JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, May 2012, Vol. 57, No. 3
doi: 10.1111/j.1556-4029.2011.02000.x
Available online at: onlinelibrary.wiley.com

TECHNICAL NOTE

ANTHROPOLOGY

Jan Kalina,[1] *Ph.D.*

Facial Symmetry in Robust Anthropometrics*

ABSTRACT: Image analysis methods commonly used in forensic anthropology do not have desirable robustness properties, which can be ensured by robust statistical methods. In this paper, the face localization in images is carried out by detecting symmetric areas in the images. Symmetry is measured between two neighboring rectangular areas in the images using a new robust correlation coefficient, which down-weights regions in the face violating the symmetry. Raw images of faces without usual preliminary transformations are considered. The robust correlation coefficient based on the least weighted squares regression yields very promising results also in the localization of such faces, which are not entirely symmetric. Standard methods of statistical machine learning are applied for comparison. The robust correlation analysis can be applicable to other problems of forensic anthropology.

KEYWORDS: forensic science, anthropology, robust image analysis, correlation analysis, multivariate data, classification

The aim of this paper is to apply robust statistical methods to image analysis of faces with possible applications in forensic anthropology. We use robust correlation analysis to search for areas with the highest level of symmetry in a database of images of faces. This gives the solution to the task of localizing the faces, which are localized as the areas with the best axial symmetry. While classical biometrical methods of forensic anthropology and anthropometrics (1,2) are vulnerable with respect to noise, occlusion, or small violation of symmetry, the robust approach yields reliable results also for rotated and asymmetric faces. The methods of statistical learning are used for comparison in both standard and nonstandard situations. Robust approaches based on robust correlation analysis turn out to be the winner of the comparisons and can be recommended for forensic anthropology applications.

The first methods of multivariate statistical analysis were developed at the beginning of the 20th century for anthropological applications in physical anthropology. The concepts of correlation analysis, classification analysis, and statistical diversity measures (distances) were developed by researchers analyzing anthropological measurements (F. Galton, K. Pearson, R. A. Fisher, R. C. Rao). Only later, these statistical methods spread to other branches of research. These methods need some refinement reflecting the current development of statistics, mainly concerning the high vulnerability of the methods to noise (outlying values). Classical multivariate statistics also assumes the biometric variables to follow the Gaussian normal distribution (3). In this paper, a new robust correlation coefficient is described as a new general statistical measure without special assumptions, and its performance in face localization in images is examined. It is based on down-weighting individual pixels allowing us to measure the violation of individual pairs of pixels from the symmetry. The weights are determined automatically by robust statistical methods.

Identification of groups is the common subject of many papers in physical anthropology (4). Both face detection and face recognition can be solved by classification analysis, which is a statistical method for group identification. Face detection classifies each part of the image as a face or nonface (area not corresponding to a face). Face recognition classifies each face to a particular person of the given database. For both contexts, the usual methods of forensic anthropology start by a dimension reduction (principal components, Fourier transform, discrete cosine transform, or wavelet transform) and feature extraction to describe the differences among images or their groups and the contribution of variables to these differences (5). The classification analysis in forensic applications is often carried out by neural networks or support vector machines.

We illustrate the nonimpugnable power of the robust correlation analysis on the task of localizing the face even without a preliminary transformation of the images. Avoiding the initial reduction of dimension and feature extraction enables also a clear interpretation without assuming a mathematical model, allowing us to explain the importance of particular parts of the face or individual pixels on the classification. Usual classification procedures (6) are not capable enough to extract information from raw images, and this is the motivation for the common usage of dimension reduction and feature extraction, which import additional variability to the data or lose some information. Standard approaches of forensic analysis of images have their advantages (robustness to illumination, size, or rotation of the face), but they lack robustness. Moreover, they are often organized as a complicated cascade of extremely simple classification methods with numerous parameters (7). Such methods are, however, tailor-made for particular problems, and their numerous parameters cannot be tuned and are not suitable for a general usage in anthropology (4). Also, the assumptions of classification methods may not be fulfilled in practical situations in anthropology, yielding results opposing the intuition. Therefore, new methods with a clear interpretation would be desirable.

[1]Center of Biomedical Informatics, Institute of Computer Science AS CR, Pod Vodárenskou věží 2, 182 07 Praha 8, Czech Republic.

*Fully supported by "Center of Biomedical Informatics," Project 1M06014 of the Ministry of Education, Youth and Sports of the Czech Republic.

Received 7 Oct. 2010; and in revised form 19 Jan. 2011; accepted 29 Jan. 2011.

