

DETECTION OF NON-AFFINE SHAPE OUTLIERS FOR MATCHED-PAIR SHAPE DATA

STANISLAV KATINA

ABSTRACT. Cleft lip/palate (CLP) is a relatively common birth defect so disfiguring that nowadays it is almost always corrected surgically as early as possible. The postnatal surgical correction does not, however, result in a normally growing upper jaw, but instead, owing to scar tissue, one that grows abnormally. It is important to decide if a clinical treatment group is homogeneous. The example involves data from digitally processed lateral X-ray films of 48 boys who have complete unilateral CLP but no other malformation. 22 landmarks were represented by their Procrustes shape coordinates, principal components of matched-pair differences were examined, and the distribution of the 48 shape changes was studied for outliers in the affine and non-affine subspaces of the full Procrustes shape and form space. To separate outliers from inliers we use bagplots. There are no outliers apparent in the affine subspace. In the non-affine subspaces, we found no outliers in the subspace of bending patterns at large scale but four outliers in the subspace of local changes at small scale. Almost the same outliers were found by form-space PCA. These latter are associated with possible *creases* of the corresponding thin-plate splines. In those cases we can use the same spline formalism to relax the outlying form to an inlier by optimal relaxation along the *curve décolletage* that weighs bending energy against Procrustes distance and stop relaxation on the fence. These maneuvers suggest a possibly novel and interesting fusion of the Procrustes-spline toolkit with outlier detection. They also have practical implications for craniofacial management of CLP follow-up as well as suggestive implications for outlier detection in applied craniometrics and anthropometrics more generally.

1. Introduction

Cleft lip/palate (CLP) is a relatively common birth defect (about 1 per 1000) so disfiguring that nowadays it is almost always corrected surgically as early

© 2012 Mathematical Institute, Slovak Academy of Sciences.

2010 Mathematics Subject Classification: Primary 62H25, 62H35, 62P10; Secondary 6209, 62H11, 62G35.

Keywords: shape outliers, relative warp analysis, matched-pair shape differences, bagplots, thin-plate splines, cleft lip/palate.

This research was supported by VEGA [2/0038/12] and Welcome Trust [WT086901MA].

as possible. The postnatal surgical correction does not, however, result in a normally growing maxilla (upper jaw), but instead, owing to scar tissue, one that grows abnormally. For the clinical biostatistician it is important to decide if a clinical treatment group is homogeneous. In this paper we comment on this question both as an example of geometric morphometrics (GMM) and as an example of the kind of question that craniofacial surgeons might propose in their role as physical anthropologists, students of the variation of human form.

2. Data

The example involves data from digitally processed lateral X-ray films (under standard conditions—focus-object distance is 370 cm, object–film distance is 30 cm, magnification is 8.1 %) of 48 boys who have complete unilateral CLP but no other malformation, born between the years 1985–1988. All were surgically corrected in infancy by the same team of surgeons at the Clinic of Plastic Surgery in Prague, Czech Republic. Primary cheiloplasty performed according to Tennison’s method at an average age of 9 month was associated with periosteoplasty (without nasal septum repositioning) using a 5–7mm wide and 15–20mm long periosteal flap obtained from the lateral maxillary segment [8]. All patients underwent palatoplasty at an average age of 5 years, always by method of push-back with pharyngeal flap surgery. The corrective procedures were implemented in the patients with persisting soft tissue deformations like lip and nose, lip or nose [9]. Data are from follow-up at two ages, 10 and 15 years (measured at both times), within the adolescent growth spurt.

3. Detection and relaxation of shape outliers

22 cephalometric points (landmarks) were represented by their Procrustes shape coordinates [1], principal components of matched-pair differences were examined, and the distribution of the 48 shape changes was studied for outliers in the affine and the non-affine subspace of the full Procrustes shape and form space [3]. To separate outliers from inliers we use bagplots (bivariate boxplots) as supplied by R software [5]. Along *curve décolletage*, we relax the outliers by *flattening the surface* in the direction to Procrustes mean shape.

