

Membership Statistical Test For Shapes With Application to Face Identification

Eugene Demidenko

Abstract—We develop a statistical test to determine if shape belongs to a given homogeneous family of shapes based on general statistical hypothesis testing. First, we illustrate the membership test on a simple random triangle model. Second we apply this test to face identification problem with the task to identify that the last photo is of the same person with different facial expression. Third, we outline how to use the membership test in more general size-irrelevant statistical shape model.

Keywords: Linear model, least squares, triangle shape, statistical test.

I. INTRODUCTION

One of the most common task of shape analysis is shape identification. Specifically, we formulate the problem as follows. Let N random shapes S_1, S_2, \dots, S_N belong to the same general population, assuming that shapes are independent and identically distributed (iid). Let a $(N + 1)$ th shape be given. We want to determine whether this shape belongs to the same general population with a given probability of the type I error, α . Herein this problem is called *the membership problem* and the according test is called *the membership test*. Conceptually, one decides that the $(N + 1)$ shape belongs to the same population if it lies within the range of the previous N shapes. To solve the membership problem, we adopt the well established statistical theory of linear models to determine the range of shape variation. In the case when shape can be reduced to one statistic the range reduces to the confidence interval. In practical setting when shape is specified by a numbers of landmarks the shape range takes the form of the normalized sum of squares. To test whether the shape is within the range the F-test is used. We illustrate the membership test by random triangles that are uniquely specified by the length of three sides. The random triangle model is applied to face identification with different facial expression. Much work has been done in the field of face identification and recognition as a part of more general video-based biometric person authentication—for example, see a recent special issue of Lecture Notes in Computer Science (2003). Unlike others, often heuristics approaches we rigorously apply classic statistical hypothesis testing that enables to control the type I and II errors. At the end we generalize the triangle model to general polygon shape model with arbitrary size.

II. MEMBERSHIP TEST FOR GENERAL LINEAR MODEL

First we remind the general linear hypothesis test in the framework of linear model and then modify the test for the

membership problem solution.

A. General linear hypothesis testing

The general linear model with Gaussian error term in a vector-matrix form is defined as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{V}), \quad (1)$$

where \mathbf{X} is the $n \times m$ design matrix of full rank and \mathbf{V} is a known, up to a scalar, $n \times n$ nonsingular covariance matrix. The unknown $m \times 1$ vector $\boldsymbol{\beta}$ is the vector of regression coefficients and $\boldsymbol{\eta}$ is the $n \times 1$ unobservable error vector with multivariate Gaussian (normal) distribution with zero mean and covariance matrix \mathbf{V} , Searle (1971), Rao (1972), Key (1998), Demidenko (2004). It is proven the best unbiased estimator of $\boldsymbol{\beta}$ is the solution to the Weighted Least Squares (WLS) and is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}), \quad (2)$$

where $'$ denotes matrix transposition. In a special case when $\mathbf{V} = \mathbf{I}$, the identity matrix, the weighted least squares reduces to ordinary least squares which minimizes the norm $\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$. The general linear statistical hypothesis is formulated as

$$H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}, \quad (3)$$

against the alternative $H_A : \mathbf{C}\boldsymbol{\beta} \neq \mathbf{0}$, where \mathbf{C} is a fixed $q \times m$ matrix of full rank $q \leq m$. To test (3) two Residual Sums of Squares (RSS) are computed,

$$\begin{aligned} RSS &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \\ RSS_0 &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_0)' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_0). \end{aligned}$$

RSS is the absolute minimum of the residual sum of squares $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ and $\hat{\boldsymbol{\beta}}$ is the WLS estimator (2). RSS_0 is the residual sum squares under the linear restriction $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ and $\hat{\boldsymbol{\beta}}_0$ is the WLS estimator under restriction (3). In many instances one can compute RSS_0 by reducing the regression under restriction (3) to a regression with $m - q$ coefficients. Then, under hypothesis H_0 ,

$$\frac{(RSS_0 - RSS)/q}{RSS/(n - m)} \sim F(q, n - m), \quad (4)$$

where the right-hand side is the F -distribution with q and $n - m$ degrees of freedom, Searle (1971), Rao and Toutenburg (1999). F-test (4) is very flexible and allows testing various statistical hypotheses on $\boldsymbol{\beta}$. Lehmann (1986) proved that the F-test is optimal, or more specifically is uniformly the most powerful invariant test.

