UNIVERSITY OF LEEDS

This is a repository copy of *Classification of form under heterogeneity and non-isotropic errors*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/111230/

Version: Accepted Version

Article:
Shuweihdi, F, Taylor, CC orcid.org/0000-0003-0181-1094 and Gusnanto, AS (2017) Classification of form under heterogeneity and non-isotropic errors. Journal of Applied Statistics, 44 (8). pp. 1495-1508. ISSN 0266-4763

https://doi.org/10.1080/02664763.2016.1214246

(c) 2016, Informa UK Limited, trading as Taylor & Francis Group. This is an Accepted Manuscript of an article published by Taylor & Francis in the Journal of Applied Statistics on 29 July 2016, available online: https://doi.org/10.1080/02664763.2016.1214246

Reuse
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

To appear in the *Journal of Applied Statistics*
Vol. 00, No. 00, Month 20XX, 2–14

Classification of Form under Heterogeneity and Non-Isotropic Errors

Farag Shuweihdi[a*] , Charles C. Taylor[a] and Arief Gusnanto[a]

[a] *Dept. of Statistics, University of Leeds,Leeds, LS2 9JT, UK*

(*v4.1 released December 2013*)

A number of areas related to learning under supervision have not been fully investigated, particularly the possibility of incorporating the method of classification into shape analysis. In this regard, practical ideas conducive to the improvement of form classification are the focus of interest. Our proposal is to employ a hybrid classifier built on Euclidean Distance Matrix Analysis (EDMA) and Procrustes Distance (PD), rather than Generalised Procrustes Analysis (GPA). In empirical terms, it has been demonstrated that there is notable difference between the estimated form and the true form when EDMA is used as the basis for computation. However, this does not seem to be the case when GPA is employed. With the assumption that no association exists between landmarks, EDMA and GPA are used to calculate the mean form and diagonal weighting matrix to build superimposing classifiers. As our findings indicate, with the use of EDMA estimators, the superimposing classifiers we propose work extremely well, as opposed to the use of GPA, as far as both simulated and real datasets are concerned.

Keywords: Data mining, classification, shape analysis, similarity, distance

1. Introduction

In reality, it is often the case that a number of species have distinct characteristics, in relation to shape. For instance, it is more frequently the case that the head of a female differs, in how it looks, from that of a male. The fact that someones hand during childhood differs from how it looks during adulthood suggests that shape may change according to factors, such as gender, growth and time. As an additional example of shape, someone may show diverse momentary facial emotional expressions, in response to a certain situation, such as joy or anger. One significant question could be raised as to whether or not an entitys shape is influenced by its position, rotation, and size (or basically, as explored in this paper, the position and rotation). Comprehending the previous phenomena requires us to create a tool that is based on statistics and through which shape can be understood. [6] developed a methodology, known as statistics-based shape analysis theory. This area has received extremely significant contributions, from outstanding sources of references about shape analysis development, such as [2] , [4] and [7]. In a broad range of fields, for instance psychology and biology, utilisation of shape analysis has been notable. As a result of great interest in it, textbooks covering a wide range of shape analysis aspects, have been produced, such as [14] and [3].
A configuration relating to geometry is described as **a form** (or size-and-shape), when it is invariant under translation and rotation. Typically, analytical investigation of form, in relation to a mathematics-related concept, requires an assortment of unique points,

*Farag Shuweihdi. Email: fstat2005@gmail.com

known as landmarks, to be depicted, by a specialist, on an object of interests image, as a first step. Putting it more explicitly, a **continuous** image is converted into a limited set of **isolated** (discrete) points (i.e landmarks). The significant characteristics of a configuration related to geometry are identified by these landmarks, denoted by a matrix in two or three dimensions (i.e coordinates). In the case of two dimensions, it is easy to denote landmarks coordinates through an actual or intricate scheme [3]. In this undertaking, real system is considered. Ordinarily, a matrix denotes a configuration, say $\mathbf{x}$, of order $k \times m$, where k is the number of landmarks, and m is a dimension such that $m \geq 2$. For an ideal alignment of matrices through a matching method, artificial variability caused by a set of unfavourable factors (i.e arbitrary location and orientation) presents itself as a key issue; commonly, this issue is addressed through restoring the factors of a predefined geometric alteration. Both the Ordinary Procrustes Analysis (OPA) and the Generalized Procrustes analysis GPA can be utilised to handle the issue through similarity alterations [4]. [1] examined the significant role individual landmarks play in the identification of shape variation.