Relative warp analysis (RWA) is modified PCA [1], where relative warps (RWs) are PCs with respect to the bending energy or inverse bending energy matrix. Let

$$\mathbf{x}_{P,i} = \text{Vec}(\mathbf{X}_{P,i}), \quad i = 1, 2, \dots, n; \quad n = 48$$

be a $2k$ -vector of Procrustes shape coordinates, where

$$\mathbf{X}_{P,i} = \left(\mathbf{x}_{P,i}^{(1)} : \mathbf{x}_{P,i}^{(2)} \right), \quad \mathbf{x}_{P,i}^{(d)} = \left(x_{i1}^{(d)}, x_{i2}^{(d)}, \dots, x_{ik}^{(d)} \right), \quad d = 1, 2.$$

DETECTION OF NON-AFFINE SHAPE OUTLIERS FOR MATCHED-PAIR SHAPE DATA

Let $\mathbf{x}_{D,i}$ be $2k$ -vectors ($k = 22$) of matched-pair differences of vectorized Procrustes shape coordinates,

$$\mathbf{x}_{D,i} = \mathbf{x}_{P,15,i} - \mathbf{x}_{P,10,i},$$

$\mathbf{x}_{P,15,i} = \text{Vec}(\mathbf{X}_{P,15,i})$ and $\mathbf{x}_{P,10,i} = \text{Vec}(\mathbf{X}_{P,10,i})$, \mathbf{S}_D be the covariance matrix of the data $\mathbf{x}_{D,i}$,

$$\overline{\mathbf{X}}_{P,10} = \left(\overline{\mathbf{x}}_{P,10}^{(1)} \vdots \overline{\mathbf{x}}_{P,10}^{(2)} \right) = (\overline{\mathbf{x}}_1, \dots, \overline{\mathbf{x}}_k)^T$$

be $k \times 2$ matrix of mean Procrustes shape coordinates $\overline{\mathbf{x}}_j$ of 10-year group, $j = 1, 2, \dots, k$. Let

$$\mathbf{L} = \begin{pmatrix} \mathbf{S} & \mathbf{1}_k & \overline{\mathbf{X}}_{P,10} \\ \mathbf{1}_k^T & \mathbf{0} & \mathbf{0} \\ \overline{\mathbf{X}}_{P,10}^T & \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad \mathbf{L}^{-1} = \begin{pmatrix} \mathbf{L}_{k \times k}^{11} & \mathbf{L}_{k \times 3}^{12} \\ \mathbf{L}_{3 \times k}^{21} & \mathbf{L}_{3 \times 3}^{22} \end{pmatrix},$$

where \mathbf{L} is symmetric positive definite, the inverse of \mathbf{S} exists as long as the landmarks are at least four in number, not all on one straight line, and also not in the same place (coincident); then inverse of \mathbf{L} exists and is equal to \mathbf{L}^{-1} , $\mathbf{S}_{js} = \phi(\overline{\mathbf{x}}_j - \overline{\mathbf{x}}_s)$; $j, s = 1, 2, \dots, k$, $\phi(\mathbf{x}) = \|\mathbf{x}\|_2^2 \log(\|\mathbf{x}\|_2^2)$, $\forall \|\mathbf{x}\|_2 > 0$, if $\|\mathbf{x}\|_2 = 0$, $\phi(\mathbf{x}) = 0$; $k \times k$ matrix $\mathbf{B}_e = \mathbf{L}^{11}$ is called *bending energy matrix* of $\overline{\mathbf{X}}_{P,10}$, $2k \times 2k$ matrix $\mathbf{B} = \mathbf{I}_{2 \times 2} \otimes \mathbf{B}_e$. There are these constrains of this matrix $\mathbf{1}_k^T \mathbf{B}_e = \mathbf{0}$, $\mathbf{X}^T \mathbf{B}_e = \mathbf{0}$, so the rank of the bending energy matrix is $k - 3$. Then the non-zero eigenvalues of $(\mathbf{B}^-)^{\alpha/2} \mathbf{S}_D (\mathbf{B}^-)^{\alpha/2}$ are \hat{l}_j with corresponding eigenvectors $\hat{\mathbf{g}}_j$ (PC loadings, RWs), $(\mathbf{B}^-)^{\alpha/2} = \sum_j \hat{\lambda}_j^{-\alpha/2} \hat{\gamma}_j^T \hat{\gamma}_j$ is Moore-Penrose generalized inverse of $\mathbf{B}^{\alpha/2}$. RW scores are defined as $r_{ij} = \hat{\mathbf{g}}_j^T (\mathbf{B}^-)^{\alpha/2} \mathbf{x}_{D,i}$. The effect of the j th RW can be viewed by plotting

