IMS Collections
Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečková
Vol. 7 (2010) 113–122
© Institute of Mathematical Statistics, 2010
DOI: 10.1214/10-IMSCOLL712

Locating landmarks using templates*

Jan Kalina

Charles University in Prague and Academy of Sciences of the Czech Republic

Abstract: This paper examines different approaches to classification and discrimination applied to two-dimensional (2D) grey-scale images of faces. The database containing 212 standardized images is divided to a training set and a validation set. The aim is the automatic localization of the mouth. We focus on template matching and compare the results with standard classification methods. We discuss the choice of a suitable template and inspect its robustness aspects.

While methods of image analysis are well-established, there exists a popular belief that statistical methods cannot handle this task. We ascertain that simple methods are successful even without a prior reduction of dimension and feature extraction. Template matching and linear discriminant analysis turn out to give very reliable results.

1. Introduction

The aim of this paper is to locate landmarks in two-dimensional (2D) grey-scale images of faces, to examine some aspects of template matching including the construction of templates and robustness aspects, and to compare different methods for locating landmarks. In contrary to standard approaches, we want to examine methods applied to raw data, without a prior reduction of dimension and feature extraction. There exists a popular belief that statistical methods cannot handle this task. We refer to [13] giving a survey of 181 recent articles on face detection and face recognition, which is still not an exhaustive survey but rather a study of selected remarkable specific approaches. Existing methods of image analysis are complicated combinations of ad hoc methods of mathematics, statistics and informatics as well as heuristic ideas which are tailor-made to suit the particular data and the particular task. These black boxes are far too complex to implement for users of the methods in all areas of applications. We point out that these reliable methods are based on extremely simple features, albeit organized in a cascade (see [10]), and furthermore simple templates are used also in complicated situations, for example in the spaces with the reduced dimension (see [9]). Our aim is also to compare template matching with methods of multivariate statistics; these turn out to yield successful results for standardized images. Reduction of dimension becomes unnecessary when very fast computers are available to analyze raw data and template matching has a clear interpretation and can be implemented routinely.

Charles University in Prague, KPMS MFF, Sokolovská 83, 186 75 Praha 8 and Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 2, 182 07 Praha 8, Czech Republic. e-mail: kalina@euromise.cz
url: http://www.euromise.org/homepage/people/kalina.html

*This work is supported by Jaroslav Hájek Center for Theoretical and Applied Statistics, project LC 06024 of the Ministry of Education, Youth and Sports of the Czech Republic.

AMS 2000 subject classifications: Primary 62H30; secondary 62H35.

Keywords and phrases: classification and discrimination, high-dimensional data.

Possible applications of detecting objects in images include also face detection for forensic anthropology, secret service or military applications, but also other applications on images with other objects than faces (weather prediction from satellite images, automatic robot vision) or even detection of events in financial time series (fraud detection).

We work with the database of images from the Institute of Human Genetics, University Clinic in Essen, Germany, which was acquired as a part of grants BO 1955/2-1 and WU 314/2-1 of the German Research Council (DFG). It contains 212 grey-scale images of faces of size 192 × 256 pixels. We divide them to a training database of 124 images and a validation database with 88 images. A grey value in the interval [0,1] corresponds to each pixel, where low values are black and large values white. The images were taken under standardized conditions always with the person sitting straight in front of the camera looking in it. While the size of the head can differ only slightly, the heads are often rotated by a small angle and the eyes are not in a perfectly horizontal position in such images. For example there are no images with closed eyes, hair over the face covering the eyes or other nuisance effects. The database does not include images with a three-dimensional rotation (a different pose).

The Institute of Human Genetics is working on interesting problems in the genetic research using images of faces. The ambitions of the research are to classify automatically genetic syndromes from a picture of a 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. Some of the results are described in the papers by [12], [9] and [1].