Robust statistical methods (8) represent such a new paradigm applicable to forensic anthropology, allowing us to obtain a robust solution with respect to noise in the image or occlusion of the face, resistant to modifications of hair style, not relying on the assumption of normal distribution of the data. While their history goes back to 1960s (9), only recently they started to penetrate to applications to different fields including biostatistics (10) and microarray image analysis for medical applications (11). They have not been widely applied to image analysis of faces (12). Here, we compare different robust measures of similarity between two images.

Forensic identification is often based on the face, which is assumed to be unique allowing us to identify a particular person (13). In this paper, the search for the best symmetry in the image is based on the fact that the intra-object variability (violation of symmetry in the face) is smaller compared to the inter-object variability (similarity of a face to a nonface or perhaps to the face of another person). Indeed, the detailed study of the robust correlation analysis is the key novel approach of this work.

Symmetry is a basic feature of faces that enhances recognition and reconstruction of faces (14). Nevertheless, fast and effective symmetry detection is still a difficult problem in computer vision (15). Symmetry of the face is one of the anthropological features invariant in time. Also, human brain evaluates symmetry of each face during the process of face recognition (16) and considers very symmetric faces to be more aesthetic. Although faces are not entirely symmetric, their symmetry is assumed, for example, in the study by Wiskott et al. (17) or Würtz (18) for the problem of face detection and recognition. Kalina (19) optimizes templates to locate landmarks in faces also assuming symmetry of the mouth and eyes with a very reliable performance of the resulting symmetric templates. In this paper, we measure the violation of symmetry in each face, and the method allows us also to interpret which pixels contribute to this violation. Locating the axis of facial symmetry is a special case of the landmark registration or shift registration applied, for example, by Ramsay and Silverman (20) in a paleopathological study of bone shapes. Symmetry is typically analyzed by shape analysis, which is based on landmarks at edges of the face (14, 21). Our approach studies the symmetry of gray intensities of the image rather than the symmetry of the contour of the face.

Forensic sciences work with uncertainty (22) and differences among individuals. Nevertheless, the symmetry measured by various robust correlation coefficients turns to be reliable for face localization for every face. Robust methods are more suitable especially for asymmetric and occluded faces, for example, a robust correlation coefficient is able to trim away pixels corresponding to asymmetric hair or occlusion. The robust approach can be applied also for the analysis of the partially destroyed skull of a decomposed corpse, where the robust approach can down-weight or ignore the problematic parts. The robust measure of similarity between two images can be perceived as an alternative to the Procrustes registration in the study by Mallett et al. (13).

Other possible applications not considered in this paper include the identification of a given face by means of the robust correlation analysis, the analysis of images in personal identity documents, or forensic stomatology (perhaps in combination with template matching) for the identification of surgical interventions, injuries, or anomalies in the X-ray image of the skull or teeth. Symmetry can be also used together with morphological and metric characteristics in the superprojection of the skull to a 2D image of the face to perform the postmortem face recognition. The study of symmetry of the face in 2D images is an important training problem and preparation for a 3D study of faces. While the (nonrobust) 3D analysis of faces is already developed (22), this paper only fills the gap of robust statistical methods in the study of symmetry in 2D images of faces.

This paper has the following structure. The next section describes the methodology used for the search of the vertical symmetry in the face, namely standard and robust versions of the correlation coefficient and methods of multivariate statistical analysis. Further sections present the results and a discussion with conclusions.

Materials and Methods

We work with a database of images of faces taken at the Institute of Human Genetics, University of Duisburg-Essen, Germany (projects BO 1955/2-1 and WU 314/2-1 of the German Research Council). This database contains 212 gray scale images of the whole faces of different persons between the ages of 18 and 35 years. The database contains 92 images of men (43%) and 120 women (57%). No two images correspond to the same person. The persons were selected as a random sample from persons with German origin.

Each of the original images is a matrix with the same size 192×256 pixels. A gray value in the interval [0,1] corresponds to each pixel, where small values are black and large values are white. Images are taken under the same conditions, and the Institute tried to have the images standardized as much as possible. There is exactly one face of a person looking straight at the camera on every image. The faces have about the same size and contain no facial expressions. Some faces are, however, rotated in the plane by small angles.

The Institute of Human Genetics uses images of faces to carry out genetic research with the aims to classify automatically genetic syndromes from an image of the face, to examine the connection between the genetic code and the size and shape of facial features, and also to visualize a face based only on its biometric measures. This research starts by a careful manual identification of the landmarks in faces by an anthropologist trained in this field. Some of the results of the genetic research are described in Böhringer et al. (23).

Correlation Analysis

In this work, the search for the face as the most symmetric area in the image is based on comparing two rectangular parts of the image. Pearson correlation coefficient is the most commonly used correlation measure in forensic sciences (24). The study by Pelin et al. (25) is an example of a study using Pearson correlation coefficient, which turns out not be statistically significant. In our study, we have weaker results with Pearson correlation coefficient as well, but obtain improved results with a robust correlation coefficient.