$$\text{Vec}(\mathbf{X}_P(c_j, j, \alpha)) = \text{Vec}(\overline{\mathbf{X}}_{P,10}) \pm c_j \mathbf{B}^{\alpha/2} \hat{\mathbf{g}}_j \hat{l}_j^{1/2}, \quad r_j = c_j \hat{l}_j^{1/2}, \quad c_j \in \mathbb{R}^+ \quad (1)$$

for various values of $r_j \in \langle 0, \max(|r_{ij}|) \rangle$ (or some magnification of $\max(|r_{ij}|)$); alternatively, fixing $c_j = 1$, magnification of $\hat{l}_j^{1/2}$ as standard deviation of PC _{j} scores). The effect of the linear combination of RW₁ and RW₂ can be viewed by plotting

$$\text{Vec}(\mathbf{X}_P(c_1, c_2, \alpha)) = \text{Vec}(\overline{\mathbf{X}}_{P,10}) \pm c_1 \mathbf{B}^{\alpha/2} \hat{\mathbf{g}}_1 \hat{l}_1^{1/2} \pm c_2 \mathbf{B}^{\alpha/2} \hat{\mathbf{g}}_2 \hat{l}_2^{1/2}.$$

A PC summary of the shape data

$$\text{Vec}(\mathbf{X}_{P,15,i}(\alpha)) = \text{Vec}(\overline{\mathbf{X}}_{P,10}) \pm \mathbf{B}_e^{\alpha/2} \sum_{j=1}^2 r_{ij} \hat{\mathbf{g}}_j = \text{Vec}(\overline{\mathbf{X}}_{P,10}) + \sum_{j=1}^2 \hat{\mathbf{g}}_j \hat{\mathbf{g}}_j^T \mathbf{x}_{D,i}.$$

If $\alpha = 1$, bending patterns are at *large scale*, if $\alpha = -1$, bending patterns are at *small scale*, and if $\alpha = 0$, then we take $\mathbf{B}^0 = \mathbf{I}$ as the $2k \times 2k$ identity matrix and the procedure is exactly the same as classical PCA of matched-pair differences.

Any value of α other than zero converts the analysis to RWA of shape covariances with respect to bending energy instead of Procrustes distance.

To find the affine component we use linear regression model

$$\bar{\mathbf{x}}_{P,10}^{(d)} + \mathbf{x}_{D,i}^{(d)} = \bar{\mathbf{x}}_{P,10}^{(d)} \beta_i^{(d)} + \epsilon_i^{(d)}, \quad d = 1, 2; \quad i = 1, 2, \dots, n,$$

then

$$\mathbf{x}_{A,i}^{(d)} = \bar{\mathbf{x}}_{P,10}^{(d)} \hat{\beta}_i^{(d)} \quad \text{and} \quad \mathbf{X}_{A,i} = \begin{pmatrix} \mathbf{x}_{A,i}^{(1)} \\ \mathbf{x}_{A,i}^{(2)} \end{pmatrix}$$

is the affine component of $\mathbf{X}_{P,i}$. In affine subspace, \mathbf{S}_{DA} stands for sample covariance matrix of $\mathbf{x}_{DA,i} = \text{Vec}(\mathbf{X}_{A,i}) - \text{Vec}(\bar{\mathbf{X}}_{P,10})$. Then PCA of \mathbf{S}_{DA} is called *affine-subspace PCA*.