All such procedures require as the first step the localization of landmarks, although this is not their primary aim. The landmarks are defined as points of correspondence (exactly defined biologically or geometrically) on each object that matches between and within populations (see [2] or [3]). Examples of landmarks include the soft tissue points located on inner and outer commisure of each eye fissure, the points located at each labial commisure, the midpoints of the vermilion line of the upper and lower lip (see [4]).

The team of genetics researchers uses two approaches to locate 40 landmarks in each face as follows [1]. One possibility is the manual identification, carefully performed by an anthropologist trained in this field. Another approach used at the institute is a semi-automatic procedure based on [12]. This starts with a two-dimensional wavelet transformation of the images and uses templates in the space of the wavelet coefficients. However it turns out to be very sensitive to slight rotations of the face. This is the motivation for our study of template matching and its robustness.

Chapter 2 is devoted to template matching applied to locating the mouth in images of the training database. We study robustness to local modifications or different lighting conditions. Chapter 3 compares different methods of classification analysis for the same task.

2. Locating the mouth using template matching

We describe our construction of templates and apply them with the aim to localize the mouth in the training database with 124 images of faces. Template matching is a tailor made method for object detection in grey-scale images using an ideal object with the ideal shape in the typical form, particularly applicable to locating

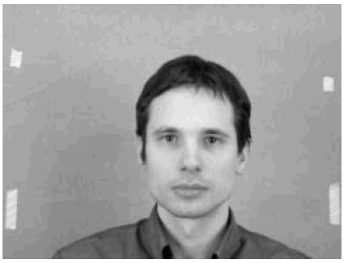

FIG 1. *An image from the database. Every image is a matrix of* 192 × 256 *pixels.*

faces or their landmarks in a single image. [13] gives a list of references on template matching.

The template is placed on every possible position in the image and the similarity is measured between the template and each part of the image, namely the grey value of each pixel of the template is compared with the grey value of the corresponding pixel of the image. The standard solution is to compute the Pearson product-moment correlation coefficient r to compare all grey values of the image ignoring the coordinates of the pixels. In the following text we consider the Pearson product-moment correlation coefficient r (shortly called correlation coefficient) and the weighted Pearson product-moment correlation coefficient (shortly called weighted correlation coefficient).

2.1. Construction of templates

In the references [13] or [10] we have found no instructions on a sophisticated construction of templates. We construct the set of mouth templates in the following way. Starting with a particular mouth with a typical appearance, we compute the Pearson product-moment correlation coefficient between this mouth of size 27 × 41 pixels and every possible rectangular area of the size 27 × 41 pixels of every image of the training set. In 16 images the maximal correlation coefficient between the template and the image exceeds 0.85 and this largest correlation coefficient is obtained each time in the mouth. The symmetrized average of the grey values of the 16 mouths is used as the first template. The process of averaging removes individual characteristics and retains typical properties of objects.

The procedure was then repeated with such initial mouth, which did not have the correlation coefficient with any of the previous mouth templates above 0.80. Some of the initial templates are rectangles including just the mouth itself and the nearest neighbourhood, others go as far downwards as to the chin. Nonstandard mouths are also included as initial templates, for example not horizontal, open with visible teeth, smiling or luminous lips after using lipstick. Therefore we subjectively select different sizes of the templates.

Altogether a set of *13 mouth templates* of different sizes was constructed. All the 13 templates together lead to correct locating the mouth in every of 124 examined images, when the correlation coefficient is used as the measure of association between the template and the image.

