)\}e_{(s,a)}(P_k(\tilde{D})). \tag{A.6}$$

If $Y_r, r = 1, \dots, n$ is given by (4.7), from (A.6) and (4.2) we obtain

$$(G_n)_{(i,a),(j,b)} = n^{-1}\{d_i(m + 1) - d_i(a)\}^{-1}\{d_j(m + 1) - d_j(b)\}^{-1} \sum_r iX_r^a jX_r^b X_r^{m+1} X_r^{m+1}, \tag{A.7}$$

which is (4.13) expressed in the selected basis, thus proving the lemma. ■

The proof of Theorem 4.1 is elementary following from Lemma 4.1, and from the observation that V_1 has a multivariate distribution with a finite covariance matrix Σ since $(\mathbb{R}P^m)^q$ is compact. For n large enough, \bar{V} has approximately a multivariate normal distribution $\mathcal{N}(\mu, \frac{1}{n}\Sigma)$, and by the delta method [28, p.45], it follows that

$$P(\bar{V}) \sim \mathcal{N}\left(P(\mu) = j_k(\mu_k), \frac{1}{n}D_{\mu}P\Sigma D_{\mu}P^T\right). \tag{A.8}$$

The range of the differential $D_{\mu}P$ is a subspace of $T_{P(\mu)}j_k((\mathbb{R}P^m)^q)$; therefore the asymptotic distribution of $P(\bar{V})$ is degenerate. If we decompose $S(m + 1)^q = T_{P(\mu)}j_k((\mathbb{R}P^m)^q) \oplus T_{P(\mu)}j_k((\mathbb{R}P^m)^q)^\perp$ into tangent and normal subspaces, then the covariance matrix of the tangential marginal distribution of $\tan P(\bar{V})$ is $\frac{1}{n}\Sigma_E$, which is nondegenerate since the generalized extrinsic covariance $\det(\Sigma_E) = \prod_{s=1}^q \lambda_s(a) > 0$. Because \bar{V} is a strongly consistent estimator of μ and S_n is a strongly consistent estimator of Σ , from Slutsky's theorems [28, p.42] it follows that G_n in (4.13) is a strongly consistent estimator of Σ_E . Let $U = [(sU_1, \dots, sU_m)_{s=1, \dots, q}]^T$ be the random vector whose components are the components of $\tan P(\bar{V})$ w.r.t. the basis $e_{(s,a)}(\tilde{D})$ given in the proof of Lemma 4.1. G_n is a consistent estimator of Σ_E . Then asymptotically $Z_n = \sqrt{n}G_n^{-\frac{1}{2}}U$ converges to $\mathcal{N}(0, I_{mq})$, and $Z_n^T Z_n$ converges to a chi-square distribution with mq degrees of freedom. If one uses the equivariance again, one gets $Z_n^T Z_n = T(\bar{Y}_{j_k, n}; \mu)$ in (4.15), which completes the proof of Theorem 4.1. ■

Table 5
Reconstructed object coordinates of the eight configurations.