Figure 1 illustrates the main idea, where two neighboring rectangles of width 30 pixels are considered and compared. In the whole image, we take any two neighboring vertical rectangular strips of the same width (30–70 pixels) and perform the mirror reflection of one of them. The similarity between the two rectangles is measured by means of various measures of similarity between two images, including robust measures. We stress that the rectangles are direct neighbors without any boundary between them. The face is not expected at the very right or left boundary of the image. We compare the following correlation measures, which are standard measures of similarity between two images (6, 21):

- Pearson product–moment correlation coefficient r, which will be shortly called correlation coefficient;

FIG. 1—*Localization of the face by searching for the best vertical symmetry in the image. The similarity between the two rectangles of width 30 pixels is compared by different statistical methods. Particularly the radial weights are connected to the marked joint midpoint of the two rectangles.*

FIG. 2—*Localization of the face by searching for the best vertical symmetry in the image. The similarity between the two rectangles of width 60 pixels is compared by different statistical methods. Ignoring the black boundary areas of the rectangles ensures a method robust to rotation of the face.*

- Weighted Pearson product–moment correlation coefficient r_W with weights w; and
- Spearman's rank correlation coefficient r_S.

The coefficient r_W is a weighted analogy of r and is equivalent to the weighted coefficient of determination in linear regression. Radial weights allow us to stress the face in comparison with other parts of the rectangles, because they down-weight pixels far from the midpoint of the middle vertical line, in other words from the midpoint of the two rectangles taken together as one large area. For example, there are two neighboring rectangles of size 192×30 pixels in Fig. 1, and the virtual midpoint is marked, which serves as a reference point for the computation of radial weights. The weight in a particular pixel is defined to be inversely proportional to its distance from the joint midpoint of the two rectangles.

Robustness of the Correlation Analysis

The vulnerability of classical correlation analysis to outliers is well known (26). The Pearson product–moment correlation coefficient r *is* vulnerable to outlying values in the data (26). Kalina (27) studies the robustness of r_W in the context of template matching and proves the method based on r_W to be robust with respect to small occlusion, small violations of symmetry, illumination changes, or rotation by a small angle.

Symmetry and Rotation

The aim is to search for the symmetry in images rotated by a small degree. Only the kind of rotation in a frontal plane will be considered, where the whole face is displayed exactly from the front. We keep the image and rotate the rectangles, which is equivalent to retaining the rectangles vertical and rotating the face. However, rotating the rectangles moves their corners outside the image, and the number of such lost pixels depends on the angle of rotation.

Therefore, instead of the rectangles (Fig. 1), we modify the approach and consider their intersections with a circle. Only such pixels belong to the picture for each rotation. Figure 2 shows an example with modified rectangles of width 60 pixels, where the black corners are ignored during the computations in this part of the study. We do not examine rectangles at the right or left edge of the image.

Robust Correlation Coefficient

We describe several robust versions of a correlation coefficient with a high resistance against noise or outliers and apply them to our study. Different proposals of robust correlation measures are presented by Shevlyakov and Vilchevski (26); these do not, however, perform well in image analysis (20). Robust versions of the correlation coefficient are defined based on robust regression estimators, mainly the least weighted squares (LWS) and least trimmed squares (LTS), which are linear regression estimation procedures with desirable properties.

The LWS regression (28) is a robust regression method resistant to a larger portion of noise or outliers in the data. The idea is to down-weight less reliable observations (possible outliers), while the most credible and typical data points obtain the largest weights. The weights are assigned to particular data points automatically during the computation of the estimator. In our case, the weights represent an important diagnostic tool explaining the violation of symmetry in individual pairs of pixels, because smaller weights correspond to pixels that contribute to the violation of the symmetry.

To be specific, let us consider the model

$$Y_i = b_0 + b_1 x_{i1} + \ldots + b_p x_{ip} + e_i, \; i = 1, 2, \ldots, n$$

For a particular value of the estimate b of the parameter β, let us denote the residual of the ith observation by $u_i(b) = Y_i - b_0 - b_1 x_{i1} - \ldots - b_p x_{ip}, i = 1, 2, \ldots, n$. Let us denote the ith-order value among the squared residuals by $u^2_{(i)}(b)$. While only the sizes of the nonnegative nonincreasing weights $w_1, w_2, \ldots, w_n$ are chosen before the computation, the LWS estimator b_{LWS} minimizes

$$\sum_{i=1}^{n} w_i u^2_{(i)}(b)$$

over all possible values of b.