In the case a certain form, one of the targets is identify the population class where it belongs. In other words, the configuration that is akin to a certain class and distinguishes it clearly from other classes is accurately identifiable. Form classification involves a basic requirement: presupposing that the errors around the k landmarks are *isotropic*. The non-fulfilment of isotropic presupposition means the presence of *non-isotropic* errors. Therefore, substandard performance is a possible outcome of the employment of any classification methodology, if this issue is not considered. Given that the calculated mean form is the basis of the superimposing classifiers suggested, it is it is important to note that, even under the condition of isotropy, [8] demonstrated that bias can be shown by the estimated mean form using GPA in the case of large errors. Additionally, when GPA is used under isotropic Gaussian errors, the calculated mean form is not a steady estimate of the true mean form [9]. The objective of classification faces a challenge as a result of these circumstances, given that it needs a technique through which mean and covariance matrix can be estimated to improve the performance of classification. Therefore, it is significant that weights showing landmarks degrees of influences are included in a classifier incorporating Procrustes distance. It is basically important that these weights adjust the information in the case of each landmark. In this research study, given the claim by [9] that GPA estimators show inconsistence under non-isotropy, Euclidean Distance Matrix Analysis (EDMA) [9] will be adopted, instead, to compute the mean form and weighting matrix to achieve a better classification precision. Incorporating Euclidean distance for inter-landmarks, [9] demonstrated that in $\mathbb{R}^2$, and under the Gaussian errors, the mean form produced by EDMA is calculated in a consistent manner, with the use of a bias corrected estimator. To examine whether or not an average form estimated by GPA and EDMA differs from the true form in cases of non-isotropy Gaussian errors, numerical examples have been provided here. Therefore, it can be established which method of computation is suitable to create a classifier. With both EDMA and GPA estimators employed to construct the superimposing classifiers, the purpose is to identify the most effective technique of computation, in relation to the improvement of the form classification performance. As claimed by [9], the superimposed methods of form or shape are unacceptable, in statistical terms. As far as the objective of classification is concerned, it is vital that the enhanced classification using a superimposing method is compared with a common technique of classification; the decision tree classifier will be adopted, with the use of transformed landmarks.

Section 2 introduces a weighted superimposing classifier employing PD. Section 3, includes our elucidation of how the variance of centred and non-centred landmarks will not be considerably different for many landmarks. A demonstration of the outcomes of the

superimposing classifiers using real and simulated configurations is included in Sections 4 and 5. Discussion is given Section 6. Lastly, conclusions are provided in Section 7.

2. Form Classification

For an assumed G labelled reciprocally exclusive form populations, say $\pi_1, \cdots, \pi_G$, let $\mathbf{X_g} = \{\mathbf{x}_{g,1}, \cdots, \mathbf{x}_{g,n_g}\}$ be a set of forms drawn i.i.d from class g, in which each $\mathbf{x}_{g,i} \in \pi_g$ and $\mathbf{x}_{g,i} \notin \pi_{g'}$ for $g \neq g'$. A configuration matrix $\mathbf{x}$ can, by means of translation and rotation, result from perturbation employing a linear Gaussian model [4] which suggests that all configurations are perturbed, translated and rotated and on a random scale:

$$\mathbf{x} = \left(\boldsymbol{\mu}_g + \mathbf{E}_g\right)\Gamma + 1_k B^T, \quad g = 1, \cdots, G. \tag{1}$$

in which $1_k B^T$ and Γ are nuisance factors (location and orientation matrix), i.e unwanted information not relevant to geometrical form information. $\boldsymbol{\mu}_g$ is the true mean and $\mathbf{E}_g$ is a $k \times m$ matrix of separate arbitrary errors, which, as per the norm, distributed with mean equal to zero and covariance matrix (form variation matrix) $\Sigma_{km,g} = \Sigma_{k,g} \otimes \Sigma_m$, in which $\Sigma_{k,g}$ is the covariance of the landmarks (the rows of $\mathbf{E}_g$), $\otimes$ is the Kronecker product and Σ_m is the covariance for the dimensions (the columns of $\mathbf{E}_g$). Our assumption here is $\Sigma_m = I_m$, which suggests that identical separate variability exists around each mean landmark. The case of $\Sigma_m = I_m$ was hypothesised in order to estimate the covariance matrix with the use of GPA by [4] and EDMA by [9]. **In this study, landmarks are assumed to be independent, and hence Σ_k is diagonal matrix**[comment] .

2.1 *Isotropic Model*

Under isotropic conditions, all classes are presupposed to share a mutual covariance matrix, say Σ, and $m = 2$ in which

$$\Sigma_g = \Sigma_{km,g}^{\text{comment 7}} = \Sigma = \sigma^2 I_{km}, \ g = 1, \cdots, G,$$

indicates sameness between populations so that $\Sigma_{k,g} = \sigma^2 I_k$. Given that $\boldsymbol{\mu}$ is unidentified, it is possible to employ GPA or EDMA to calculate the true mean. OPA distance for centred estimators is employed for the assigning of a new centred object into an identified class.

2.2 *Non-Isotropic Model*

For a mitigation of the non-isotropys negative impact on superimposition classifier performance, *weighted* PD is used. An initial allusion to this matter is Lissitz et all [11]. To directly obtain weighted PD, the least-squares technique is applied to two weighted landmark configurations. Different from PD, in which all landmarks are equally weighted, weighted PD provides less influence to the less accurate landmarks, but designates more weight to extremely precise landmarks−that is, linking the weight magnitude to the landmark quality.

This paper proposes a methodology that depends on the employment of a mean and covariance matrix calculated by EDMA in a superimposing classifier. In fact, given its dependence on the weighting matrix estimation method, no *unique* solution to the non-isotropic error problem exists. For an assessment of the advantages of using EDMA in

classification, the GPA-based calculation using GPA is employed to compute mean forms and weighting matrices. Additionally, for simulated configurations, the true parameters are utilised to build a superimposing classifier.
According to the Gaussian perturbation model, the mean form and variance-covariance matrix Σ_k^c **for centred configuration**[comment 1] can be evaluated consistently [9]. Therefore, for $\boldsymbol{\mu}_g$ and $\Sigma_{k,g} = \text{diag}(\sigma_{1g}^2, \cdots, \sigma_{kg}^2)$, EDMA is employed to identify the evaluated mean forms $\{\hat{\mathbf{m}}_g\}_{g=1}^G$ and covariance matrices $\left\{\hat{\Sigma}_{k,g}^c\right\}_{g=1}^G$ for centred configurations. The landmark weights are calculated by $(\hat{\Sigma}_{k,g}^c)^{-1}$ where the diagonal is parallel to the diagonal of the evaluated weights matrix $\mathbf{w}$.

