

Robust Image Analysis of Faces for Genetic Applications

Jan Kalina¹

¹Centre of Biomedical Informatics, Institute of Computer Science AS CR, Prague, Czech Republic

Summary

This paper is devoted to automatic localization of objects (eyes, mouth) in two-dimensional (2D) grey scale images of faces. Motivated by a practical problem in human genetics, the output of the localization of objects in the given database of images is needed for further tasks in the genetic research. A robust filter is applied on the image to ensure denoising. Templates are used as the main method. The mouth and both eyes are localized jointly using the weighted Pearson product-moment correlation coefficient or its robust analogy based on robust regression methods. In the database with 212 images of faces the method allows to locate the mouth and eyes correctly in 100 % of cases. Also the robust correlation coefficient based on the least weighted squares regression localizes the mouth and both eyes in 100 % of images of the given database. Robustness aspects of the method are examined with respect to rotation, noise, occlusion and asymmetry in the image. The joint localization of the mouth and both eyes produces the method invariant to rotation of any degree. This work is tailor made for the given images with expected usage of the methods in genetic applications.

Keywords: object localization, template matching, eye or mouth detection, robust correlation analysis, image denoising

1. Introduction

The primary motivation for this work is the automatic location of landmarks in images of human faces motivated by genetic research. We are working with a database of images which come from 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 grey value images of the size 192 times 256 pixels, each image corresponding to a different person. The faces have about the

same size but are rotated in the plane by small angles. Our aim is to find a robust solution with respect to rotation, occlusion, noise in the image and asymmetry of the face, while allowing for a clear interpretation of the method.

Template matching is a tailor made method for object detection in grey scale images. A template is a model, a typical form, an ideal object. It 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. [1] gives references on template matching applied to face detection and face recognition. Nevertheless standard image analysis procedures are formed as a cascade of extremely simple classifiers [2]. On the other hand ignoring the usual procedures of dimension reduction and feature extraction can ensure a clear interpretation and allows theoretical robustness considerations [3].

[4] is an example of template matching applied to the data set of irises of 64 persons with the task to assign a new iris to the correct person from the training database. The method maximizes the mutual information as the similarity measure between the template and the red channel of the color image. [5] considers templates as elements of a tree-structured hierarchy. More sophisticated algorithms of image analysis combine ad hoc methods of mathematics, statistics and informatics [2] accenting high computational speed rather than convenient theoretical properties [3]. Recent papers on image analysis replace pixels by features or patches. Patches [6] are homogeneous areas of pixels and the patch-based approach analyzes the image as a set of individual patches. On the other hand features typically correspond to edges or objects with heterogeneity or

discontinuity. [7] was the first paper reducing the dimension by replacing pixel intensities by a feature set. [8] and [9] are recent works on human or face detection by extracting features with the aim to propose methods possibly robust to illumination changes, different pose or facial expressions. [10] studies variance matrices of features, because variance is robust to illumination changes.

We work with the database containing images which are matrices with the size 192*256 pixels. A grey value in the interval [0,1] corresponds to each pixel, where low values are black and large values white. Images are photographed under the same conditions, with the person sitting straight against the camera looking straight at it. The Institute of Human Genetics tried to have the images standardized as much as possible. For example there are no images with closed eyes, hair over the face covering the eyes or other nuisance effects. Still the faces in the images happen to be rotated by a small angle. The eyes are not in a perfectly horizontal position in such images. 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 [11]. 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 [12]. Images of patients can be classified to one of 10 groups according to a genetic malformation deforming the face. For different syndromes the success rate lies between 75 % and 80 %, which is considered remarkably successful.

Locating the landmarks is always the first step of all such procedures, however not the primary goal of the study. The landmarks are prominent parts of the face, for example the corners of the eyes and the mouth, the midpoint of the top and the bottom edges of the lips or significant points of the nostrils and eyebrows. The team of genetics researchers uses two approaches to locate 40 landmarks in each face. One possibility is the manual identification, which is carefully and accurately performed by an anthropologist trained in this field. As the second approach the Institute uses an automatic method, based on the algorithm [13], which will be now described.