Based on these templates we created a new set of templates. We selected one particular template and averaged such mouths, which have the correlation coefficient with it over 0.80. The symmetrized mean becomes one of the new templates. Then we selected another of the previous templates, symmetrized it and performed

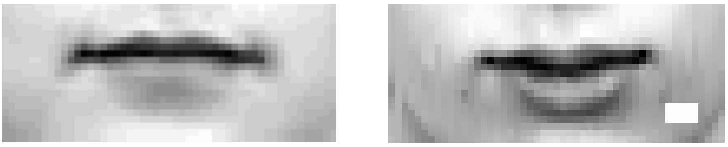

FIG 2. *Left: one of the templates for the mouth. Right: a mouth with a plaster.*

the same procedure. The selection of templates from the set of 13 templates was subjective and we have tried to select templates, which would be very different from those selected in previous steps. When the number of these new templates reached 7, it was possible to locate all the mouths in the whole database. Therefore our final set includes *7 mouth templates* with different sizes, namely two templates with a beard and five without it.

One of the templates is shown in Figure 2 (left). This template has the size 21 × 51 pixels. It locates the mouth in 99 % images of the training database, when using the correlation coefficient r as the measure of similarity between the template and the image. It is also the best template in the following sense. In a particular image the separation between the mouth and all non-mouths can be measured in the form

$$\frac{\max\{r(\text{template, mouth});\ \text{all positions of the mouth}\}}{\max\{r(\text{template, non-mouth});\ \text{all non-mouths}\}}. \tag{2.1}$$

The worst separation (2.1) over all the 124 images is a measure of the quality of a template. The best such result is obtained for the non-bearded template in Figure 2 (left).

2.2. Results of the template matching

The references on image analysis (for example [6] or [7]) describe the Pearson product-moment correlation coefficient as the standard and only recommendable measure of similarity between the template and the image. The importance of the lips or the central area of the template can be underlined properly if the weighted Pearson product-moment correlation coefficient

$$r_w(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^{n} w_i(x_i - \bar{x}_W)(y_i - \bar{y}_W)}{\sqrt{\sum_{i=1}^{n}[w_i(x_i - \bar{x}_W)^2] \sum_{j=1}^{n}[w_j(y_j - \bar{y}_W)^2]}}. \tag{2.2}$$

is used with radial weights $\mathbf{w}^R$. Let both the template and the weights be matrices of size $n_1 \times n_2$ pixels. The idea is to define the radial weight w_{ij}^R of a pixel with coordinates $[i, j]$ inversely proportional to its distance from the midpoint $[i_0, j_0]$. Formally let us firstly define

$$w_{ij}^* = \frac{1}{\sqrt{(i - i_0)^2 + (j - j_0)^2}}. \tag{2.3}$$

If n_1 and n_2 are odd numbers, then $w_{i_0 j_0}^*$ is not defined and we define additionally $w_{i_0 j_0}^* = 1$. The radial weights $\mathbf{w}^R$ are defined as

$$w_{ij}^R = \frac{w_{ij}^*}{\sum_{k=1}^{n_1} \sum_{l=1}^{n_2} w_{k\ell}^*}, \quad i = 1, \ldots, n_1, \quad j = 1, \ldots, n_2. \tag{2.4}$$

TABLE 1

Percentages of images with the correctly located mouth using different templates. Comparison of the Pearson product-moment correlation coefficient, weighted Pearson product-moment correlation coefficient with radial weights and Spearman's rank correlation coefficient. The templates have different sizes.

Template with description	r	r_w	r_S	Size of the template
All 7 templates	1.00	1.00	0.94	
1. Non-bearded	0.99	0.99	0.83	21 × 51
2. Non-bearded	0.93	0.94	0.80	27 × 41
3. Non-bearded	0.94	0.91	0.82	21 × 41
4. Non-bearded	0.92	0.69	0.83	21 × 41
5. Non-bearded	0.95	0.96	0.60	26 × 41
6. Bearded	0.91	1.00	0.50	26 × 56
7. Bearded	0.62	0.78	0.43	29 × 56

Weighted correlation coefficient with equal weights corresponds to classical Pearson correlation coefficient without weighting.

Now we examine the performance of particular mouth templates in locating the mouth over the training set of 124 images of the database using the classical correlation coefficient r, weighted correlation coefficient r_w with radial weights and Spearman's rank correlation r_S as the similarity measures between the template and the image. The results are summarized in Table 1 as percentages of correctly localized mouths over the database with 124 images. The top of the table gives results with 7 templates from Section 2.1. Further, the table contains results of locating the mouth with just one template at the time.