We define the robust correlation coefficient r_{LWS} based on the LWS as the weighted correlation coefficient r_W with the weights determined by the LWS. These properties of the coefficient follow from the high robustness of the LWS regression (28). For noncontaminated data sets, its properties resemble those of Pearson correlation coefficient. For the computation of the LWS-based correlation coefficient, an approximative algorithm is obtained as the weighted analogy of Rousseeuw and Van Driessen (29). In our computations, we use linearly decreasing weights or radial weights. A two-stage adaptive procedure for the weight selection is proposed by Rudin and Inman (30).

The LTS estimator (8) is a special case of the LWS estimator with weights equal to 1 or 0, while the weight 1 is assigned to the total number of h ($n/2 < h < n$) data points. This corresponds to the least squares regression on the h data points ignoring completely the remaining data. The advantage is the automatic detection of the points that should be trimmed away. The robust correlation coefficient r_{LTS} based on the LTS is defined in the same way as r_W with weights that are equal to 1 or 0 only.

For comparison, we also use the M-estimator correlation coefficient proposed by Shevlyakov and Vilchevski (26); however, M-estimators are not very robust in the sense of the breakdown point (8), which is a statistical measure of sensitivity against noise or outliers in the data. Robust correlation coefficient is exploited also in Kalina (11) and can be recommended also to other image analysis in forensic applications (30). The following robust correlation coefficients are used in our numerical study to locate the facial symmetry in the images:

- r_{LWS} with weights determined by the LWS regression with linearly decreasing weights or radial weights;
- r_{LTS} after removing outliers detected by the LTS regression with $h = 0.95n$ or $h = 0.6n$ (here n is the number of pixels); and
- M-estimator correlation coefficient based on Huber's M-estimator.

Methods of Statistical Learning

We use the following standard classification methods (5) to localize the face by means of the axis of the best vertical symmetry in images of faces. They are popular especially in the image analysis of faces. Classification trees have a hierarchical structure comparing the variables with a threshold in each step. Support vector machines maximize the margin between classes. Neural networks include different methods containing numerous parameters, while their fitting is a combination of art and science requiring to choose numerous parameters or to avoid overfitting the data. The linear discriminant analysis is a standard statistical method discriminating between two groups of multivariate data assuming multivariate normal distribution with equal covariance structures; a nice application to criminology is presented by Ramsay and Silverman (20). Agglomerative hierarchical clustering classifies multivariate data into clusters (subsets, groups), starting with individual images as clusters and merging recursively selected pairs of clusters into a single cluster.

Results

Correlation Analysis

We apply the methods of the previous section to locate the axis of the best vertical symmetry in all 212 images of the database. Table 1 lists the percentages of correctly localized faces by this method for different widths of the rectangles (70, 60, 50, 40, or 30 pixels each). The width of the head in different images is usually between 80 and 85 pixels. We consider such results of the classification to be correct, when the estimated symmetry line intersects the nose and separates the nostrils. The results are compared for the Pearson product–moment correlation coefficient r, its weighted counterpart r_W with radial weights, Spearman's rank correlation coefficient r_S, and robust correlation coefficients.

TABLE 1—*Percentages of correct results obtained with the correlation analysis with different standard and robust measures of correlation. Two neighboring areas are compared, where each of them has a width of 70, 60, 50, 40, or 30 pixels. The weights or parameters of the methods are summarized in parentheses.*

Correlation coefficient	70	60	50	40	30
r	0.98	0.96	0.95	0.91	0.85
r_W (radial)	1.00	1.00	0.98	0.98	0.92
r_S	1.00	0.99	0.96	0.86	0.72
r_{LWS} (linearly decreasing)	0.99	0.98	0.99	0.94	0.89
r_{LWS} (radial)	0.98	0.98	1.00	0.99	0.95
r_{LTS} ($h = 0.95n$)	0.98	0.98	0.97	0.95	0.84
r_{LTS} ($h = 0.6n$)	0.99	1.00	1.00	0.99	0.57
r_M (Huber)	0.96	0.96	0.94	0.91	0.86

The Pearson product–moment correlation coefficient r has the tendency to find the best symmetry in a homogeneous area in the right or left part of the image. The most symmetric nonfaces detected by r in the image are located in the background (80% of cases), although it is contaminated by severe noise. Only less commonly, they contain the shifted face divided in an asymmetric way to the two rectangles (in 18%) or a shoulder with a smaller part of the face (2% of cases). Also, the position of the shoulders or possible asymmetry of hair contributes to incorrect classification results. The weighted correlation r_W with radial weights localizes the face correctly; also when the head is namely not perfectly straight, the (often asymmetric) shirt or hair can have a nuisance effect strongly influencing the correlation coefficient.