Landmark no.		1	2	3	4	5	6	7	8	9	10
Reconstructed 3D Config 1	x	-0.09	-0.18	-0.08	0.01	-0.19	-0.09	0.00	-0.24	-0.08	0.05
	y	0.37	0.31	0.25	0.31	0.18	0.12	0.18	0.15	0.06	0.15
	z	2.51	2.44	2.27	2.36	2.53	2.35	2.44	2.49	2.23	2.35
	w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Landmark no.		11	12	13	14	15	16	17	18	19	
Reconstructed 3D Config 1	x	-0.25	-0.09	0.04	-0.35	-0.09	0.13	-0.36	-0.10	0.14	
	y	-0.04	-0.12	-0.03	-0.11	-0.22	-0.08	-0.45	-0.53	-0.38	
	z	2.62	2.36	2.49	2.55	2.13	2.33	2.78	2.35	2.56	
	w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Landmark no.		1	2	3	4	5	6	7	8	9	10
Reconstructed 3D Config 2	x	0.43	0.11	0.27	0.59	0.11	0.26	0.59	-0.05	0.17	0.66
	y	1.39	1.30	1.13	1.22	0.99	0.81	0.91	0.93	0.69	0.81
	z	6.89	6.78	6.52	6.65	6.90	6.64	6.79	6.76	6.47	6.68
	w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Landmark no.		11	12	13	14	15	16	17	18	19	
Reconstructed 3D Config 2	x	-0.07	0.16	0.65	-0.38	-0.00	0.81	-0.39	-0.01	0.80	
	y	0.47	0.23	0.35	0.39	-0.02	0.19	-0.36	-0.77	-0.58	
	z	6.99	6.69	6.86	6.89	6.32	6.66	7.15	6.61	6.97	
	w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Landmark no.		1	2	3	4	5	6	7	8	9	10
Reconstructed 3D Config 3	x	0.22	0.06	0.11	0.27	0.05	0.09	0.26	-0.03	0.03	0.27
	y	0.62	0.61	0.54	0.55	0.45	0.38	0.39	0.43	0.33	0.35
	z	3.28	3.30	3.16	3.16	3.30	3.17	3.15	3.30	3.10	3.08
	w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Landmark no.		11	12	13	14	15	16	17	18	19	
Reconstructed 3D Config 3	x	-0.04	0.02	0.25	-0.21	-0.10	0.29	-0.22	-0.11	0.26	
	y	0.20	0.10	0.13	0.18	0.01	0.06	-0.18	-0.35	-0.28	
	z	3.30	3.09	3.07	3.29	2.96	2.93	3.31	2.96	2.90	
	w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Landmark no.		1	2	3	4	5	6	7	8	9	10
Reconstructed 3D Config 4	x	0.00	-0.31	-0.11	0.20	-0.32	-0.11	0.20	-0.48	-0.16	0.29
	y	1.47	1.36	1.15	1.27	1.03	0.83	0.94	0.96	0.69	0.84
	z	8.10	7.90	7.38	7.61	8.17	7.61	7.80	8.09	7.29	7.56
	w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Landmark no.		11	12	13	14	15	16	17	18	19	
Reconstructed 3D Config 4	x	-0.50	-0.18	0.30	-0.83	-0.27	0.47	-0.87	-0.28	0.51	
	y	0.46	0.21	0.36	0.34	-0.06	0.19	-0.57	-0.89	-0.63	
	z	8.55	7.70	7.98	8.37	7.10	7.43	9.23	7.69	8.11	
	w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Landmark no.		1	2	3	4	5	6	7	8	9	10
Reconstructed 3D Config 5	x	-0.13	-0.43	-0.19	0.10	-0.46	-0.19	0.11	-0.62	-0.23	0.20
	y	1.72	1.69	1.43	1.50	1.36	1.11	1.17	1.30	0.99	1.07
	z	8.10	8.19	7.35	7.51	8.69	7.74	7.76	8.57	7.49	7.44
	w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Landmark no.		11	12	13	14	15	16	17	18	19	
Reconstructed 3D Config 5	x	-0.68	-0.25	0.23	-1.06	-0.31	0.43	-1.19	-0.34	0.51	
	y	0.77	0.47	0.59	0.67	0.24	0.42	-0.46	-0.81	-0.52	
	z	9.48	8.14	8.22	9.45	7.62	7.54	11.02	9.01	8.88	
	w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Landmark no.		1	2	3	4	5	6	7	8	9	10
Reconstructed 3D Config 6	x	0.00	-0.16	-0.07	0.10	-0.19	-0.09	0.07	-0.28	-0.13	0.12
	y	0.52	0.47	0.33	0.38	0.32	0.18	0.24	0.29	0.08	0.17
	z	2.99	2.98	2.83	2.85	3.12	2.97	2.99	3.12	2.89	2.92
	w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Landmark no.		11	12	13	14	15	16	17	18	19	
Reconstructed 3D Config 6	x	-0.32	-0.17	0.08	-0.50	-0.25	0.17	-0.56	-0.31	0.12	
	y	0.07	-0.13	-0.04	0.00	-0.33	-0.17	-0.36	-0.70	-0.51	
	z	3.31	3.09	3.10	3.31	2.94	2.98	3.64	3.28	3.31	

(continued on next page)