2.3 *Superimposing Classifier*

The algorithm below constitutes the basis of a method through which a configuration matrix $\mathbf{x}_o$ coming from an unidentified form class can be assigned:

Algorithm 1 Superimposing Classifier

Estimate $\hat{\mathbf{m}}_g$ and $(\hat{\Sigma}_{k,g}^c)^{-1}$, $g = 1, \cdots, G$, using EMDA. Then, set $\mathbf{w}_g = \left[\text{diag}(\hat{\Sigma}_{k,g}^c)\right]^{-1}$.
Decomposing the diagonal matrix $\mathbf{w}_g$ so that $\mathbf{w}_g = \mathbf{w}_g^{1/2}\mathbf{w}_g^{1/2}$, the weighted PD, $d_{(\text{w})}$, between two weighted forms

$$d_{(\text{w}),g}^2(\mathbf{x}_o^c) = \text{tr}(\mathbf{w}_g^{1/2}\hat{\mathbf{m}}_g - \mathbf{w}_g^{1/2}\mathbf{x}_o^c\Lambda)^T(\mathbf{w}_g^{1/2}\hat{\mathbf{m}}_g - \mathbf{w}_g^{1/2}\mathbf{x}_o^c\Lambda), \quad g = 1, \cdots, G. \tag{2}$$

Where Λ is a rotation matrix estimated by OPA.
Assign **a new configuration, say $\mathbf{x}_o$, needs to be classified**[comment 9] to predicted class $g^*(\mathbf{x}_o)$ according to

$$g^*(\mathbf{x}_o) = \arg\min_{g=1,\cdots,G} d_{(\text{w}),g}^2(\mathbf{x_o}). \tag{3}$$

The common classifier employing PD is noticeably acquired through the replacement of $\mathbf{w}$ of (2) by $\mathbf{I}_{km}$ (identity matrix). Despite the fact that, unlike Procrustes algorithm, the estimators are simple, in terms of computation, $\hat{\mathbf{m}}$ acquired by EDMA can cause a reflection of $\hat{\mathbf{m}}$; and with the employment of EDMA to calculate the mean form, it is necessary that we compute $k(k-1)/2$, non-zero parameters for Euclidean distance matrix of order $k \times k$ acquired by inter-distance between k landmarks for the configuration matrix. This is a very high figure, as compared with the number of k parameters required for estimating the mean form, especially when k is high.

2.4 *Decision Tree*

Through the classification decision tree, which is a non-parametric technique, we are able to categorise new objects into one (and only one) predefined sub-class, commonly through the use of axis-parallel decision boundaries [5]. Although it requires extensive computation, this technique is extremely easy to comprehend. Given that it is a flexible method, the decision tree works well in the case of data being linearly separable ($\Sigma_g = \Sigma,\ g = 1, \cdots, G,$) and non-linearly separable ($\Sigma_g \neq \Sigma_{g'},\ g \neq g' = 1, \cdots, G,$).

Additionally, it does not require any probability assumptions. Given the nuisance parameters, it is important that the tree classifier is learnt on altered form landmarks imposed on a general mean form, say $\mathbf{x}'$ as follows:

Algorithm 2 Decision tree based on transformed landmarks

Apply GPA to a training set of configuration matrices of order $k \times m$, $\mathbf{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ to estimated transformed landmarks $\mathbf{X}^{'} = \{\mathbf{x}'_1, \cdots, \mathbf{x}'_n\}$ and general mean form $\hat{\mathbf{m}}$.
Each transformed matrix say $\mathbf{x}'_{k\times m}$, is arranged in a row vector,

$$\mathbf{x}'_{k\times m} \rightarrow (x'_1, \cdots, x'_{km}) = \boldsymbol{x}'_{1\times km}.$$

Use a matrix of transformed landmark variables $\boldsymbol{X}^{'}_{n\times km}$ to construct tree classifier.
Apply OPA to align a new configuration $\mathbf{x}_o$ on $\hat{\mathbf{m}}$.
The tree classifier is employed to allocate $\boldsymbol{x}'_o$ to an estimated class.

3. Variance of Centred Landmarks

For a set of configurations of size n **with independent landmarks,** $i.e$**, such that** $cov(x_j, x'_j) = \sigma_{jj'} = 0, \quad j \neq j' = 1, \cdots, k$**,** [main comment] the consequence of centring on landmark variance, say $\hat{\Sigma}^c_k$ is examined through the perturbation model,

$$\mathbf{x}_i = \boldsymbol{\mu} + \mathbf{E}_i, \quad i = 1, \cdots, n, \tag{4}$$

where

$$\mathbf{x}_i = \begin{bmatrix} x_{i1} & y_{i1} \\ \vdots & \vdots \\ x_{ik} & y_{ik} \end{bmatrix},$$

and $\mathbf{E}_i$ denotes an error matrix after a normal distribution with mean zero and diagonal covariance matrix $\boldsymbol{\Sigma}_k$. **Let** $(s^2_{x_j}, s^2_{y_j})$ **and** $(\bar{x}_j$**,**$\bar{y}_j)$ **,** $j = 1, \cdots, k$**, represent the sample variance and mean of landmark** j**.** [comment 20] Given that this research study examines Σ_k as being a diagonal matrix, the landmarks coordinates are separate, $cov(x_j, x_{j'}) = 0, \cdots, j \neq j'$. **The coordinates of landmarks are independent and identical to each other,** $\Sigma_m = I_m$**, which suggests that identical separate variability exists around each mean landmark.**[comment 21] Currently, in the event of the configurations acquired from model (4) being centred, we get