The template in Figure 2 (left) with radial weights yields the best results over non-bearded templates in terms of the separation (2.1), where the correlation coefficient r is replaced by weighted correlation r_w with radial weights. The improvement in locating the mouth with radial weights compared to equal weights is remarkable in images with a different size or rotation of the face. Other attempts to define templates or other combinations of several templates were less successful. Spearman's rank correlation coefficient r_S has a low performance in locating the mouth.

The validation set contains 88 images taken under the same conditions as the training set. The set of 7 templates locate the mouth correctly in 100 % of images of the validation set with both equal and radial weights. The non-bearded template has the performance 100 % also for both equal and radial weights for the weighted correlation coefficient.

Robust modifications of the correlation coefficient in the context of image analysis of templates were inspected by [8]; the best performance was obtained with a weighted Pearson product-moment correlation coefficient with weights determined by the least weighted squares regression [11]. The next section 2.3 studies robustness aspects of template matching. Although the literature is void of discussions about robustness aspects in the image analysis context, we will see in Section 3 that also some non-robust classification methods perform very successfully in comparison with template matching with the weighted Pearson product-moment correlation coefficient r.

2.3. Robustness of the results

An important aspect of the methods for locating objects in images is their robustness with respect to violations of the standardized conditions. This study goes beyond the study of sensitivity to asymmetry of the image by [8].

To examine the local sensitivity of the classical and weighted correlation coefficient, we study the effect of a small plaster similarly with Figure 2 (right). Grey values in a rectangle of size 3×5 pixels are set to 1. Every mouth in the database is modified in this way placing the plaster 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.

We use the set of 7 templates and different weights to search for the mouth in such modified images. Equal weights localize the mouth correctly in 88 % out of the 124 images. Radial weights $\mathbf{w}^R$ are robust to such plaster and locates the mouth correctly in 100 % of images.

Now we study theoretical aspects of the robustness of the template matching.

We need the notation $\bar{t}_w$ and $\bar{x}_w$ for the weighted means of the template $\mathbf{t}$ and an image (mouth or non-mouth) $\mathbf{x}$ respectively, for example $\bar{x}_w = \sum_{i=1}^{n} w_i x_i$. The weighted variance $S_w^2(\mathbf{x}; \mathbf{w})$ of $\mathbf{x}$ with weights $\mathbf{w}$ is defined by

$$S_w^2(\mathbf{x}) = \sum_{i=1}^{n} w_i (x_i - \bar{x}_w)^2 \tag{2.5}$$

and an analogous notation $S_w^2(\mathbf{t})$ is used for the weighted variance of grey values of the template $\mathbf{t}$ with weights $\mathbf{w}$. The weighted covariance $S_w(\mathbf{x}, \mathbf{t})$ between $\mathbf{x}$ and $\mathbf{t}$ equals

$$S_w(\mathbf{x}, \mathbf{t}) = \sum_{i=1}^{n} w_i (x_i - \bar{x}_w)(t_i - \bar{t}_w). \tag{2.6}$$

The following practical theorem studies the robustness of $r_w(\mathbf{x}, \mathbf{t})$ with respect to an asymmetric modification of the image, for example a part of the image can have a different illumination, in the matrix notation $\mathbf{x}^* = (x_{ij}^*)_{i,j}$ with $x_{ij}^* = x_{ij}$ for $j < j_0$ and $x_{ij}^* = x_{ij} + \varepsilon$ for $j \geq j_0$ for some j_0 for every i.