The algorithm starts by manual location of the set of 40 landmarks in a training set of 83 images of faces. These landmarks are called fiducial points and they together are placed on all positions in the image as one large template retaining fixed distances between the landmarks. Two-dimensional Gabor wavelet transformations with different values of the two-dimensional scale parameter are applied on all the training images and also on a new image in which the landmarks are to be located. The jets (Gabor wavelet coefficients) in each landmark of the training image and the jets in the corresponding pixels of the new image are compared. We can understand the jet of each of the training images as a (multi-dimensional) template. The correlation coefficient between the vectors of wavelet coefficients (or only their magnitudes) is computed and their sum over all 40 landmarks is used as the similarity measure between the training image and the new image. However, such an approach turns out to be vulnerable to small rotations of the face.

The aim of our work is to search for the mouth and eyes in images of faces using templates. We propose an algorithm for localizing each of the eyes separately and also jointly both eyes and the mouth using templates. We inspect robustness properties of the described method, for example the robustness to noise or rotation of the image. The paper has the following structure. Chapter 2 describes an initial denoising of images obtained by applying a robust filter. Templates are used as the main method to localize the eyes

(Chapter 3) and jointly both eyes and the mouth (Chapter 4). Robustness aspects of the method are examined and robust analogies of the correlation coefficient are inspected in Chapter 5.

2. Methods

Our approach begins with image denoising. We describe an approach for locating the eyes in the image and then the joint localization of both eyes and the mouth. Finally we compare different methods of robust correlation analysis for the same task of localizing both eyes and the mouth jointly. We must admit that we lose the advantage of feature-based methods described in Chapter 1, which allow to work with different sizes of the images. Nevertheless our approach has a clear interpretation and we are able to obtain a rotation-invariant approach. The results are presented in Chapter 3.

2.1 Image denoising

Denoising, filtering or robustification is a transformation often used to remove noise from images. We summarize arguments in favor of image denoising from references and describe our method based on the least trimmed squares or least weighted squares estimators. The motivation for such procedure is to remove noise from images while retaining the facial features well recognizable. Another artefact in our database of images is the reflection of light bulbs at different positions in the eyes of persons. This nuisance effect caused by the method of photographing at the Institute of Human Genetics is also removed by the image denoising. We describe these very robust estimators for a linear regression context, because they are used also later in Chapter 2.4 to define a robust correlation coefficient.

[14] describes filters (two-dimensional operators) for denoising and prefers the trimmed mean and other L-estimators [15] to the median. [16] proposes an M-estimator correlation coefficient, gives theoretical arguments in favor of the combining the robustness and efficiency for Gaussian white noise and applies it to the image analysis of templates. Denoising is applicable also to molecular genetics images [17] or alternatively [18] based on robust statistics.

We proceed to the definition of the robust statistical estimators, which are used for the image denoising in our work. Let us consider the linear regression model in the form

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + e_i, i = 1, 2, \dots, n, \quad (1)$$

or in the matrix notation $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$. The least weighted squares (LWS) regression estimator proposed by [19] is one of robust regression methods with a high breakdown point. It requires the specification of the sizes of the non-negative weights w_1, w_2, \dots, w_n . However these are assigned to particular data observations only after a permutation, which is determined in an implicit way during the computation of the estimator. This permutation depends on residuals $u_1(\mathbf{b}), \dots, u_n(\mathbf{b})$ corresponding to a particular value $\mathbf{b} = (b_0, b_1, \dots, b_p)^T$ of the estimator of the vector parameter $\boldsymbol{\beta}$, where

$$u_i(\mathbf{b}) = Y_i - b_0 - b_1 x_{i1} - \dots - b_p x_{ip}, i = 1, 2, \dots, n. \quad (2)$$

Typically the weights w_1, w_2, \dots, w_n are chosen as a non-increasing sequence. Denoting arranged squares residuals as

$$u_{(1)}^2(\mathbf{b}) \leq u_{(2)}^2(\mathbf{b}) \leq \dots \leq u_{(n)}^2(\mathbf{b}), \quad (3)$$

the estimator of $\boldsymbol{\beta}$ is defined as

$$\hat{\mathbf{b}}_{LWS} = \underset{\mathbf{b}}{\operatorname{argmin}} \sum_{i=1}^n w_i u_{(i)}^2(\mathbf{b}).$$

A popular choice [20] is to use linearly decreasing weights $w_i = 1 - (i-1)/n, i = 1, \dots, n$; another possibility is a two-stage procedure of [21] for the computation of adaptive weights, allowing to determine also the sizes of the weights automatically.