In many practical situations the covariance matrix \mathbf{V} is *unknown* and depends on a finite-dimensional unknown parameter vector θ , so we write $\mathbf{V} = \mathbf{V}(\theta)$. For example, assuming equi-correlation structure we have $V_{ij} = \rho$ if $i \neq j$ and $V_{ii} = 1$. A common approach to simultaneously estimate β and θ is the maximum likelihood approach with the criterion

$$l(\beta, \theta) = -\frac{1}{2} [\ln |\mathbf{V}(\theta)| + (\mathbf{y} - \mathbf{X}\beta)' \mathbf{V}^{-1}(\theta) (\mathbf{y} - \mathbf{X}\beta)].$$

Maximization of l over β and θ yields the maximum likelihood estimates (MLE) $\hat{\beta}$ and $\hat{\theta}$. It is easy to obtain that when θ is known the maximization over β leads to (2). We still can apply F-test (3) after substituting θ with an estimate $\hat{\theta}$ but then it loses its optimal properties. However, asymptotically, when the number of observations grows, this test regains its optimal properties, Cox and Hinkley (1974). We shall use the maximum likelihood approach for the random triangle model in the next section.

B. Membership test for univariate data

To start, we apply the F-test to the following simple membership problem: let y_1, y_2, \dots, y_n be an iid sample from a general population and y_{n+1} is a new observation, independent of the previous n observations. Does y_{n+1} belong to the same population, i.e. is y_{n+1} a member? This is a typical question of the identification problem. For example, in medical diagnostics $\{y_i\}$ are observations of normal patients and y_{n+1} is the observation of a new patient. If y_{n+1} is within “normal” range the patient is disease-free otherwise further diagnostics or treatment is required.

To simplify, we assume that observations have a normal distribution and the same variance, or more precisely $y_i \sim \mathcal{N}(\mu, \sigma^2)$ for $i = 1, 2, \dots, n$ and $y_{n+1} \sim \mathcal{N}(\mu_*, \sigma^2)$. Then the question of whether y_{n+1} is a member reduces to the hypothesis $H_0 : \mu = \mu_*$ with the alternative $H_A : \mu \neq \mu_*$. Applying the general test (4) it is easy to check that in this case $q = 1$, $RSS = \sum_{i=1}^n (y_i - \bar{y})^2$ and $RSS_0 = \sum_{i=1}^{n+1} (y_i - \bar{y}_{n+1})^2$, where $\bar{y} = \sum_{i=1}^n y_i / n$ and $\bar{y}_{n+1} = \sum_{i=1}^{n+1} y_i / (n+1)$. After elementary algebra one finds

$$RSS_0 - RSS = \frac{n}{n+1} (y_{n+1} - \bar{y})^2,$$

and therefore, if y_{n+1} is a member, then

$$\frac{n}{n+1} \frac{(y_{n+1} - \bar{y})^2}{s_y^2} \sim F(1, n-1), \quad (5)$$

where $s_y^2 = RSS/(n-1)$. Taking the square root, the F-test (5) becomes equivalent to the t-test. Thus, one accepts the membership if

$$|y_{n+1} - \bar{y}| < s_y \sqrt{\frac{n+1}{n}} t_{1-\alpha/2, n-1}, \quad (6)$$

where $t_{1-\alpha/2, n-1}$ is the $1 - \alpha/2$ quantile of the t-distribution with $n - 1$ degrees of freedom (df). As follows from (6) for large n , one accepts the membership if y_{n+1} belongs to the $1 - \alpha$ confidence interval constructed based on the sample $\{y_i, i = 1, \dots, n\}$.

C. Membership test for multivariate data

Now we generalize the membership test to a sequence of linear models defined by (1). It is assumed that shape is defined by m landmarks or generally points on the plane or space; to be specific we assume that the shape is two-dimensional and can be represented as a polygon. It is assumed that the order of the polygon sides is known and a vertex is chosen to be the initial point. Following Bookstein (1991) and Small (1996) a polygon with n vertices can be equivalently represented by $2m - 3$ numbers: m sides plus $m - 3$ diagonals from the initial vertex and diagonals. Importantly, this representation is rotation invariant but size dependent (we will address the size variation later). Then, if N shapes are represented by $2m - 3$ vectors $\{\mathbf{y}_i, i = 1, \dots, N\}$ and a \mathbf{y}_{N+1} vector (shape) is given we want to know if it belongs to the same population. It is worthwhile to mention that this shape reduction to distances is not the only way to reduce a shape to a linear model. For example, to eliminate size one could take angles.