We study how adding a constant ε to a part of the image effects the weighted correlation coefficient of such image with the original template and original weights. Here the notation $\mathbf{x}+\varepsilon$ with $\mathbf{x} = (x_1, \dots, x_n)^T$ stands for $(x_1+\varepsilon, x_2+\varepsilon, \dots, x_n+\varepsilon)^T$. We also use the following notation. The image $\mathbf{x}$ is divided to two parts and $\sum_I$ or $\sum_{II}$ denote the sum over the pixels of the first or second part, respectively. Dividing the image $\mathbf{x}$ to three parts, the sums over particular parts are denoted by $\sum_I$, $\sum_{II}$ and $\sum_{III}$.

Theorem 2.1. *Let* $\mathbf{t}$ *denote the template,* $\mathbf{x}$ *the image and* $\mathbf{w}$ *the weights. We assume these matrices to have the same size. Then the following formulas are true.*

1. *For* $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)^T$ *and* $\mathbf{x}^* = (\mathbf{x}_1, \mathbf{x}_2 + \varepsilon)^T$, $r_w(\mathbf{x}^*, \mathbf{t}) =$

$$= \frac{S_w(\mathbf{x}, \mathbf{t}) + \varepsilon \sum_{II} w_i t_i - \varepsilon v_2 \bar{t}_w}{S_w(\mathbf{t}) \sqrt{S_w^2(\mathbf{x}) + v_2(1 - v_2)\varepsilon^2 + 2\varepsilon(2v_2 - 1)(\sum_{II} w_i x_i - v_2 \bar{x}_w)}}, \tag{2.7}$$

where $v_2 = \sum_{II} w_i$.

2. *For* $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)^T$ *and* $\mathbf{x}^* = (\mathbf{x}_1 + \varepsilon, \mathbf{x}_2 - \varepsilon)^T$, $r_w(\mathbf{x}^*, \mathbf{t}) =$

$$= \frac{S_w(\mathbf{x}, \mathbf{t}) + \varepsilon(\sum_I w_i t_i - \sum_{II} w_i t_i) - \varepsilon v \bar{t}_w}{S_w(\mathbf{t}) \sqrt{S_w^2(\mathbf{x}) + \varepsilon^2(1 - v)^2 - 2\varepsilon v \bar{x}_w + 2\varepsilon(\sum_I w_i x_i - \sum_{II} w_i x_i)}}, \tag{2.8}$$

where $v = \sum_I w_i - \sum_{II} w_i$.

3. *For* $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)^T$ *and* $\mathbf{x}^* = (\mathbf{x}_1, \mathbf{x}_2 + \varepsilon, \mathbf{x}_3 - \varepsilon)^T$, $r_w(\mathbf{x}^*, \mathbf{t}) =$

$$= \frac{S_w(\mathbf{x}, \mathbf{t}) + \varepsilon \bar{t}_w(w_3 - w_2) + \varepsilon(\sum_{II} w_i t_i - \sum_{III} w_i t_i)}{S_w(\mathbf{t})\sqrt{S_w^2(\mathbf{x}) + t + \varepsilon^2\left[w_2 + w_3 - (w_2 + w_3)^2\right]}}, \tag{2.9}$$

where $w_2 = \sum_{II} w_i$ *and* $w_3 = \sum_{III} w_i$ *and*

$$t = \varepsilon\left[\sum_{II} w_i x_i - \sum_{III} w_i x_i + \bar{x}_w(w_3 - w_2)\right]. \tag{2.10}$$

4. *Let* $\boldsymbol{\varepsilon}$ *denote a matrix of the same size as* $\mathbf{x}$ *containing constants* $(\varepsilon_{ij})_{ij}$. *Then*

$$r_w(\mathbf{x} + \boldsymbol{\varepsilon}, \mathbf{t}) = \frac{S_w(\mathbf{x}, \mathbf{t}) + S_w(\mathbf{t}, \boldsymbol{\varepsilon})}{S_w(\mathbf{t})\sqrt{S_w^2(\mathbf{x}) + S_w^2(\boldsymbol{\varepsilon}) + 2S_w(\mathbf{x}, \boldsymbol{\varepsilon})}}. \tag{2.11}$$