Table 5 (continued)

	w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Landmark no.		1	2	3	4	5	6	7	8	9	10
Reconstructed 3D Config 7	x	−0.01	−0.12	0.10	0.20	−0.17	0.05	0.15	−0.23	0.09	0.25
	y	0.80	0.70	0.58	0.68	0.47	0.36	0.45	0.41	0.25	0.40
	z	4.44	4.31	4.06	4.18	4.41	4.14	4.24	4.35	3.95	4.16
	w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Landmark no.		11	12	13	14	15	16	17	18	19	
Reconstructed 3D Config 7	x	−0.31	0.02	0.17	−0.44	0.11	0.35	−0.56	−0.01	0.25	
	y	0.06	−0.07	0.08	−0.06	−0.26	−0.01	−0.63	−0.78	−0.51	
	z	4.48	4.10	4.25	4.39	3.74	4.01	4.65	3.96	4.21	
	w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Landmark no.		1	2	3	4	5	6	7	8	9	10
Reconstructed 3D Config 8	x	−0.04	−0.06	−0.03	−0.01	−0.06	−0.04	−0.02	−0.08	−0.04	−0.00
	y	0.11	0.11	0.11	0.12	0.08	0.08	0.09	0.07	0.08	0.09
	z	0.56	0.56	0.60	0.59	0.59	0.63	0.62	0.59	0.65	0.64
	w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Landmark no.		11	12	13	14	15	16	17	18	19	
Reconstructed 3D Config 8	x	−0.08	−0.04	−0.01	−0.11	−0.04	0.02	−0.12	−0.05	0.02	
	y	0.03	0.03	0.04	0.02	0.01	0.04	−0.06	−0.09	−0.05	
	z	0.64	0.69	0.68	0.64	0.74	0.72	0.72	0.84	0.81	
	w	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	

Appendix B

Table 4 displays the coordinates of the 19 landmarks in each of 16 camera images. The camera images are organized as pairs, each of which were used to reconstruct a 3D object in \mathbb{R}^3 such as the ones displayed in the Fig. 4. Homogeneous coordinates of the eight reconstructed objects, using the standard reconstruction algorithm [16, p. 188] are displayed in Table 5.