$$\mathbf{x}^c_i = \begin{bmatrix} x_{i1} - \bar{x}^*_i & y_{i1} - \bar{y}^*_i \\ \vdots & \vdots \\ x_{ik} - \bar{x}^*_i & y_{ik} - \bar{y}^*_i \end{bmatrix} = \begin{bmatrix} x^c_{i1} & y^c_{i1} \\ \vdots & \vdots \\ x^c_{ik} & y^c_{ik} \end{bmatrix},$$

where

$$\bar{x}^*_i = \frac{\sum_{j=1}^k x_{ij}}{k} \quad \text{and} \quad \bar{y}^*_i = \frac{\sum_{j=1}^k y_{ij}}{k}.$$

Thus, the sample variance for a centred landmark, say l, is

$$\begin{aligned} s_l^{c2} =& \frac{1}{n}\sum_{i=1}^{n}(x_{il}^c - \bar{x}_l^c)^2 \\ =& \frac{1}{n}\sum_{i=1}^{n}\left[\left(x_{il} - \frac{\sum_{j=1}^{k} x_{ij}}{k}\right) - \frac{1}{n}\sum_{i=1}^{n}\left(x_{il} - \frac{\sum_{j=1}^{k} x_{ij}}{k}\right)\right]^2 \\ =& \frac{1}{n}\sum_{i=1}^{n}\left[\frac{(k-1)}{k}\left(x_{il} - \frac{1}{n}\sum_{i=1}^{n} x_{il}\right) - \frac{1}{k}\sum_{j=1,j\neq l}^{k-1}\left(x_{ij} - \sum_{i=1}^{n}\frac{x_{ij}}{n}\right)\right]^2 \\ =& \frac{(k-1)^2}{k^2}s_l^2 + \frac{1}{k^2}\sum_{j\neq l=1}^{k-1} s_j^2 + \frac{k-1}{k^2}\text{covariance terms.} \end{aligned} \tag{5}$$

In relation to separate variables, the final term on the right hand side of Equation (5) is the covariance between landmarks. Since we consider the case where the landmarks are independent, the term is practically zero, i.e. $\sigma_{jj'} = 0, j \neq j' = 1, \cdots, k$. Given this, it is interesting to note that in Equation (5), s_l^{c2} will tend to s_l^2 when k increases, provided that $\sum_{j\neq l=1}^{k-1} s_j^2 \ll 1/k^2$ as k increases. This means that, under the previously mentioned condition, the sample variance of the centred landmark will be closer to the non-centred landmark when the number of landmarks increases to a large number. main comment

4. Description of Experiments and Analysis

In order to establish how precise the classification techniques proposed are, experiments are initially conducted with the use of simulation method. Triangles (a few landmarks) form the basis of two experiments, whilst facial expressions (many landmarks) form the basis of one experiment. Every experiment involves two classes of configuration matrices ($G = 2$), produced by a perturbation linear model 1. In relation to simulated data, by considering the true mean and covariance matrix, we will additionally be able to develop a greater insight about how effective EDMA and GPA are, as far as the performance of classification is concerned. In other words, through the utilisation of different methods in the estimation of parameters and creation of classifiers, one method can be preferred to another. A total number of nine classifiers will be built; superimposing classification using PD and WPD will from the basis of eight of them , whilst a decision tree will form the basis of one classifier. Real configurations involve two classes of facial expressions and gorilla skulls.

(1) **A**: Heterogeneity and non-isotropic errors, $d^2_{(\mathrm{w}),g}(\mathbf{x}) = \mathrm{tr}(\hat{\mathbf{m}}_g^c - \mathbf{x}^c\Lambda)^T(\hat{\Sigma}_{k,g}^c)^{-1}(\hat{\mathbf{m}}_g^c - \mathbf{x}^c\Lambda)$, where $\hat{\mathbf{m}}_{k,g}^c$ and $\hat{\Sigma}_{k,g}^c$ are estimated using EDMA.

(2) **B**: Homogeneity and non-isotropic errors, $d^2_{(\mathrm{w}),g}(\mathbf{x}) = \mathrm{tr}(\hat{\mathbf{m}}_g^c - \mathbf{x}^c\Lambda)^T(\hat{\Sigma}_{k}^c)^{-1}(\hat{\mathbf{m}}_g^c - \mathbf{x}^c\Lambda)$, where $\hat{\Sigma}_{k}^c$ is a common matrix estimated by EDMA.

(3) **C**: Homogeneity and isotropic errors, $d^2_{(\mathrm{w}),g}(\mathbf{x}) = \mathrm{tr}(\hat{\mathbf{m}}_g^c - \mathbf{x}^c\Lambda)^T(\hat{\mathbf{m}}_g^c - \mathbf{x}^c\Lambda)$, where $\hat{\mathbf{m}}_{k,g}^c$ is estimated using EDMA.

(4) **D**: Heterogeneity and non-isotropic errors, $d^2_{(\mathrm{w}),g}(\mathbf{x}) = \mathrm{tr}(\hat{\mathbf{m}}_g^c - \mathbf{x}^c\Lambda)^T(\hat{\Sigma}_{k,g}^c)^{-1}(\hat{\mathbf{m}}_g^c - \mathbf{x}^c\Lambda)$, where $\hat{\mathbf{m}}_g$ and $\hat{\Sigma}_{k,g}^c$ are estimated using GPA.