For the special case with the symmetric mouth, symmetric template and symmetric weights we can formulate the following corollary of Theorem 2.1, where we can express $r_w^*(\mathbf{x}, \mathbf{t})$ as a function of $r_w(\mathbf{x}, \mathbf{t})$. In this special case the weighted correlation coefficient $r_w^*(\mathbf{x}, \mathbf{t})$ always decreases compared to $r_w(\mathbf{x}, \mathbf{t})$, and the theorem expresses the level of the decrease and thus proves the template matching to be reasonably robust to small modifications of the template.

Theorem 2.2. *Let us consider a particular template* $\mathbf{t}$, *image* $\mathbf{x}$ *and weights* $\mathbf{w}$. *We assume that all these matrices have the same size and are symmetric along the vertical axis. Then the following formulas are true.*

1. *Let* $\mathbf{t}$, $\mathbf{x}$ *and* $\mathbf{w}$ *have an even number of columns. Let us perform the following modification* $\mathbf{x}^*$ *of the mouth* $\mathbf{x}$. *Grey values on one side of the axis are equal to those of* $\mathbf{x}$ *and the remaining are increased by* ε *compared to those from* $\mathbf{x}$. *Then the weighted correlation coefficient between the template and the modified mouth* $\mathbf{x}^*$ *can be expressed by*

$$r_w(\mathbf{x}^*, \mathbf{t}) = r_w(\mathbf{x},\, \mathbf{t})\frac{S_w(\mathbf{x})}{\sqrt{S_w^2(\mathbf{x}) + \frac{\varepsilon^2}{4}}}. \tag{2.12}$$

2. *Let* $\mathbf{t}$, $\mathbf{x}$ *and* $\mathbf{w}$ *have an even number of columns. Let us perform the following modification* $\mathbf{x}^*$ *of the mouth* $\mathbf{x}$. *Grey values on one side of the axis are increased by* ε *and the remaining are decreased by* ε *compared to those from* $\mathbf{x}$. *Then the weighted correlation coefficient between the template and the modified mouth* $\mathbf{x}^*$ *can be expressed by*

$$r_w(\mathbf{x}^*, \mathbf{t}) = r_w(\mathbf{x}, \mathbf{t})\frac{S_w(\mathbf{x})}{\sqrt{S_w^2(\mathbf{x}) + \varepsilon^2}}. \tag{2.13}$$

3. *Let us perform the following modification* $\mathbf{x}^*$ *of the mouth* $\mathbf{x}$. *For a specific number* k *in* $\{0, 1, \ldots, j/2\}$, *grey values in columns* $1, \ldots, k$ *are increased by* ε *and in columns* $k - j + 1, \ldots, k$ *are decreased by* ε *compared to those from* $\mathbf{x}$. *The remaining grey values are equal to those in* $\mathbf{x}$. *Then the weighted correlation coefficient between the template and the modified mouth* $\mathbf{x}^*$ *can be expressed by*

$$r_w(\mathbf{x}^*, \mathbf{t}) = r_w(\mathbf{x}, \mathbf{t})\frac{S_w(\mathbf{x})}{\sqrt{S_w^2(\mathbf{x}) + 2v\varepsilon^2}}, \tag{2.14}$$

where $v = \sum_{i=1}^{n}\sum_{j=1}^{k} w_{ij}$.

Table 2

Percentages of correctly classified images using different classification methods implemented in R *software. The classification rule is learned over the training data set with 124 images and further applied to the validation set with 88 images. The template matching uses 7 templates with radial weights.*

	Results over the		
Classification method	training set	validation set	R library
Linear discriminant analysis	1.00	1.00	`neural`
Support vector machines	0.90	0.85	`e1071`
Hierarchical clustering	0.53	-	`cluster`
Classification tree	0.97	0.90	`tree`
Neural network – multilayer	1.00	1.00	`neural`
Neural network – Kohonen	0.98	0.96	`kohonen`
Template matching	1.00	1.00	-

3. Locating the mouth using classification methods

This section compares classification methods applied to locating the mouth in the original images. This has not been inspected in this context without the usual prior steps of dimension reduction and feature extraction because of a high computational complexity.