References

- [1] O.D. Faugeras, What can be seen in three dimensions with an uncalibrated stereo rig? in: Proc. European Conference on Computer Vision, in: LNCS, vol. 588, 1992, pp. 563–578.
- [2] R.I. Hartley, R. Gupta, T. Chang, Stereo from uncalibrated cameras, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 1992.
- [3] S. Sugathadasa, Affine and projective shape analysis with applications, Ph.D. Thesis, Texas Tech University, 2006.
- [4] R.I. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, 2nd edition, Cambridge University Press, 2004.
- [5] X. Liu, V. Patrangenaru, S. Sugathadasa, Projective Shape Analysis for Noncalibrated Pinhole Camera Views. To the Memory of W.P. Dayawansa, Technical Report M983, Florida State University Department of Statistics, 2007.
- [6] S.J. Maybank, Classification based on the cross ratio, in: J.L. Mundy, A. Zisserman, D. Forsyth (Eds.), Applications of Invariance in Computer Vision, in: Lecture Notes in Comput. Sci., vol. 825, Springer, Berlin, 1994, pp. 433–472.
- [7] K.V. Mardia, C. Goodall, A.N. Walder, Distributions of projective invariants and model based machine vision, Adv. Appl. Probab. 28 (1996) 641–661.
- [8] C. Goodall, K.V. Mardia, Projective shape analysis, J. Graph. Comput. Statist. 8 (1999) 143–168.
- [9] V. Patrangenaru, New large sample and bootstrap methods on shape spaces in high level analysis of natural images, Comm. Statist. 30 (2001) 1675–1693.
- [10] J.L. Lee, R. Paige, V. Patrangenaru, F. Ruymgaart, Nonparametric density estimation on homogeneous spaces in high level image analysis, in: R.G. Aykroyd, S. Barber, K.V. Mardia (Eds.), Bioinformatics, Images, and Wavelets, Department of Statistics, University of Leeds, 2004, pp. 37–40. <http://www.maths.leeds.ac.uk/Statistics/workshop/leeds2004/temp>.
- [11] R. Paige, V. Patrangenaru, F. Ruymgaart, W. Wang, Analysis of projective shapes of curves using projective frames, in: S. Barber, P.D. Baxter, K.V. Mardia, R.E. Walls (Eds.), Quantitative Biology, Shape Analysis, and Wavelets, Leeds University Press, Leeds, 2005, pp. 71–74. <http://www.maths.leeds.ac.uk/Statistics/workshop/leeds2005/temp>.
- [12] K.V. Mardia, V. Patrangenaru, Directions and projective shapes, Ann. Statist. 33 (4) (2005) 1666–1699.
- [13] J.T. Kent, K.V. Mardia, A new representation for projective shape, in: S. Barber, P.D. Baxter, K.V. Mardia, R.E. Walls (Eds.), Interdisciplinary Statistics and Bioinformatics, Leeds University Press, Leeds, 2006, pp. 75–78. <http://www.maths.leeds.ac.uk/lasr2006/proceedings/>.
- [14] A. Munk, R. Paige, J. Pang, V. Patrangenaru, F. Ruymgaart, The one and multisample problem for functional data with applications to projective shape analysis, J. Multivar. Anal. 99 (2008) 815–833.
- [15] R.N. Bhattacharya, V. Patrangenaru, Large sample theory of intrinsic and extrinsic sample means on manifolds – Part II, Ann. Statist. 33 (3) (2005) 1211–1245.
- [16] Y. Ma, S. Soatto, J. Kosecka, S.S. Sastry, An Invitation to 3-D Vision, Springer, New York, 2006.
- [17] R.I. Hartley, Projective reconstruction and invariants from multiple images, Preprint, 1993.
- [18] V. Patrangenaru, Moving projective frames and spatial scene identification, in: K.V. Mardia, R.G. Aykroyd, I.L. Dryden (Eds.), Proceedings in Spatial-Temporal Modeling and Applications, Leeds University Press, 1999, pp. 53–57.
- [19] R.N. Bhattacharya, V. Patrangenaru, Large sample theory of intrinsic and extrinsic sample means on manifolds—I, Ann. Statist. 31 (1) (2003) 1–29.
- [20] M. Spivak, A Comprehensive Introduction to Differential Geometry, vol. I, Second edition, Publish or Perish, Inc., Wilmington, Del., 1979.
- [21] I. Dimitric, A note on equivariant embeddings of Grassmannians, Publ. Inst. Math. (Beograd) (N.S.) 59 (1996) 131–137.
- [22] G.S. Watson, Statistics on Spheres, in: University of Arkansas Lecture Notes in the Mathematical Sciences, vol. 6, A Wiley-Interscience Publication, John Wiley and Sons Inc., New York, 1983.
- [23] M.J. Prentice, A distribution-free method of interval estimation for unsigned directional data, Biometrika 71 (1984) 147–154.
- [24] R.N. Bhattacharya, J.K. Ghosh, On the validity of the formal Edgeworth expansion, Ann. Statist. 6 (1978) 434–451.
- [25] C. Longuet-Higgins, computer algorithm for reconstructing a scene from two projections, Nature 293 (1981) 133–135.

- [26] A. Bhattacharya, R. Bhattacharya, Nonparametric statistics on manifolds with applications to shape spaces, in: S. Ghoshal, B. Clarke (Eds.), *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, in: IMS Collections, vol. 3, 2008, pp. 282–301.
- [27] P. Hall, *The Bootstrap and Edgeworth Expansion*, Springer Series in Statistics, New York, 1997.
- [28] T.S. Ferguson, *A Course in Large Sample Statistics*, Chapman and Hall, 1996.
- [29] V. Balan, M. Crane, V. Patrangenaru, X. Liu, Projective shape manifolds and coplanarity of landmark configurations. A nonparametric approach (2009) (in press).
- [30] B. Efron, Bootstrap methods: Another look at the jackknife, *Ann. Statist.* 7 (1) (1979) 1–26.
- [31] J.T. Kent, Projective shape analysis, in: Summer 2007 Program on the Geometry and Statistics of Shape Spaces, July 7–13, 2007, Speaker Abstracts.
- [32] K.V. Mardia, V. Patrangenaru, Directions and projective shapes, Technical Report No. 02/04, Dept. Statistics, Univ. Leeds, 2002.
- [33] S.J. Maybank, Probabilistic analysis of the application of the cross ratio to model based vision: Misclassification, *Int. J. Comput. Vis.* 14 (3) (1995).
- [34] S. Birchfield, <http://www.ces.clemson.edu/~stb/projective/node19.html>, 2003.