Let $\mathbf{X}_{DF} = (\mathbf{X}_D; \mathbf{x}_{size})$, be an $n \times (2k + 1)$ matrix with the rows equal to

$$\mathbf{x}_{DF,i} = \left(\mathbf{x}_{D,i}^T, \ln(CS_i) \right)^T, \quad i = 1, 2, \dots, n,$$

and

$$\mathbf{x}_{size}^T = (\ln(CS_1), \dots, \ln(CS_n)).$$

Let \mathbf{S}_{DF} be the covariance matrix of the data $\mathbf{x}_{DF,i}$. Then PCA of \mathbf{S}_{DF} is called *form-space PCA*. The first PC represents *allometry*—shape change during growth.

Visualization of interpolated shape changes can be done via thin-plate spline (TPS) deformation grids [1], field of vectors (within the convex hull of reference shape $\bar{\mathbf{X}}_{P,10}$, where longer vectors show stronger deformation in the specific direction of the shape change) superimposed with the grid of gray-scale rectangles with colors corresponding to the Procrustes distances (regions showing milder deformation are lighter, regions with stronger deformation are darker; the surface does not show the direction—but only the size—of some shape change).

Let $\mathbf{x} = \text{Vec}(\mathbf{X})$ be a $2k$ -vector of vectorized Procrustes shape coordinates. To relax a configuration \mathbf{x} onto the mean configuration \mathbf{x}_m , seek the configuration \mathbf{x}_r that minimizes the weighted sums of two energies. One term is the bending energy $(\mathbf{x}_r - \mathbf{x}_m)^T \mathbf{B} (\mathbf{x}_r - \mathbf{x}_m)$ of \mathbf{x}_r considered as a deformation of the mean form. The other term $(\mathbf{x}_r - \mathbf{x})^T (\mathbf{x}_r - \mathbf{x})$ is the sum of squares that corresponds to squared Procrustes distance between the relaxed landmark configuration \mathbf{x}_r and the vector \mathbf{x} actually encountered. For any regularization parameter (relative weight) λ , we seek the relaxed configuration \mathbf{x}_r that minimizes the weighted sum $(\mathbf{x}_r - \mathbf{x})^T (\mathbf{x}_r - \mathbf{x}) + \lambda (\mathbf{x}_r - \mathbf{x}_m)^T \mathbf{B} (\mathbf{x}_r - \mathbf{x}_m)$ of the energies. Setting the gradient of this expression to zero straightforwardly generates the solution

$$\mathbf{x}_r \equiv \mathbf{x}_r(\lambda) = (\mathbf{I}_{2k} + \lambda \mathbf{B})^{-1} (\mathbf{x} + \lambda \mathbf{B} \mathbf{x}_m). \quad (2)$$

As λ varies, $\mathbf{x}_r(\lambda)$ traverses a smooth curve in the space of landmark configurations, named *curve décolletage* [2]. Note that when $\lambda = 0$ then $\mathbf{x}_r(0) = \mathbf{x}$,

in the absence of any penalty for deviation from the mean configuration \mathbf{x}_m , the best fit to any data is the selfsame data. In the other limit $\lambda \rightarrow \infty$, \mathbf{x}_r tends to the configuration of zero bending having the least Euclidean cost (ordinary least square fit to the data, in the affine subspace). Because \mathbf{B} is a function of \mathbf{x}_m , this formulation is not symmetric in \mathbf{x} and \mathbf{x}_m . By minimizing quadratic variation, the relaxation is trying to *flatten the surface* (to make the surface flat, linear) to the maximum extent consistent with the data.