We assume that statistical model for the i th shape is defined via a set of general linear models, similarly to (1),

$$\mathbf{y}_i = \mathbf{X}_i \beta + \boldsymbol{\eta}_i, \quad \boldsymbol{\eta}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{V}), \quad i = 1, 2, \dots, N. \quad (7)$$

Typically, \mathbf{X}_i is the identity matrix and then β specifies the true shape. However, if one has two types of shapes, such as gender related, the data can be combined under umbrella (7) and matrix \mathbf{X}_i may have dummy variables that represent gender. Moreover, coefficients at gender variable indicate the difference in gender shape and may serve for the appropriate significance testing. The $(N+1)$ th shape also follows a linear mode but possibly with different beta-coefficients,

$$\mathbf{y}_{N+1} = \mathbf{X}_{N+1} \beta_* + \boldsymbol{\eta}_{N+1},$$

where $\boldsymbol{\eta}_{N+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$. To simplify, we shall assume that covariance matrix \mathbf{V} is known. Then in the hypothesis testing language, the membership problem can be translated into the hypothesis

$$H_0 : \beta = \beta_*. \quad (8)$$

We start with RSS_0 . If $\beta = \beta_*$, then all $N + 1$ patients belong to the same population and for example applying maximum likelihood gives the estimates $\hat{\beta}_{N+1}$ and $\hat{\mathbf{D}}$. Then the minimal weighted sum of squares under (8) is $RSS_0 = \sum_{i=1}^{N+1} (\mathbf{y}_i - \mathbf{X}_i \hat{\beta}_{N+1})' \mathbf{V}^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\beta}_{N+1})$. Now we find the minimal weighted sum of squares when (8) is not true, i.e. β and β_* are unrelated. Thus we want to find the minimum of the weighted sum of squares

$$\sum_{i=1}^N (\mathbf{y}_i - \mathbf{X}_i \beta)' \mathbf{V}^{-1} (\mathbf{y}_i - \mathbf{X}_i \beta) + (\mathbf{y}_{N+1} - \mathbf{X}_{N+1} \beta_*)' \mathbf{V}^{-1} (\mathbf{y}_{N+1} - \mathbf{X}_{N+1} \beta_*). \quad (9)$$

Two situations may occur. In the first situation there exists $\hat{\beta}_*$ such that $\mathbf{y}_{N+1} = \mathbf{X}_{N+1} \hat{\beta}_*$. In particular this happens when the length of \mathbf{y}_{N+1} is less than m and matrix \mathbf{X}_{N+1} has full rank. Then the second term of (9) vanishes and the minimal sum of squares reduces to $RSS = S_1 = \sum_{i=1}^N (\mathbf{y}_i -$

$\mathbf{X}_i \hat{\beta})' \mathbf{V}^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\beta})$, where $\hat{\beta}$ is the weighted least squares estimator,

$$\hat{\beta} = \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{V}^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{V}^{-1} \mathbf{y}_i \right)$$

In the second situation, the minimum of the second term of (9) is $S_2 = (\mathbf{y}_{N+1} - \mathbf{X}_{N+1} \hat{\beta}_*)' \mathbf{V}^{-1} (\mathbf{y}_{N+1} - \mathbf{X}_{N+1} \hat{\beta}_*)$ and it is expected to be positive, where $\hat{\beta}_* = (\mathbf{X}_{N+1}' \mathbf{V}^{-1} \mathbf{X}_{N+1})^{-1} \mathbf{X}_{N+1}' \mathbf{V}^{-1} \mathbf{y}_{N+1}$ so that $RSS = S_1 + S_2$. Hence, in both situations, according to the F-test, we say the new patient is normal if

$$\frac{(RSS_0 - RSS)/m}{RSS/(\sum n_i - m)} \sim F(m, \sum n_i - m). \quad (10)$$

Further, we will refer (10) to as the membership test. This test is exact when matrix \mathbf{V} is known. Typically \mathbf{V} is unknown and is estimated along with β by maximum likelihood (see the next section). Then test (10) has an approximate significance level. However, since estimates of β and \mathbf{V} do not correlate in large sample, test (10) regains its optimal properties asymptotically although it becomes equivalent to the χ^2 -test, Rao (1973).