Locating the mouth in the whole images without a preliminary reduction of dimension is a task with an enormous computational complexity. Therefore we consider the mouth and only one non-mouth from every image of the training set with 124 images, always with the size 21×51 pixels; this is the size of the template in Figure 2 (left). We select such non-mouth which has the largest correlation coefficient with the template in Figure 2 (left). A shifted mouth was not considered to be a non-mouth, so the non-mouths are required be at least five pixels distant (in the Euclidean sense) from the mouth. All mouths and non-mouths are selected in such position that the correlation coefficient with the template in Figure 2 (left) is larger than the correlation coefficient between the template and the same image (mouth or non-mouth) shifted aside; this ensures the images to have centered in the same way, treating the fact that the midpoint of the template does not correspond to the midpoint of the lips.

Such training database for the next work contains 248 images (a group of 124 mouths and a group of 124 non-mouths) with the aim to classify these images to groups. We apply linear discriminant analysis, support vector machines, hierarchical clustering, classification trees and neural networks to this task. These methods were selected as standard for classification analysis (see [5]). We point out that the dimension of the data much larger than the number of data.

Now we discuss the results of particular methods summarized in Table 2, which describes the results of the classification over the training set with 248 images. The resulting classification rule was further used on the validation set to examine the reliability of the classification rules, which had been learned over the described training set. The validation set was created from the original validation database of 88 images in the same way again as a set containing the mouth and only one non-mouth from each image in the same way as before, so it contains 176 images (88 mouths and 88 non-mouths). We use additional libraries of the R software (`http://cran.r-project.org`) for the computation of standard classification methods; the libraries are listed in Table 2.

The linear discriminant analysis yielding 100 % correct results consists in computing the classification score and classifying based on the inner product of the image with the score. The classification yields correct results without error. Influential values of the score appear in the top corners. This corresponds to the intuition, because the top corners have the lowest variability in the images of both mouths and non-mouths.

Results of the support vector machines classifier with a radial basis kernel were not convincing, although the classification is based on 136 support vectors, which indicates the complexity of this classification problem. Such classification rule is based on 136 closest images to the nonlinear boundary between the group of mouths and the group of non-mouths.

The hierarchical clustering with the average linkage method with the Euclidean distance measure giving two clusters as the output yields poor output. One cluster contained 58 non-mouths and the other contained 190 remaining images, namely 66 non-mouths and all 124 mouths. The method is not able to classify correctly such worst non-mouths which visually resemble a mouth. While there is a much larger variability among the non-mouths than among mouths, the method perceives the mouths to be a large and rather heterogeneous group. Non-mouths very different from mouths are classified as non-mouths, while problematic non-mouths are classified as mouths. Hierarchical clustering is an agglomerative (bottom-up) method starting with individual objects as individual clusters and merges recursively a selected pair of clusters into a single cluster; therefore it does not allow to classify a new observation from the validation set.

The classification tree is based only on 6 pixels, which can be found outside lips. It relies too strongly on specific properties of the training set and can hardly be accepted as a practical classification rule.

For neural networks we use two different approaches. The multilayer perceptron networks with 4 neurons as an example of supervised methods yields 100 % correct results in classifying the images as mouths or non-mouths. Kohonen self-organizing maps are an example of unsupervised methods based on mapping the multivariate data down onto a two-dimensional grid, while its size is a selectable parameter. We were not able to find any value of this size, for which 100 % correct results would be obtained.

The validation set also contains one atypical face. This is an older lady with an unusually big mouth, which is at the same time affected by small rotation, nonsymmetry and a light grimace. Nevertheless the classifiers either localize the mouth correctly in this image, or they fail also in several other faces (Table 2).