To generalize the median to higher-dimensional settings, a variety of different maximum depth estimators have been introduced. They extend, for example, the halfplane location depth of Tukey [7]. Let X_1, \dots, X_n be independent and identically distributed bivariate random variables, $X_i \in \mathbb{R}^2$, $i = 1, 2, \dots, n$. For given observations $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$, $\mathbf{x}_i \in \mathbb{R}^2$, we write $\mathbf{x}_i = (x_{i1}, x_{i2})^T$, where x_{ij} are RW_{*j*} scores (here $j = 1, 2$). The *halfplane location depth* of an arbitrary point $\theta \in \mathbb{R}^2$ relative to \mathbf{x}_i is defined as $d_i(\theta, \mathbf{x}) = \min_H \# \{i : \mathbf{x}_i \in H\}$, $i = 1, \dots, n$, where H ranges over all closed halfplanes of which the boundary line passes through θ . The *deepest location* is defined as $\hat{\theta}_{dl} = \arg \max_{\forall \theta} d_i(\theta, \mathbf{x})$ and often called the *Tukey median* [6].

The *depth region of depth s* is defined as the set D_s of points θ with $d_i(\theta, \mathbf{x}) \geq s$. Equivalently, D_s is the intersection of all closed halfplanes that contains at least $n - s + 1$ observations, hence D_s is a bounded convex polygon, and $D_{s+1} \subset D_s$. The boundary of D_s is a convex polygon, which is called the *contour of depth s* (*sth depth contour*, *sth hull*). Therefore, each vertex of a depth contour is the intersection point of two lines, each passing through two observations.

The univariate boxplot is based on the ranks since the box goes from the observation with the rank $\lfloor \frac{n}{4} \rfloor$ to that with the rank $\lceil \frac{3n}{4} \rceil$, and the central bar of the box is drawn at the median. A natural generalization of ranks to multivariate data is the notion of halfspace depth [7]. Using this concept in shape analysis, we propose a *bivariate boxplot (bagplot)* with the *bag*, that contains 50% of the data points \mathbf{x}_i , a *fence* that separates inliers from outliers and a *loop* indicating the points outside the bag but inside the fence [6]. The loop plays the same role as the two whiskers in one dimension, so we could call this plot also “*bag-and-bolster plot*” to stress the analogy with “*box-and-whisker plot*”. Like the univariate boxplot, bagplot visualizes several characteristics of the data—location, spread, correlation, skewness, and tails. The construction of bag B is as follows. Let $\#D_s$ denote the number of points contained in D_s . One first determines the value s for which $\#D_s \leq \lfloor \frac{n}{2} \rfloor < \#D_{s-1}$ and then interpolates linearly between D_s and D_{s-1} , relative to $\hat{\theta}_{dl}$, to obtain the set B (the bag B is a convex hull). The fence is obtained by inflating B , relative to $\hat{\theta}_{dl}$, by a factor 3 [8]. The loop contains all points between the bag and the fence and its outer boundary is the convex hull of the bag and the non-outliers. The points outside the fence are flagged as the outliers.

Above mentioned Tukey depth and depth contours can be directly applied to PC_1 and PC_2 scores to detect possible outliers [4]. Furthermore, the bagplot fence can be used as stop-rule for relaxation of the configuration \mathbf{x} , where this rule can be seen as *multivariate winsorization*.

4. Results and conclusions

In the non-affine (non-uniform) subspaces using RWA, we find no outliers in the subspace of bending patterns at large scale but four outliers in the subspace of local changes at small scale (Fig. 1; see also equation (1)).