Further, we examine the capability of the correlation coefficient in the task of face recognition, that is, we compare a half of the face with halves of faces of other persons. We consider rectangles (Fig. 1) of width 60 pixels. For a particular face, its left half is considered together with the set of 212 right halves of faces of all persons in the database. It turns out that in 96% of cases, the largest value of r is obtained for the situation that both halves correspond to the face of the same person. The weighted correlation coefficient with radial weights gives 100% performance. This reveals that although the face is not symmetric, the intra-person variability is still smaller than the inter-person variability (the resemblance between two halves of the face is larger than that between one-half and a half of the face of another person).

Robustness of the Correlation Analysis

An important aspect of methods of image analysis is their robustness with respect to violations of the standardized conditions. To examine the local sensitivity of the classical and weighted Pearson product–moment correlation coefficient, we consider rectangles of width 60 and search for the axial symmetry in images with occlusion, illumination changes, different size, or shifted rectangles. This study illustrates theoretical finding of Böhringer et al. (23) for the context of (weighted) correlation analysis of faces, and the performance of both r and r_W is presented in Table 2.

To study the effect of a small occlusion by occluding every image in the database, we set the gray intensities in a rectangle of size 3×5 pixels to 1. Every face in the database is modified in

TABLE 2—*Performance in face localization for specific situations in modified images by occlusion, changes in illumination, size of the face, or shift in the image. The results reveal the robustness of the method to atypical situations.*

Modification of the Standard Images	Results of r	Results of r_W (Radial Weights)
Occlusion	1.00	1.00
Illumination (column-dependent noise)	1.00	1.00
Illumination (radial noise)	1.00	1.00
Size	1.00	1.00

this way by placing the occlusion always on the same position to the bottom right corner of the mouth, below the midpoint of the mouth by 7 to 9 rows and on the right from the midpoint by 16 to 20 columns.

Further, we examine the effect of moderate illumination changes or size of the image. First, we add column-dependent noise to every face in the database so that the left rectangle is retained, and a constant is added to the gray intensities in the right columns so that the columns far to the right obtain a larger value. Next, we also consider radial noise, retaining again the left rectangle and adding noise to gray intensities with the variability directly proportional to the distance of every pixel from the midpoint of the face. Another study retains the image and increases or decreases the size of the rectangles by 10%. All the results are given in Table 2.

Finally, we examine the robustness of the method to a small shift by 1 pixel. We compute the weighted correlation coefficient between a particular rectangle with size 192×60 pixels and the same rectangle shifted by 1 pixel aside. Although the gray intensities in particular pixels may vary greatly, the value of r_W over the whole database lies in the interval [0.936; 0.991]. This is ensured by a large size of the rectangles, and the method can be declared robust to a small shift. A shift by half of a pixel yields even a larger reliability of the correlation. Here, the gray value in a particular pixel in the shifted image is computed as the mean of four neighboring pixel in the original image. The correlation coefficient between the rectangle and its shifted counterpart typically exceeds 0.99, which follows from the large size of the rectangles and also large autocorrelation in the images.

Symmetry and Rotation

Next, we examine the robustness of the method to rotation up to ±10 degrees. The modified rectangles (section Methods) of width 60 pixels (Fig. 2) obtained by intersecting the original rectangles with the circle ignore 8% of the total number of pixels in the rectangles. We use the modified rectangles rotated by three different angles (−10, 0, and 10 degrees) together with the weighted correlation coefficient with radial weights to locate the best axial symmetry in the images. We select such of the three possible rotations, which leads to the largest value of the weighted correlation coefficient. In all images, the highest level of symmetry turns out to be present in the face, which is localized correctly as the eyes and nostrils are correctly separated in 100% of images, even if the largest weighted correlation coefficient is attained for an incorrect rotation of the modified rectangles. Particularly, the largest value of the three weighted correlation coefficients was attained for the nonrotated modified rectangles in 90% of images. In the remaining 10% of cases, the rotation of the face is mis-specified, often because of a nonsymmetric position of the shoulders with respect to the face, but still the face is correctly localized.

FIG. 3—*Weights assigned to a particular face by the LWS-based correlation coefficient. Black pixels contribute more to the symmetry and obtain larger weights. The symmetry is violated in white pixels (hair, collar), which are down-weighted in the computation of the robust correlation coefficient.*

Robust Measures of Correlation

The robust correlation coefficients based on the LWS or LTS regressions are now applied to search for the area with the best symmetry in the images of the database.

The LWS-based correlation coefficient r_{LWS} with linearly decreasing or radial weights is computed using the weighted analogy of the approximative algorithm (29). The result with radial weights is even more robust than with linearly decreasing weights (31), because the radial weights assign smaller values to outlying pixels compared to the linearly decreasing weights. In any case, the LWS-based method outperforms other correlation coefficients.