III. STATISTICAL ANALYSIS OF RANDOM TRIANGLES

We start with the simplest two-dimensional (planar) shape, triangle. We deal with *random triangles*, so that the length of the side is a random variable. It is assumed that triangles are independent and identically distributed (iid). Since the triangular shape is uniquely specified by the length of its sides we reduce the statistical analysis to the analysis of n iid triples. The triples are independent and belong to the same general population. Statistical analysis of triangles is a relatively old shape problem and was considered by a number of authors, Watson (1986), Mardia (1989), Goodall and Lange (1989), to name a few. However, unlike previous studies we assume that triangles come just slightly rotated and scaled so that we know the order of the sides. In fact, in many applied problems we know the top and the base of triangles, so that the order of the sides is known. For instance, this assumption holds for our face identification problem below.

In particular, we solve the following membership problem: given a sequence of N iid triangles determine whether the $(N+1)$ th triangle belongs to the same general population.

Thus, let us have N iid random triangles with the length of the sides specified by the following *triangle model*

$$\mathbf{s}_i = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, N, \quad (11)$$

where $\mathbf{s}_i = (s_{i1}, s_{i2}, s_{i3})'$ denotes the vector of the length of the three sides and vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)'$ is unknown. The error vector $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3})' \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$, where \mathbf{V} is the 3×3 covariance matrix. Vector $\boldsymbol{\mu}$ uniquely specifies the mean triangular shape and matrix \mathbf{V} specifies the covariance matrix of the error vector, both unknown and to be estimated (unstructured parameterization). The correlation between sides may reflect the fact that the sides were computed from vertices coordinates that are random. If the randomness of the length of the sides is due to the measurement error one may assume that all ε_{ij} have the same variance that lead to a more compact

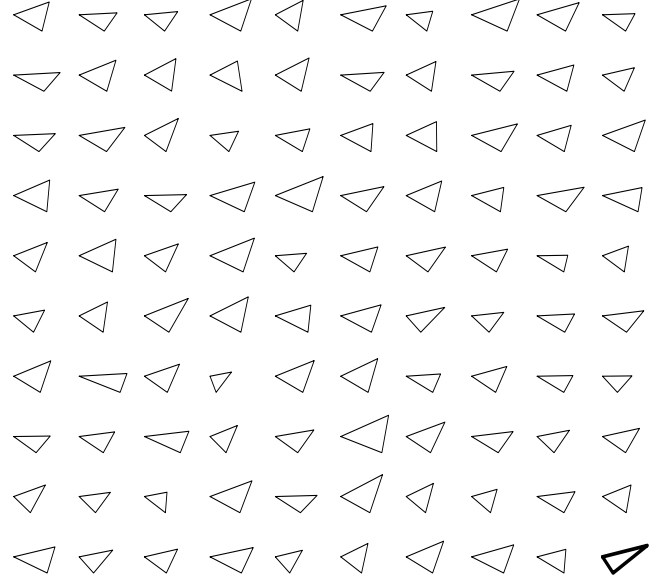


Fig. 1. First 99 independent triangles belong to the same general population. Does the 100th triangle (bold) belong to the same population?

parameterization, $\mathbf{V} = \sigma^2 \mathbf{R}$, where the correlation matrix, \mathbf{R} is given by

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}. \quad (12)$$

An alternative (isotropic) parameterization would be to assume that all three sides correlate at the same magnitude, $\rho_{12} = \rho_{13} = \rho_{23} = \rho$. This model has a nice interpretation as a model to account for intra-cluster (intra-triangle) correlation. Indeed, let us assume that the length is subject to a pure measurement error but triangles are different in size that can be captured by the model $s_{ij} = \mu_j + b_i + \varepsilon_{ij}$, where $\varepsilon_{ij} \sim \mathcal{N}(0, \tau^2)$ is associated with the measurement error and $b_i \sim \mathcal{N}(0, \tau^2 d)$ reflects the difference in size (d is the scaled variance). It is elementary to check that in the previous notation $\sigma^2 = \tau^2(1 + d)$ and $\rho = d/(1 + d)$. The theory of linear mixed effects model and balanced growth curve model readily applies, Demidenko and Stukel (2002).