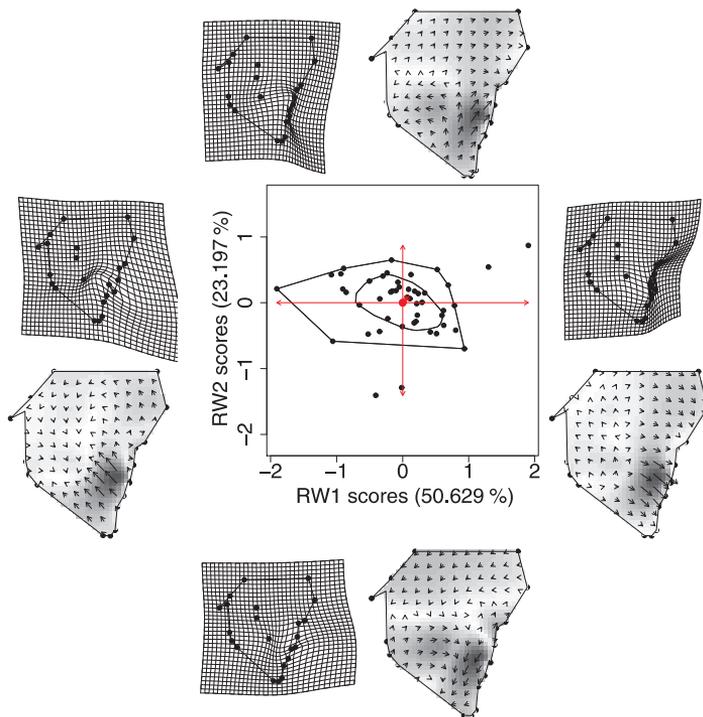


FIGURE 1. Results of RWA of matched-pair differences in non-affine subspace of local changes with small scale ($\alpha = -1$; RW_1 and RW_2 scores with the bag and the fence of the *bagplot* (middle); TPS deformation grids and field of vectors superimposed with the grid of gray-scale rectangles with colors corresponding to the Procrustes distances of the mean shape $\bar{\mathbf{X}}_{P,10}$ to the $\text{Vec}(\mathbf{X}_P(c_j, j, \alpha))$ (see equation (1)) indicated by the tips of the arrow heads in the middle figure).

Almost the same outliers are found by form-space PCA (Fig. 2). These latter are associated with possible *creases* (extrema of directional derivative; [2])

DETECTION OF NON-AFFINE SHAPE OUTLIERS FOR MATCHED-PAIR SHAPE DATA

of the corresponding TPS. In those cases we can use the same spline formalism to relax the outlying form to an inlier by optimal relaxation along the *curve décolletage* that weighs bending energy against Procrustes distance (Fig. 3; see also equation (2)).

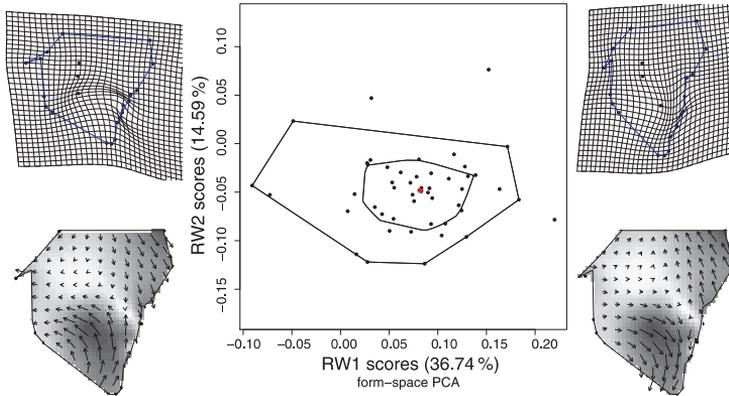


FIGURE 2. Results of form-space PCA of matched-pair differences; PC_1 and PC_2 scores with the bag and the fence of the *bagplot* (middle); TPS deformation grids and field of vectors superimposed with the grid of gray-scale rectangles with colors corresponding to the Procrustes distances of the mean shape $\bar{\mathbf{X}}_{P,10}$ to the $\text{Vec}(\mathbf{X}_P(c_1, 1))$ in PC_1 .