The contribution of pairs of pixels to the value of r_{LWS} is influenced by the sizes of the weights. The weights determined by r_{LWS} for the face in Fig. 1 are shown in Fig. 3. The boundary of the hair or the collar obtains smaller weights, because the violate the symmetry. Therefore, the value of r_{LWS} is not intrigued by this local asymmetry and gives a more reliable measure of symmetry than Pearson correlation coefficient.

We compute the LTS-based correlation coefficient r_{LTS} choosing $h = 0.95n$ or $h = 0.6n$. The results for the width 50 or 60 pixels are 100% correct for $h = 0.6n$. Not only the face itself but also the background in a semicircle around the head contributes to the symmetry. The outliers are detected in the background that contains both noise and spatial artifacts. As the images were taken, the light was coming from the front, so a halo around the head is produced. The LTS-based correlation coefficient down-weights these important pixels and focuses on the very center of the face, while the whole face should in fact have larger weights. Too narrow rectangles not containing the background next to the head are not capable to locate the face correctly in some images. A small value of the trimming constant h does not bring a big difference from the least squares, while a larger trimming increases the values of the correlation coefficient closer to 1, not only when the rectangles are symmetric. Then, the separation of the face from nonfaces is not very reliable.

Good results of the robust correlation analysis can be explained by the presence of large homogeneous areas in the image. To quantitate this, Moran (32) proposed a 2D autocorrelation coefficient as a measure of autocorrelation comparing the intensities in particular pixels with their neighbors. The value of Moran's (32) coefficient I lies between 0.96 and 0.98 in every image of our database, which explains that the intensity in a pixel is very strongly similar to the intensity in a neighboring pixel. To measure the autocorrelation in a robust way, we define an analogy of I based on the LWS regression with linear weights allowing us to down-weight the noise in the images. This robust measure has the mean value 0.9995 in every image. The measure shows that violations of the homogeneousness of the images truly correspond to boundaries and edges between almost homogeneous areas.

Methods of Statistical Learning

For the purpose of this section, we selected the training database of 124 faces and 124 nonfaces and the validation database of 88 faces and 88 nonfaces. We use the classification rules of statistical learning to learn over the training database, and the performance is verified on the validation database. The images are used as pairs of vertical neighboring rectangles 192×30 pixels. The nonfaces are selected as those nonfaces that have the largest symmetry measured by Pearson correlation coefficient.

The task is to classify each part of the image into two groups (faces and nonfaces), and for each classification method, we calculate the percentage of correctly classified images (pairs of neighboring rectangles). We use standard methods of statistical learning implemented by the R software (or its additional packages), which is a free software environment popular not only for statistical computing (33) Classification results for various statistical learning methods are presented in Table 3.

TABLE 3—*Percentages of correct results in classification of faces over the training and validation databases using standard methods of statistical learning. Two neighboring areas in the images are compared, where each of them has the width of 30 pixels.*

Methods of Statistical Learning	Results over the Training Database	Results over the Validation Database
Support vector machines	1.00	1.00
Classification trees	0.98	0.95
Neural networks	0.93	0.89
Linear discriminant analysis (on robust principal components)	1.00	1.00
Hierarchical cluster analysis	0.53	(not possible)

Support vector machines are trained using the function *svm* in library *e1071*. While the performance is 100% correct over the training database, the algorithm uses 142 support vectors, which means that the classification rule is based on 142 images. These are faces similar to nonfaces or nonfaces similar to faces. The method aims at distinguishing between two variability sources, namely variability between the two clusters and variability among images within clusters, and therefore, it requires a large computational complexity relatively to the size of the database. Only this high complexity of the classification rule allows us to obtain reliable results.

Classification trees in R software package (function *tree* in library *tree*) work only for smaller problems and collapse with our high-dimensional data. We performed the computation in Matlab (The MathWorks, Natick, MA). The resulting tree is based on 10 pixels of the total number of $192 \times 30 = 5760$ pixels. The resulting tree classifies correctly 98% of images. In the remaining six cases, the face is classified to be a nonface. The tree considers just gray intensities in single pixels, studying the difference between two groups only on a small number of individual pixels. The fitted tree relies too strongly on specific properties of the training set, and we find it a controversial classification rule. Moreover, they are rather unstable to small changes in the data.

Neural networks do not give persuasive results. The multilayer perceptron and radial basis function networks (library *neural*) do not converge to any result. Kohonen self-organizing map (library *kohonen*) yields two mixed groups as output, where one contains 95% of all faces and 9% of all nonfaces and the other contains the remaining 5% of the faces and 91% of all nonfaces. Also perfectly symmetric faces are classified incorrectly.

We compute the linear discriminant analysis by the function *lda* in library *MASS*. While the sizes of the images are too large to perform this classification on original images, we were able to compute robust principal components by the projection pursuit algorithm (34) and apply the linear discriminant analysis on five main components.