A problem may emerge with model (11) when the measurements are very imprecise because the sum of two sides may be less than the other side. However, we shall ignore this possibility.

One hundred random triangles are shown in Figure 1. The question we pose is: “Assuming that the first 99 triangles are iid and follow model (11), does the last triangle (bold) belong to the same general population?”

We start with estimation of $\boldsymbol{\mu}$ and \mathbf{V} given data $\{\mathbf{s}_i, i = 1, \dots, N\}$. The log-likelihood function for model (11) is

$$l(\boldsymbol{\mu}, \mathbf{V}) = -\frac{1}{2} \left\{ N \ln |\mathbf{V}| + \sum_{i=1}^N (\mathbf{s}_i - \boldsymbol{\mu})' \mathbf{V}^{-1} (\mathbf{s}_i - \boldsymbol{\mu}) \right\}.$$

The maximum of this function admits a closed-form solution under unstructured parameterization because model (11) is

balanced (no missing measurements),

$$\bar{\mathbf{s}} = \frac{1}{N} \sum_{i=1}^N \mathbf{s}_i, \quad \hat{\mathbf{V}}_{ML} = \frac{1}{N} \sum_{i=1}^N (\mathbf{s}_i - \bar{\mathbf{s}})(\mathbf{s}_i - \bar{\mathbf{s}})'. \quad (12)$$

Also, we can use an unbiased method of moments estimator of \mathbf{V} ,

$$\hat{\mathbf{V}}_{MM} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{s}_i - \bar{\mathbf{s}})(\mathbf{s}_i - \bar{\mathbf{s}})'. \quad (13)$$

When matrix \mathbf{V} is defined as $\sigma^2 \mathbf{R}$, where matrix \mathbf{R} is parametrized by three parameters (12) there is no closed-form solution to the MLE. However we can use the method of moments and estimate $\hat{\rho}_{jk}$ as the correlation coefficient between s_{ij} and s_{ik} , and $\hat{\sigma}^2 = \sum (\mathbf{s}_i - \bar{\mathbf{s}})' \hat{\mathbf{R}}^{-1} (\mathbf{s}_i - \bar{\mathbf{s}}) / (3N)$.

Now we turn our attention to the membership problem: if a $M = N + 1$ triangle (triple) is given, does that triangle belong to the same triangle population? We assume that $\mathbf{s}_{N+1} = \boldsymbol{\mu}_* + \boldsymbol{\varepsilon}_{N+1}$, where $\boldsymbol{\varepsilon}_{N+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$. In other words, the sides of the triangle may be different but the covariance matrix is the same. We rephrase the membership problem in terms of the statistical hypothesis testing $H_0 : \boldsymbol{\mu}_* = \boldsymbol{\mu}$ with the alternative that at least for one j we have $\mu_{*j} \neq \mu_j$ ($j = 1, 2, 3$). Under the null (the last triangle belongs to the population) $\bar{\mathbf{s}} - \mathbf{s}_{N+1} \sim \mathcal{N}(\mathbf{0}, \frac{N+1}{N} \mathbf{V})$ and replacing \mathbf{V} with $\hat{\mathbf{V}}$ we obtain that approximately

$$S = \frac{N}{N+1} (\bar{\mathbf{s}} - \mathbf{s}_{N+1})' \hat{\mathbf{V}}^{-1} (\bar{\mathbf{s}} - \mathbf{s}_{N+1}) \sim \chi^2(3), \quad (14)$$

the chi-squared distribution with three degrees of freedom. Thus, if the test statistic S takes the value greater than $q_{1-\alpha}$, the $(1-\alpha)$ th quantile of the chi-squared distribution, then H_0 is rejected.

In Figure 1 first 99 random triangles are generated with

$$\boldsymbol{\mu} = \begin{bmatrix} 2.5 \\ 2.5 \\ 3.0 \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} 0.20 & 0.05 & 0.07 \\ 0.05 & 0.10 & 0.03 \\ 0.07 & 0.03 & 0.10 \end{bmatrix},$$

and the 100th triangle was generated with $\boldsymbol{\mu}_* = (1.5, 2.0, 3.0)$ so that it does not belong to the same population (all triangles are slightly rotated). The chi-squared test (14) rejects H_0 for the sample depicted in Figure 1 at $\alpha = .05$ with the p-value 0.036. The last triangle is different.