(5) **E**: Homogeneity and non-isotropic errors, $d^2_{(\mathrm{w}),g}(\mathbf{x}) = \mathrm{tr}(\hat{\mathbf{m}}^c_g - \mathbf{x}^c\Lambda)^T(\hat{\Sigma}^c_k)^{-1}(\hat{\mathbf{m}}^c_g - \mathbf{x}^c\Lambda)$, where $\hat{\Sigma}^c_k$ is a common matrix estimated by GPA.
(6) **F**: Homogeneity and isotropic errors, $d^2_g(\mathbf{x}) = \mathrm{tr}(\hat{\mathbf{m}}^c_g - \mathbf{x}^c\Lambda)^T(\hat{\mathbf{m}}^c_g - \mathbf{x}^c\Lambda)$, where $\hat{\mathbf{m}}_g$ is the mean form estimated by GPA.
(7) **G**: A tree classifier learned on transformed form landmarks.
(8) **H**: Homogeneity and isotropic errors, $d^2_g(\mathbf{x}) = \mathrm{tr}(\boldsymbol{\mu}^c_g - \mathbf{x}^c\Lambda)^T(\boldsymbol{\mu}^c_g - \mathbf{x}^c\Lambda)$, where $\boldsymbol{\mu}^c_g$ is known parameters.
(9) **I** : Heterogeneity and non-isotropic errors, $d^2_{(\mathrm{w}),g}(\mathbf{x}) = \mathrm{tr}(\boldsymbol{\mu}^c_g - \mathbf{x}^c\Lambda)^T(\Sigma_{k,g})^{-1}(\boldsymbol{\mu}^c_g - \mathbf{x}^c\Lambda)$, where $\boldsymbol{\mu}^c_g$ and $\Sigma_{k,g}$ are unknown parameters.

The use of GPA and EDMA for the calculation of the true mean form for each class can lead to the computed form mean being considerably different from the true form mean, given the technique of computation; it is therefore likely that the superimposing classifiers performance will differ. In relation to B simulated datasets, a sample of Goodall's F test is conducted to establish whether the total calculated mean form $\hat{\hat{\mathbf{m}}}$ produced by averaging B sample means of the simulated datasets, say $\hat{\mathbf{m}}^1, \cdots, \hat{\mathbf{m}}^B$, is equivalent to the true mean $\boldsymbol{\mu}_o$. The purpose is to examine the null hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_o$ for every class. Based on this test, a more effective method of computation will demonstrate that $\hat{\hat{\mathbf{m}}}$ acquired by GPA or EDMA has, systematically, a greater similarity to the true mean, $\boldsymbol{\mu}_o$.

4.1 *Simulated Triangles*

In this example, we consider two different triangles, where $k = 3$ and $m = 2$. The true means, $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, are

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 5 & 1 \end{bmatrix}, \quad \boldsymbol{\mu}_2 = \begin{bmatrix} 1.15 & 1.15 \\ 1.15 & 4.58 \\ 3.43 & 1.15 \end{bmatrix}.$$

$\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are set as such, in order for the distinction between the population mean forms is little; and, therefore, the techniques of calculation (GPA and EDMA) can be held to account for any statistical difference and classifier performance. Additionally, this experiment involves two covariance matrices for the purpose of examining the non-isotropic errors impact on the calculated mean forms and the behaviour of the classifiers. With the use of model (1), $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are perturbed by landmark error matrices with mean zero and covariance matrices:

(1) **Heterogeneity and non-isotropic errors**, $\Sigma_{k,1} = \mathrm{diag}(0.2, 0.8, .9)$, and $\Sigma_{k,2} = \mathrm{diag}(0.9, 0.2, .001)$, respectively.
(2) **Low heterogeneity and high non-isotropic errors**, $\Sigma_{k,1} = \mathrm{diag}(0.001, 0.4, 2.1)$ and $\Sigma_{k,2} = \mathrm{diag}(.005, .5, 1.9)$, respectively. Here, the degree of non-isotropy is higher than the first matrices.

The assessment of each classifier is performed with the use of a training and test set, each comprising 150 objects chosen from each class, in which $n_1 = n_2 = 150$. The total error rate of misclassification is assessed from 50 simulations.

The mean form is computed for every simulated dataset of size 150. The number of simulated dataset is 50, so the generated form means will be 50. The Goodall's F-test is then applied to the sample of size 50 generated form means. As shown by the findings presented in Table 1, the computed mean form using EDMA is, in statistical terms, the same as the true mean for the non-isotropic cases; meanwhile, the GPA estimator

Table 1. One sample of Goodall's F-test for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_o$ in terms of triangle configurations. **Significance keys**:* significant at 5%, ** significant at 1% and *** highly significant.