Agglomerative hierarchical clustering is computed using the function *hclust* in library *cluster*. We use the average linkage method with the Euclidean distance measure. The method yields very poor results in classifying our 248 images. One of the resulting two clusters contains eight nonfaces, and the other contains 240 remaining images (116 nonfaces and all 124 faces). The explanation is the severe asymmetry of the worst nonfaces, which include very dark hair or a very dark pullover. This method learns over the training database without the possibility to be applied to a validation database.

Discussion

The aim of this paper was the localization of faces in images by means of localizing the axis of the best vertical symmetry in images. In this discussion, we also aim at displaying possible areas of application in forensic anthropology. Using the database of 212 images of faces, two neighboring rectangles in the image are considered while comparing each pixel from one rectangle with a corresponding pixel from the other rectangle. While the Pearson product–moment correlation coefficient does not yield reliable results, it is outperformed by its weighted counterpart with radial weights. Nevertheless, both coefficients are sensitive to outliers, which have a high influence on the results. We prefer to use the newly proposed robust correlation coefficient based on the LWS regression, which yield reliable results over a wide range of situations.

The method of this paper can be applied in a combination with other methods for the tasks of face detection and face recognition. The robust approach can be applied to robustify existing approaches, which do not include a procedure for outlier detection. The most common face analysis methods include feature extraction or template matching, which need to handle symmetry; detecting symmetry as one of the features may be a reasonable approach. Nevertheless, typical solutions for face detection and recognition are based on a combination of several methods. This removes the disadvantages of individual approaches, which are manifested in an insufficient or false detection. Further symmetry detection methods can also help tracking people in a video sequence.

Robust statistical methods are very suitable for analyzing data in a robust way, without special assumptions. In our situation, they are resistant with respect to noise in the image, occlusion in the face, or to asymmetric hair style. They are able to detect the rather delicate asymmetric illumination in the background, while standard methods are masked by outliers and intrigued by the asymmetry in the background. The hair, shoulders, or collar happens to be asymmetric in the images and violates the assumption of symmetry of the image, although the face itself is relatively symmetric. The robust correlation coefficient assigns smaller weights to these pixels. Also, a head possibly rotated with respect to the shoulders does not disqualify the robust correlation coefficient, which is delicate enough to down-weight the pixels of the shoulders.

Both r_{LWS} and r_{LTS} outperform r and r_{W} with radial weights. The LWS method assigns weights to important pixels which contribute to the separation between a face and a nonface. For wider rectangles, these pixels with larger weights are located in the middle and at the top of the rectangles to pick up the face and the homogeneous background, possibly with a circle around the head caused by the frontal illumination. For narrower rectangles, the larger weights obtained with the LWS underline the middle part, the very top and the very bottom. The outliers detected by r_{LTS} typically appear in the hair or background.

The choice of the width of the rectangles is examined, and width between 70 and 50 pixels for each rectangle seems to be a reasonable choice for this particular data analysis task. To summarize the results, there is no uniformly optimal method for each width of the rectangles, but rather, different methods can be recommended for different widths. However, the paper offers more general conclusions regarding the choice of the classification methods and robustness aspects. The LWS-based correlation coefficient is a new method in this context, and the paper brings arguments in favor of this robust method and convinces that robust image analysis has a strong potential to obtain robust results in anthropology.

Standard classification methods are applied on a data set with 212 faces and 212 nonfaces divided into a training set and a validation set. Some of them ignore the typical appearance of a face, while other methods fail to be computed for our too large images. The methods based on machine learning must fail because they are trying to learn all possible appearances of nonfaces, which is a too heterogeneous set compared to the set of all faces. Successful results are obtained with the linear discriminant analysis (after computing the robust principal component analysis) and support vector machines. The verification is not needed for the correlation analysis, which does not learn any parameters on the training set. Moreover, the work with raw images itself uproots the usual prejudices about the necessity of reducing the dimension of the images.

As a future forensic research, we plan to analyze 3D image information using the superprojection of a skull to a 2D image allowing the person identification. Such standard search of connections between landmarks (prominent points) in the skull and landmarks in the face can be improved by applying robust methods, while robustness aspects of standard methods should be also examined. A possible asymmetry is an important feature of a skull, which determines also an asymmetry of the face. The future work in forensic anthropology can be inspired by successful results with the LWS in 2D images in this paper or robust image analysis in molecular genetic context (11).