IV. FACE IDENTIFICATION

Face recognition and identification is a classic problem of pattern recognition. Numerous literature exists on the topic and many scientific conferences and meetings are devoted to the subject. For a recent review, the interested reader is referred to a collection of papers of Third International Conference on person authentication that includes face, speech, and fingerprints imaging processing and recognition, Bigun and Smeraldi (2001). Surveys of methods on face recognition can be found in Samal and Iyengar (1992) and Challapa et al. (1995).

The basis for our approach is the observation that regardless of the facial expression the distance between eyes and the tip of the nose does not change, thus our triangle model (11) can be applied. Particularly, we want to solve the following

membership problem: given a sequence of photos of one person, with possible different facial expression, we want to determine whether the next photo is of the same person.

We use the Yale Face Database which contains 165 grey images of 15 individuals at

<http://cvc.yale.edu/projects/yalefaces/yalefaces.html>

There are 11 facial expression per individual. For illustrative purpose we consider only 11 photos of the first person and one photo of the second person, Figure 2. Since the lower part of the face moves at different face expression we chose the following landmark triangle to quantify the image: right eye (left on the photo), left eye, the tip of the nose. This triangle is specified by the length of its sides. Thus, s_{11} is the distance between eyes of the first photo in Figure 2, s_{12} is the distance between the left eye and nose, and s_{13} is the distance between nose and the right eye of the “centerlight” photo. The landmark points are identified manually and therefore are subject to measurement error. Below we show the mean distances for the first person ($i = 1, \dots, 11$), the vector of distances of the second person ($i = 12$), standard errors of the mean, and the correlation matrix.

$$\hat{\boldsymbol{\mu}}_{11} = \begin{bmatrix} 56.57 \\ 41.65 \\ 41.94 \end{bmatrix}, \quad \mathbf{s}_{12} = \begin{bmatrix} 53.08 \\ 42.80 \\ 45.00 \end{bmatrix},$$

$$SE(\hat{\boldsymbol{\mu}}_{11}) = \begin{bmatrix} .43 \\ .43 \\ .71 \end{bmatrix}, \quad cor(\hat{\boldsymbol{\mu}}_{11}) = \begin{bmatrix} 1 & .19 & .29 \\ .19 & 1 & -.01 \\ .29 & -.01 & 1 \end{bmatrix}.$$

We used an unbiased MM estimator (13) for matrix $\boldsymbol{\Omega}$ with $N = 11$. We decide that the last photo is of the same person if vector \mathbf{s}_{12} is within “normal” range of $\hat{\boldsymbol{\mu}}_{11}$. More precisely, we compute statistic (14) and make decision upon its value. Since under the assumption that the last photo is of the same person statistic S has the chi-squared distribution with three degrees of freedom, the threshold value is 7.81 with significance level (the error to reject the face when it is really of the same person) $\alpha = 0.05$. Our data gives $S = 10.1$ and therefore we reject the null with the p-value 0.018.

V. SCALE-IRRELEVANT POLYGON SHAPE

Now we consider a generalization of the triangle model (11) where size does not matter. Thus triangles follow the linear model (11) but may have different size. One may still be interested in whether the 100th triangle in Figure 1 belongs to the population of the previous 99 triangles. When size does not matter instead of the side length one may consider the angle so the model (11) remains the same but \mathbf{s}_i is the two-dimensional vector of angles (we omit from consideration the third angle because the sum of angles is π).

When the shape is defined by $n > 3$ landmarks the statistical model for a scale-irrelevant shape is defined as

$$\mathbf{s}_i = c_i \boldsymbol{\mu} + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, N, \quad (15)$$

where $\mathbf{s}_i = (s_{i1}, \dots, s_{in})'$ is the $n \times 1$ vector of shape observations, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$ defines the true unknown shape, c_i is a positive scale coefficient associated with size, and $\boldsymbol{\varepsilon}_i$