	Class 1		Class 2	
Experiment	EDMA	GPA	EDMA	GPA
Heterogeneity and non-isotropic errors	1.313	211.78***	1.487	266***
Low heterogeneity and high non-isotropic errors	1.301	1814***	.398	1209***

differs considerably from the true mean. **As shown by [9], employing an empirical example, using GPA, variance may be artificially increased at some landmarks and decreased at others.** comment 24

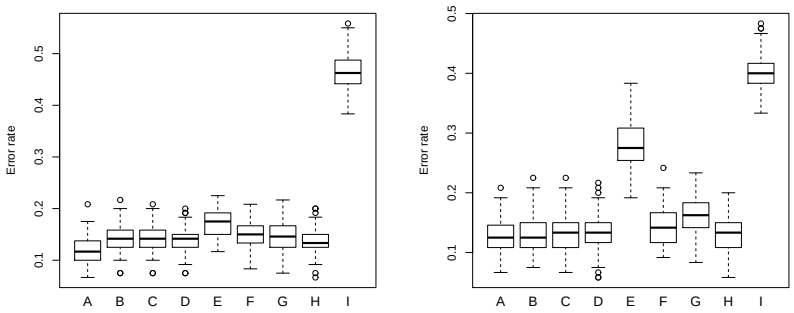

Figure 1. LHS: test error rate for two groups of simulated triangles for Heterogeneity and non-isotropic errors, whereas RHS: test error rate for Low heterogeneity and high non-isotropic errors.

Figure 1 summaries the result for the nine classifiers. The error rate of Classifier A, which is founded on the computation of two covariance matrices using EDMA, is the lowest among all classifiers. Additionally, employing a common covariance matrix, Classifier B, which is founded on EDMA estimators, displays better performance than Classifier E, which is founded on GPA estimators. In the event of superimposing classifications ignoring weighting matrices, Classifier H, which is founded on the true mean form, displays the lowest error rate; meanwhile, the precision of Classifier C employing the EDMA estimators mean from is marginally close to Classifier F employing the GPA mean form. Consequently, comparing between EDMA and GPA, the performance of classification will be improved by the addition of weights, through the use of EDMA, to superimposing classifiers, in non-isotropic cases. On the other hand, in the case of unweighted superimposing classifiers, it does not appear to make any significant difference which technique of calculation should be adopted for the mean form computation, despite the fact that the Goodall's F test prioritises the EDMA estimator. This is because the form of two classes may be different enough to lead to a close result, which is not the case when superimposing classifiers involve weighting matrices. Classifier I, employing the true covariance matrix, displays the worst performance, given the utter difference between the true covariance matrix and the covariance matrix of centred landmarks, when k is small as demonstrated by (5), where, for instance, a slight variation becomes more significant under the influence of a post-centring larger variation . Putting it more explicitly, centred configurations defined by the variance of a few landmarks, such as triangle example, may seriously be increased, in an artificial manner, at some landmarks, and reduced at others,

as opposed to the processes generating the variance owing to translation. Consequently, the true covariance matrix will become an inappropriate weighting matrix.
It is established that the performance of the decision tree technique is akin to unweighted superimposing classifiers employing EDMA and GPA estimators.

4.2 *Simulated Facial Expressions*

The previous two instances presented the simplest case of shape classification in relation to the number of landmarks. Two simulated facial expressions comprising $k = 22$ landmarks in dimension $m = 2$ are deemed as two classes of interest. While the first class is denoted by a *laugh*, the second is denoted by a *big smile*. Each simulated groups true mean is formulated through the calculation of the mean form of a real dataset of matrices [10] and [12]. Both expressions may involve quite similar forms, but the difference will be mainly lie in the perturbation technique, in which the focus is placed on low heterogeneity between classes and high non-isotropy within classes, see Figure 2. Every mean matrix of points is separately perturbed, through the use of the linear perturbation Gaussian model. In every class, $n_1 = 300$ and $n_2 = 300$. From all the datasets, 300 configurations are chosen in an arbitrary manner to train classifiers, whilst test data depend on the remaining configurations. The calculation of the total error rate is performed over 50 simulations.

Table 2. One sample of Goodall's F-test for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_o$ with respect to facial expressions. **Significance keys**:* significant at 5%, ** significant at 1% and *** highly significant.

Experiment	Big Smile 1		Laugh 2	
	EDMA	GPA	EDMA	GPA
Heterogeneity and non-isotropic errors	.874	1.631	1.149	1.643***

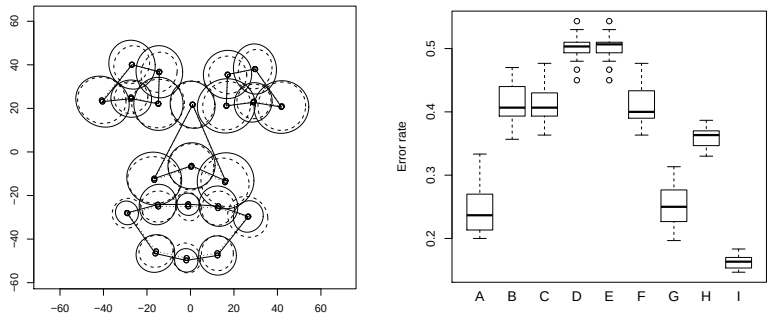

Figure 2. Left plot shows the variance contour of simulated facial expressions around the superimposed true mean forms for laugh (black lines) and big smile (dotted-red lines) class. Right plot shows the test error rate for simulated.

Table 2 includes the initial analysis in which Goodall's F-test was employed; again, this tests findings demonstrate that, with the use of EDMA, we accept the null hypothesis for the both classes. As indicated by the calculation of the mean form through the use of the GPA, only for lough class, we accept the other hypothesis. Generally, it is our observation that employing GPA for the computation of the mean form of facial expression produces almost the same precision as using EDMA, something that perhaps be due to the large size of landmarks and the small sample size ($B = 50$).