References

1. Iscan MY, Helmer RP. Forensic analysis of the skull: craniofacial analysis, reconstruction, and identification. New York, NY: Wiley-Liss, 1993.
2. Pickering RB, Bachman DC. The use of forensic anthropology. Boca Raton, FL: CRC Press, 1997.
3. Rao CR. Linear statistical inference and its applications. New York, NY: Wiley, 1973.
4. Van Vark GN, Howells WW, editors. Multivariate statistical methods in physical anthropology. A review of recent advances and current developments. Dordrecht, The Netherlands: Reidel Publishing Company, 1984.
5. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. New York, NY: Springer, 2001.
6. Yang MH, Kriegman DJ, Ahuja N. Detecting faces in images: a survey. IEEE Trans Pattern Anal Machine Intel 2002;24(1):34–58.
7. Viola P, Jones MJ. Robust real-time face detection. Int J Comp Vision 2004;57:137–54.
8. Rousseeuw PJ, Leroy AM. Robust regression and outlier detection. New York, NY: Wiley, 1987.
9. Stigler SM. The changing history of robustness. Am Statist 2010;64(4): 277–81.
10. Heritier S, Cantoni E, Victoria-Feser MP, Copt S. Robust methods in biostatistics. New York, NY: Wiley, 2009.
11. Kalina J. Robust image analysis in the evaluation of gene expression studies. ERCIM News Eur Res Consort Inf Math 2010;82:52.
12. Kalina J. Robust image analysis of faces for genetic applications. Eur J Biomed Inf 2010;6(2):6–13.
13. Mallett XDG, Dryden I, Bruegge RV, Evison M. An exploration of sample representativeness in anthropometric facial comparison. J Forensic Sci 2010;55(4):1025–31.
14. Zabrodsky H, Peleg S, Avnir D. Symmetry as a continuous feature. IEEE Trans PAMI 1995;17(12):1154–66.
15. Chunyu L, Changshui Z, Fang W, Pingfan Y. Principle component analysis based symmetry detection. Acta Electronica Sinica 1999;27(5):25–8.
16. Enquist M, Arak A. Symmetry, beauty and evolution. Nature 2002;372: 169–72.
17. Wiskott L, Fellous JM, Krüger N, von der Malsburg C. Face recognition by elastic bunch graph matching. IEEE Trans Pattern Anal Machine Intel 1997;19(7):775–9.
18. Würtz RP. Object recognition robust under translations, deformations, and changes in background. IEEE Trans Pattern Anal Machine Intel 1997;19(7):769–75.
19. Kalina J. Locating the mouth using weighted templates. J Appl Math Stat Inf 2007;3(1):111–25.
20. Ramsay JO, Silverman BW. Functional data analysis, 2nd edn. New York, NY: Springer, 2006.
21. Zelditch M, Swiderski D, Sheets DH, Fink W. Geometric morphometrics for biologists. San Diego, CA: Elsevier Academic Press, 2004.
22. Bronstein AM, Bronstein MM, Kimmel R. Three-dimensional face recognition. Int J Comp Vision 2005;64(1):5–30.
23. Böhringer S, Vollmar T, Tasse C, Würtz RP, Gillessen-Kaesbach G, Horsthemke B, et al. Syndrome identification based on 2D analysis software. Eur J Hum Genet 2006;14:1082–9.
24. Lucy D. Introduction to statistics for forensic scientists. Chichester, U.K.: Wiley, 2005.
25. Pelin C, Zagyapan R, Yazici C, Kürkcüoglu A. Body height estimation from head and face dimensions: a different method. J Forensic Sci 2010;55(5):1326–30.
26. Shevlyakov GL, Vilchevski NO. Robustness in data analysis: criteria and methods. Utrecht, The Netherlands: VSP, 2002.
27. Kalina J. Locating landmarks using templates. Nonparametrics and robustness in modern statistical inference and time series analysis: a

Festschrift in honor of Professor Jana Jureckova. IMS Collections 2010;7:113–22.
28. Víšek JÁ. The least weighted squares I. Bull Czech Econometric Soc 2001;9(15):1–28.
29. Rousseeuw PJ, Van Driessen K. A fast algorithm for the minimum covariance determinant estimator. Technometrics 1999;41(3):212–23.
30. Rudin N, Inman K. An introduction to forensic DNA analysis, 2nd edn. Boca Raton, FL: CRC Press, 2002.
31. Čížek P. Efficient robust estimation of time-series regression models. Appl Math 2008;53(3):267–79.
32. Moran PAP. Notes on continuous stochastic phenomena. Biometrika 1950;37:243–51.
33. Ihaka R, Gentleman R. R: a language for data analysis and graphics. J Comput Graph Stat 1996;5:299–314.
34. Croux C, Filzmoser P, Oliveira MR. Algorithms for projection-pursuit robust principal component analysis. Chemometr Intell Lab 2007;87: 218–25.

Additional information and reprint requests:
Jan Kalina, Ph.D.
Center of Biomedical Informatics
Institute of Computer Sciences AS CR
Pod Vodárenskou věží 2
182 07 Praha 8
Czech Republic
E-mail: kalina@euromise.cz