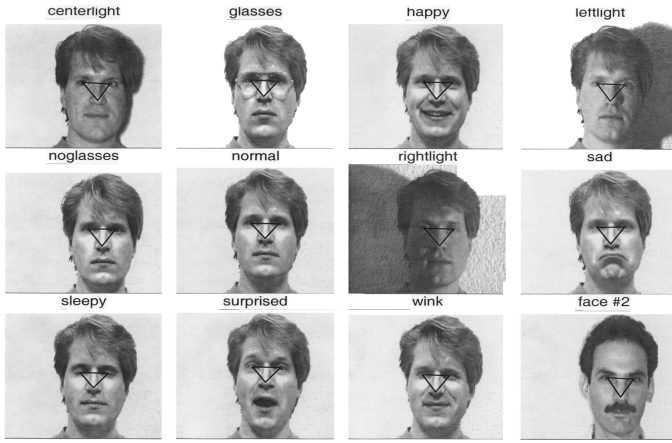


Fig. 2. 12 faces from Yale Face Database. The first 11 photos are of the same person with different face expression. The last photo is of a different person. We apply the triangle membership test of Section ?? to identify that the last photo is of a different person. This test seems adequate because face expression does not change the distance between the eyes and the tip of the nose.

is the $n \times 1$ random error term. For example, a quadrangle (a polygon with four sides) can be defined by the length of four sides plus one diagonal, $n = 5$ (this representation is called *triangulation*, see Small 1996). Again, we assume that the order of sides is known. For clarity of presentation we shall assume that the components of the error term are independent and normally distributed random variables with zero mean and variance σ^2 . If $c_i = 1$ and $n = 3$ we come to the triangle model (11) with $\mathbf{V} = \sigma^2 \mathbf{I}$. Regarding the scale coefficients $\{c_i\}$ we assume that they are fixed and unknown and we treat them as nuisance parameters (below we also consider the model where sizes are random). Since they are unknown we have to restrict the length of the vector $\boldsymbol{\mu}$ because otherwise $\boldsymbol{\mu}$ would be nonidentifiable. Thus, it is assumed $\|\boldsymbol{\mu}\| = 1$. The log-likelihood function for model (15) takes the form

$$l(\boldsymbol{\mu}, \sigma^2, c_1, \dots, c_N) = -\frac{1}{2} \left\{ nN \ln \sigma^2 + \frac{1}{\sigma^2} \sum_{i=1}^N \|\mathbf{s}_i - c_i \boldsymbol{\mu}\|^2 \right\}. \quad (16)$$

Maximization over c_i gives $\hat{c}_i = \mathbf{s}_i' \boldsymbol{\mu}$ and after some algebra we express the sum of squares as

$$\begin{aligned} \sum \|\mathbf{s}_i - \hat{c}_i \boldsymbol{\mu}\|^2 &= \sum \|\mathbf{s}_i - \boldsymbol{\mu} \boldsymbol{\mu}' \mathbf{s}_i\|^2 \\ &= \text{tr} \left[(\mathbf{I} - \boldsymbol{\mu} \boldsymbol{\mu}')^2 \sum \mathbf{s}_i \mathbf{s}_i' \right] = \text{tr} [(\mathbf{I} - \boldsymbol{\mu} \boldsymbol{\mu}')^2 \mathbf{M}] \\ &= \text{tr} [(\mathbf{I} - \boldsymbol{\mu} \boldsymbol{\mu}') \mathbf{M}] = \text{tr} \mathbf{M} - \boldsymbol{\mu}' \mathbf{M} \boldsymbol{\mu}, \end{aligned}$$

where $\mathbf{M} = \sum_{i=1}^N \mathbf{s}_i \mathbf{s}_i'$, a fixed matrix. Hence, the maximum likelihood estimate of the true shape, $\hat{\boldsymbol{\mu}}_{ML}$ is the dominant eigenvector of matrix \mathbf{M} and

$$\hat{\sigma}_{ML}^2 = (\text{tr} \mathbf{M} - \hat{\boldsymbol{\mu}}_{ML}' \mathbf{M} \hat{\boldsymbol{\mu}}_{ML}) / (Nn).$$

This model can be used for further statistical inference, for example, membership test to test whether the $(N+1)$ th shape belongs to the same size-irrelevant family of shapes.