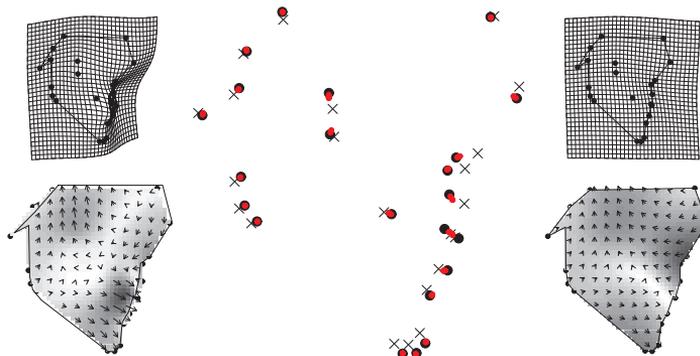


FIGURE 3. Relaxation of Procrustes shape coordinates; TPS deformation grids and field of vectors superimposed with the grid of gray-scale rectangles with colors corresponding to the Procrustes distances of the shape $\mathbf{X}_{P,10,29}$ to the shape $\mathbf{X}_{P,15,29}$ (left) and to the final relaxed shape $\mathbf{X}_{P,15,29}$ (right); *curve décolletage* of the shape $\mathbf{X}_{P,15,29}$ (\times —shape $\mathbf{X}_{P,10,29}$, big \bullet —shape $\mathbf{X}_{P,15,29}$, small red \bullet —relaxed shapes $\mathbf{X}_{P,15,29}$; middle); the difference of $\mathbf{X}_{P,15,29}$ and $\mathbf{X}_{P,10,29}$ in the subspace of the first two RWs is visualized in the scatter-plot of Fig. 1 as the most right \bullet .

STANISLAV KATINA

For biostatisticians, these maneuvers suggest a possibly novel and interesting fusion of the Procrustes-spline toolkit as the part of GMM with another theme, outlier detection, of the standard modern data-analytic toolkit. They also have practical implications for craniofacial management of CLP follow-up as well as suggestive implications for outlier detection in applied craniometrics and anthropometrics more generally.

Acknowledgements. The author thanks A. Bowman, F. L. Bookstein and P. O'Higgins for comments and J. Veleminská for data acquisition, pre-processing, and valuable comments.

REFERENCES

- [1] BOOKSTEIN, F. L.: *Morphometric Tools for Landmark Data—Geometry and Biology*. Cambridge University Press, Cambridge, 1991.
- [2] BOOKSTEIN, F. L.: *Creases as local features of deformation grids*, Med. Image Anal. **4** (2000), 93–110.
- [3] DRYDEN, I. L.—MARDIA, K. V.: *Statistical Shape Analysis*. John Wiley, New York, 1999.
- [4] KATINA, S.: *Geometrical Aspects of Statistical Size and Shape Analysis: Multivariate Statistical Models and Statistical Inference*. Habilitation Thesis, Comenius University, Bratislava, 2008.
- [5] R DEVELOPMENT CORE TEAM: *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, 2011.
- [6] ROUSSEEUW, P. J.—RUTS, I.—TUKEY, J. W.: *The bagplot: a bivariate boxplot*, Amer. Statist. **53** (1999), 382–387.
- [7] TUKEY, J. W.: *Mathematics and the picturing of data*, Proc. Int. Congress Math. **2** (1975), 523–531.
- [8] VELEMÍNSKÁ, J.—KATINA, S.—ŠMAHEL, Z.—SEDLÁČKOVÁ, M.: *Analysis of facial skeleton shape in patients with complete unilateral cleft lip and palate: Geometric morphometrics*, Acta Chirurgiae Plasticae **48** (2006), 26–32.
- [9] VELEMÍNSKÁ, J.—ŠMAHEL, Z.—KATINA, S.: *Development prediction of sagittal intermaxillary relations in patients with complete unilateral cleft lip and palate during puberty*, Acta Chirurgiae Plasticae **49** (2007), 41–46.

Received October 31, 2011

*School of Mathematics and Statistics
The University of Glasgow
University Gardens
G12 8QW Glasgow
SCOTLAND, UK*

*Department of Applied Mathematics and
Statistics
Comenius University
Mlynská dolina
SK-842-48 Bratislava
SLOVAKIA
E-mail: stanislav.katina@glasgow.ac.uk*