In relation to the findings related to classification, Classifier I which is founded on the true form means and covariance matrices displayed the lowest error rate. This discovery differs from the triangle instance, which displayed substandard performance with the use of true covariance matrices. The reason for this can be elucidated by (5), in which, with independence between landmarks permitted, the variance of centring landmarks will be extremely akin to that of non-centred landmarks, as long as a high number of landmarks exists. It is generally the case that superimposing classifiers using EDMA estimators display a far better performance than when utilising GPA estimators. With two weighting matrices employed, Classifier A employing EDMA produces the best result, as opposed to other classifiers employing EDMA or GPA, especially classifier D using GPA. In relation to a common covariance matrix, Classifier B employing EDMA produces a far better precision than Classifier E employing GPA does. Likewise, unweighted Classifier C employing EDMA outperforms Classifier F employing GPA. The decision tree classifier outperforms the GPA superimposing classifiers, which suggests that the tree is apparently less sensitive to heterogeneity and non-isotropy than to superimposing classifiers GPA estimators. It is noticeable that in the case of the tree algorithm, GPA is only employed to convert the dataset, rather than for classification.

5. Sexual dimorphism in gorilla, pan and pongo skulls

Sexual species dimorphism constitutes one of the key sources of variation in appearance between males and females belonging to the same species. The superimposing classifiers of interest and the decision tree technique will be adopted to differentiate between sexual dimorphism for three skull datasets: (1). Gorilla configurations, the sample size of gorilla configurations is 59 configurations, 29 male and 30 female, (2). Orangutan (pongo) the sample size is 60, 30 male and 30 female and (3). Chimpanzee (pan), the sample size is 54, 28 male and 26 female. Every configuration of skull comprises 8 labelled landmarks assessed in two dimensions ($m = 2$). The datasets are supplied by [13].

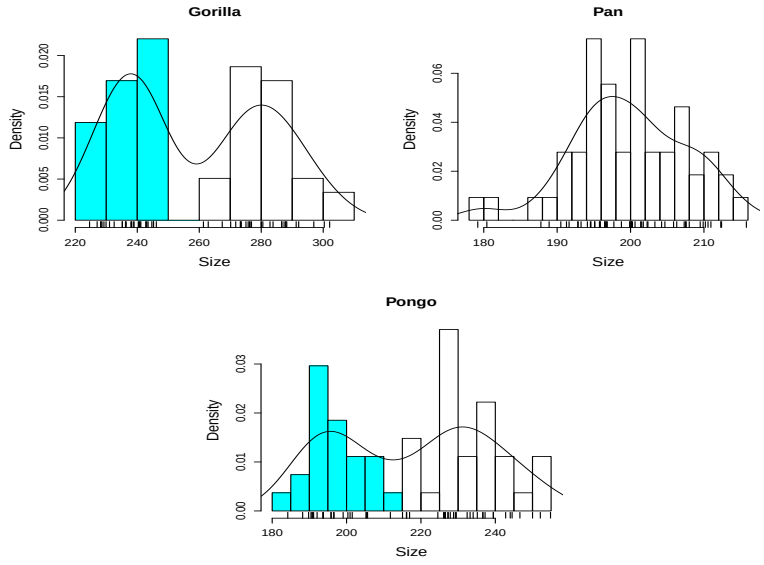

Figure 3. Histogram with estimated density function for the size of skulls. For the gorilla and pongo plots, colored bars represent female, whilst the uncolored bars represent males.

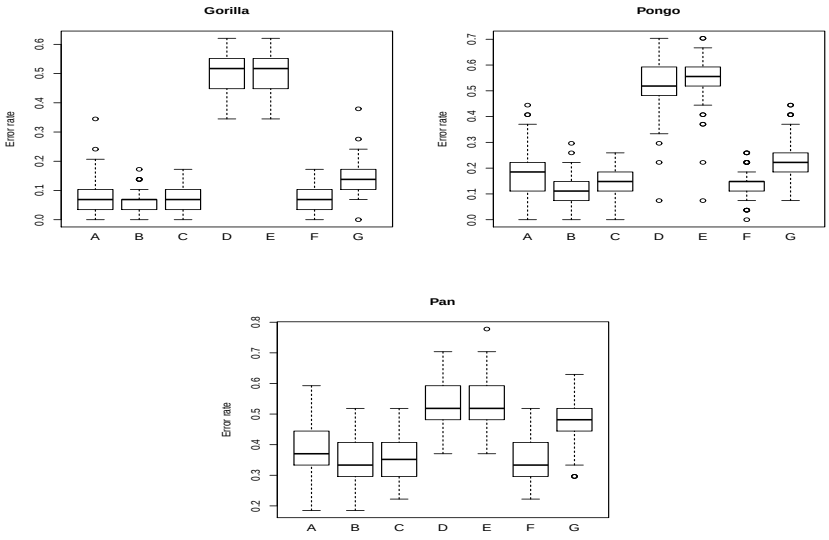

Figure 4. The test error rate for Gorilla, pongo and pan configurations.