4. Conclusions

The aim of this work was to study different methods for the automatic localization of the mouth in two-dimensional grey-scale images of faces. Standard approaches start with an initial transformation of the image, for example Procrustes superimposition or even principal components analysis used in the right circumstances. These reduce the dimension of the image, so that the ultimate analysis is done on shape and shape alone. However the templates applied to raw data have not been examined from the statistical point of view.

Chapter 2 of this papers describes our approach to the construction of templates. A set of 7 mouth templates is able to localize the mouth in all 124 images of the training database; here the weighted Pearson product-moment correlation coefficient was used with radial weights. It is presented theoretically how this weighted correlation coefficient varies for distorted images.

Chapter 3 presents an experiment comparing different classification methods. Classification trees are rather controversial for these data; they are based on a very small number of pixels. This instability could be solved by using large patches (e.g. patch mean) or some other features (e.g. Haar-like features) rather than pixel intensities. Neural networks represent a black box, for which we are not able to analyze the result in a transparent and explanatory way. Results of support vector machines (SVM) and hierarchical clustering were not satisfactory. The SVM depend on several parameters to be tuned to perform optimally; an inexperienced practitioner using default parameter settings would however not obtain successful results. Therefore we praise template matching, linear discriminant analysis and multilayer neural networks, which yielded correct results in 100 % of images of both the training and validation databases.

Non-robust methods turn out to be able to attain the best results, which is the case of the template matching and linear discriminant analysis. At the same time template matching and linear discriminant analysis allow for a nice and clear interpretation.

The author is thankful to two anonymous referees for valuable comments and tips for improving the paper.

References

[1] Böhringer, S., Vollmar, T., Tasse, C., Würtz, R. P., Gillessen-Kaesbach, G., Horsthemke, B., and Wieczorek, D. (2006). Syndrome identification based on 2D analysis software. *Eur. J. Human Genet.* **14** 1082–1089.

[2] Bookstein, F. L. (1991). *Morphometric tools for landmark data. Geometry and biology.* Cambridge University Press, Cambridge.

[3] Dryden, I. L. and Mardia, K. V. (1999). *Statistical shape analysis.* John Wiley, New York.

[4] Farkas, L. (1994). *Anthropometry of the head and face.* Raven Press, New York.

[5] Härdle, W. and Simar, L. (2003). *Applied multivariate statistical analysis.* Springer, Berlin.

[6] Jain, A. K. (1989): *Fundamentals of digital image processing.* Prentice-Hall, Englewood Cliffs.

[7] James, M. (1987): *Pattern recognition.* BSP Professional books, Oxford.

[8] Kalina, J. (2007). Locating the mouth using weighted templates. *Journal of Applied Mathematics, Statistics and Informatics* **3** 111–125.

[9] Loos, H. S., Wieczorek, D., Würtz, R. P., Malsburg von der, C., and Horsthemke, B. (2003). Computer-based recognition of dysmorphic faces. *Eur. J. Human Genet.* **11** 555–560.

[10] Viola P. and Jones M.J. (2004). Robust real-time face detection. *Int. Journal of Comp. Vision* **57** 137–154.

[11] Víšek, J. Á. (2001). Regression with high breakdown point. In J. Antoch, G. Dohnal (Eds.): *ROBUST 2000, Proceedings of the 11-th summer school JČMF, Nečtiny, September 11-15, 2000,* JČMF and Czech Statistical Society, Prague, 324–356.

[12] Wiskott, L., Fellous, J. M., Krüger, N., and Malsburg von der, C. (1997). Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Anal. Machine Intel.* **19** 775–779.

[13] Yang, M.-H., Kriegman, D. J., and Ahuja, N. (2002) Detecting faces in images: A survey. *IEEE Trans. Pattern Anal. and Machine Intel.* **24** 34–58.