Model (15) has a large number of nuisance parameters which increases with the sample size (shapes). Consequently,

we cannot use standard maximum likelihood theory because $\{\mathbf{s}_i\}$ are not identically distributed. Also, parameter c_i cannot be estimated consistently when N goes to infinity and fixed n . Since our primary interest is $\boldsymbol{\mu}$ the question arises if the mean shape $\hat{\boldsymbol{\mu}}_{ML}$ is consistent. The answer is positive. Indeed, let us assume that $\{|c_i|\}$ are bounded and $\lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N c_i^2 = A > 0$. Then with probability 1 from the law of large numbers $N^{-1} \mathbf{M} \rightarrow \sigma^2 \mathbf{I} + A \boldsymbol{\mu} \boldsymbol{\mu}'$. The dominant eigenvector of the latter matrix is unique and equal $\boldsymbol{\mu}$ because $(\sigma^2 \mathbf{I} + A \boldsymbol{\mu} \boldsymbol{\mu}') \boldsymbol{\mu} = (\sigma^2 + A) \boldsymbol{\mu}$. Since the unique dominant eigenvector is a continuous function of the matrix we obtain that $\hat{\boldsymbol{\mu}}_{ML} \rightarrow \boldsymbol{\mu}$ with probability 1. It is easy to show that $\hat{\sigma}_{ML}^2$ is consistent as well.

The scale-irrelevant shape model may be modified to account for correlation between neighboring points, as in the triangle model. Thus, let us assume that $\text{cov}(\boldsymbol{\varepsilon}_i) = \mathbf{V}$. Now, instead of $\boldsymbol{\mu}' \boldsymbol{\mu} = 1$ using a normalizing restriction $\boldsymbol{\mu}' \mathbf{V}^{-1} \boldsymbol{\mu} = 1$ we obtain that $\hat{\boldsymbol{\mu}}_{ML} = \mathbf{V}^{1/2} \boldsymbol{\tau}$, where $\boldsymbol{\tau}$ is the dominant eigenvector vector of matrix $\mathbf{V}^{-1/2} \mathbf{M} \mathbf{V}^{-1/2}$. If all elements of matrix \mathbf{V} are unknown it requires a large number of shapes because the total number of unknown parameters is $N + n + n(n+1)/2$. For example, a parsimonious specification would be to assume that the correlation between neighboring points is the same.

REFERENCES

- [1] Audio- and Video-Based Biometric Person Authentication, Proceedings (2003). *Lecture Notes in Computer Science*, 2688.
- [2] Bigun, J. and Smeraldi, F. (Eds). (2001). *Audi- and Video-Based Biometric Person Authentication*. Berlin: Springer-Verlag.
- [3] Bookstein, F.L. (1991). *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge: Cambridge University Press.
- [4] Chellapa, R., Wilson, C.L., and Sirohey, S. (1995). Human and machine recognition of faces. *Proceedings of the IEEE* **83**, 705-740.
- [5] Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*, London: Chapman and Hall.
- [6] Demidenko, E. and Stukel, T. (2002). Efficient estimation of general mixed effects models. *Journal of Statistical Planning and Inference* **104**, 197-219.
- [7] Demidenko, E. (2004). *Mixed Models: Theory and Applications*. Hoboken: Wiley.
- [8] Goodall, C.R. and Lange, N. (1989). Growth curve models for correlated triangular shapes. In: *Computer Science and Statistics*, pp. 445-452. Alexandria: American Statistical Association.
- [9] Key, S.M. (1998). *Fundamentals of Statistical Signal Processing. Detection Theory*. Upper Saddle River: Prentice Hall.
- [10] Lehmann, E.L. (1986). *Testing Statistical Hypothesis*. New York: Wiley.
- [11] Mardia, K.V. (1989). Shape analysis of triangles through directional techniques. *Journal of the Royal Statistical Society, ser B.* **51**, 449-458.
- [12] Rao, C.R. (1973). *Linear Statistical Inference and its Application*, 2nd Edition. New York: Wiley.
- [13] Rao, C.R. and Toutenburg, H. (1999). *Linear Models. Least Squares and Alternatives*. New York: Springer.
- [14] Samal, A. and Iyengar, P. (1992). Automatic recognition and analysis of human faces and facial expressions: A survey. *Pattern Recognition* **25**, 65-77.
- [15] Searle, S. R. (1971). *Linear Models*. New York: Wiley.
- [16] Small, C.G. (1996). *The Statistical Theory of Shapes*. New York: Springer-Verlag.
- [17] Watson, G.S. (1986). The shape of a random sequence of triangles. *Advances in Applied Probability* **18**, 159-169.