Noticeably, the sexual dimorphism for the gorilla and pongo configurations involves a significant level of variation in size, with no overlapping, (see Figure

6. Discussion

As Equation 5 illustrates, when landmarks are only a few (such as the triangle instance), there is a significant difference between the variance of landmarks before and after centring. Consequently, based on Figure 1, Model I was weighted by the true variance matrix, but it displayed the highest error rate, because the calculated variance of centred landmarks and the true variance are considerably different. A high error rate was, therefore, displayed by Model I. Nonetheless, in relation to simulated facial configurations (Figure 2), there are 22 landmarks, a number that is bigger than the triangle instance; and based on Equation 5, in the event of a large number of landmarks, there could be equivalence in the variance before and after entering confutation. Therefore, in relation to simulated facial expressions, Model I displayed a better performance than that of the triangle instance (Figure 1). In relation to Figure 4, given that the data are real, Model H and Model I were not employed, because they are both founded on the true mean and covariance matrix. With regard to Model D and Model E (using GPA), in relation to the real data (sexual dimorphism in gorilla, pan and pongo skulls), given that a number of landmarks could be linked owing to biological variation, there will be a change in the true direction of variability around landmarks, as a result of employing GPA which, is not the case for EDMA. The employment of a weighting matrix produced by GPA therefore displayed a higher error rate than the other classifiers did. On the other hand, EDMA depends on a form matrix, which is a matrix of distances between all landmarks.

Given its flexibility, the decision tree method offers a good performance, whether in the case of the data being linearly separable ($\Sigma_g = \Sigma,\ g = 1, \cdots, G,$) non-linearly separable ($\Sigma_g \neq \Sigma_{g'},\ g \neq g' = 1, \cdots, G,$) , correlation between variables, and also it does necessitate any probability assumptions (strong against model specification).

To achieve an improvement in facial recognition, a number of advanced methods have

been utilised; these include PCA (Principal Component Analysis), ICA (Independent Component Analysis), LDA (Linear Discriminant Analysis), SVM (Support Vector Machine), and HMM (Hidden Markov Model). Comparing these techniques to the shape classifiers adopted in this study would be a thought-provoking step. Similarly, it would be stimulating to investigate the sample size impact and using different of. This will create a space for potential investigation of other aspects out of the realm of this paper. This study is mainly focused on investigating form classification under heterogeneity and non-Isotropic Errors, with the use of two distinct shape computation techniques (EDMA and GPA).

7. Conclusion

In this paper, the methods explored serve as tools for form classification, through the use of Procrustes distance in the non-isotropic cases. EDMA and GPA were adopted to calculate the mean form and weighting matrix. All the considered weighting matrices were covariance matrices, with the key reason being that the optimum method in which landmark quality is assessed can be produced when the variation around a mean landmark is considered. Additionally, the decision tree, which is a non-parametric tool, was considered in the evaluation of the conventional multivariate analysis performance.

Different from GPA, with EDMA employed to estimate the mean form, Goodall's F-test revealed that the sample mean, in statistically terms, was akin to the true form employing simulation instances of triangle and facial expression.

With GPA employed, the variance increased, in an artificial manner, at a number of landmarks and declined at others, as opposed to the processes generating the variance under the influence of translation or rotation. On the other hand, EDMA is founded on a orm matrix which is a matrix of distances between all landmarks that is unchanging in relation to translation or rotation (regardless of the orientation in which the landmark coordinate system was, the distances between the landmarks are still the same). When EDMA is not employed, the estimators of mean form and the variance covariance structure employing GPA are extremely intensive, in computational terms.

In relation to the separate landmarks, it was demonstrated that the centred landmarks variance was quite akin to parallel non-centred landmarks, in the event of a high number of landmarks. In relation to every dataset examined, with the addition of a weighting matrix through the use of EDMA, superimposing classifiers outperformed GPA and decision tree classifiers.

References

[1] M. T. Albert, H. Le, and C. G. Small, *Assessing landmark influence on shape variation*, Biometrika. 90 (3) (2002), pp. 669–678.
[2] F. L. Bookstein, *A statistical methods for biological shape comparison*, J. Theor. Biol. (107) (1984), pp. 475–520.
[3] [3] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis*, Wiley, New York,1998.
[4] C. R. Goodall, *Procrustes methods in the statistical analysis of shape*, J. Roy. Statist. Soc. (53) (1991), pp. 285–339.
[5] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, New York, 2001.
[6] D.G. Kenall, *The diffusion of shape*, Adv. in Appl. Probab. (9) (1977), pp. 428–430.

[7] J. T. Kent and K. V. Mardia, *Consistency of Procrustes estimators*. J. Roy. Statist. Soc. (59) (1997), pp. 285–399.
[8] J. T. Kent, *The complex bingham distribution and shape analysis*, J. Roy. Statist. Soc. (59) (1994) pp. 281–290.
[9] S. Lele, *Euclidean distance matrix analysis*, Math. Geol. (25) (1993) pp. 573–602.
[10] A. Linney, D. McDonald and A. Wright, *Alter ego-live video classification of facial expression*, in *Interdisciplinary Statistics and Bioinformatic*, S. Barber, P.D. Baxter, eds, Leeds University Press, UK, 2006, pp. 31-34.
[11] R. W. Lissitz, P. H. Schonemann, and J. C. Lingoes, *A solution to the weighted Procrustes problem in which the translation is in agreement with the loss function*, Psychometrika. (41) (1978), pp. 547–550.
[12] F. Shuweihdi, *Clustering and Classification with Shape Examples*, Ph.D. diss.,School of Mathematics, Department of Statistics, University of Leeds, 2009.
[13] P. O'iggins and I. L. Dryden, *Sexual dimorphism in hominoids: further studies of craniofacial shape differences in Pan, Gorilla, Pongo*, J. Hum. Evol. (24) (1993), pp. 183–205.
[14] C. G. Small, *The Statistical Theory of Shape*, Springer, New York,1996.