The least weighted squares regression combines the robustness and efficiency [21]. A fast approximative algorithm for computing the LWS can be obtained as a weighted analogy of [22]. [23] proposed the least trimmed squares (LTS) regression, which is a special case of the LWS regression with weights equal to 1 or 0 only. It must be specified that exactly h of the data points will have the weight equal to 1.

We apply these robust estimators in the context of image denoising. Their advantage is the high breakdown point [15] ensuring high resistance against outliers. Instead of the linear regression, only the location model is relevant. The least median of squares (LMS) estimator [23] is equivalent to the mean of the shortest half of the data. The LTS estimator corresponds to the mean of such half of the data (or a group of some $h < n$ observations), which has the smallest variance. The LWS estimator is the weighted mean of the data with such permutation of the weights yielding the smallest weighted variance of the data [20].

For each pixel we take the grey values from its circular neighborhood and compute the least median of squares (LMS), least trimmed squares (LTS) or least weighted squares (LWS) estimator. The information about the coordinates is lost. Poor results are obtained with the median because it removes contrast and the resulting image is rather greyish. The performance of the LTS and LWS estimators is reliable with a small radius of the circular neighborhood, however not strongly influenced by the choice of h for the LTS or by the choice of weights for the LWS estimator.

Therefore we use the LTS filter on each image in the database. For each pixel we consider its four direct neighbors and compute the LTS estimator with $h = 3$, which is computed as the mean of the shortest triplet obtained from the values arranged in ascending order. This removal of possible extreme outliers removes also the light reflected in eyes to some extent. To examine the effect of denoising, we examined the residuals of this transformations. Large absolute values of the residuals indicate a larger local effect of denoising. This effect of the transformation is remarkably attained at edges in the image, such as between the hair or shirt and the background or at the boundary of the nostrils.

2.2 Locating eyes

We use template matching to search for the right and left eye in each image separately. It is popular to use the mean of eyes as the template for the eye (compare [24], [25]). Here we construct a set of 6 eye

templates for the right eye and their reflections in the axial symmetry are used for the other eye. The templates are obtained as the mean of real eyes of the same size of different persons; one of them is shown in Figure 1. The templates have different sizes between 26×28 pixels and 36×30 pixels. The Pearson product-moment correlation coefficient (further called simply correlation coefficient) or the weighted Pearson product-moment correlation coefficient r_w (further called weighted correlation coefficient) with suitable weights is used as the similarity measure between the template and the image.



Fig. 1. An eye template.

All of the templates contain eyebrows, which could possibly complicate the recognition. Nevertheless the area of eyebrows will be down-weighted by the

weighted correlation coefficient with radial weights. Figure 1 shows such an eye template, which yields the best performance in locating the right eye; in this case this template alone was used together with the weighted correlation coefficient with radial weights.

We compute the correlation coefficient between one of the images and the eye template from Figure 1. The correlation coefficient is shown in Figure 2, where black areas have a large value of the correlation coefficient. These areas include both eyes and also the mouth or parts of the hair. It happens for example that the right eye of a particular person has a larger correlation with one of the left eye templates than with any of the right eye templates. We use this phenomenon in the algorithm in the following way. All the twelve templates are placed on every possible position in the image. We do not distinguish between left and right eye templates. Firstly the area with the largest correlation coefficient over all templates is found. That is one candidate eye. Now the whole image without this candidate eye and its nearest neighborhood is considered. Again the area with the largest correlation with any of the eye templates is found. That is the other candidate eye.



Fig. 2. The correlation coefficient of a given image and the template for the right eye was computed. Areas of the image with a large correlation coefficient are shown black.

2.3 Locating eyes and mouth

To localize jointly the mouth and both eyes, we start with a region which is candidate to contain the mouth and search for the eyes in a certain region above the candidate mouth. We use 6 templates for the right eye, their reflections in the axial symmetry and 7 symmetric mouth templates. Each mouth template is used to find several areas, which have the weighted correlation coefficient with any of the mouth templates above a certain threshold. This can be different in different images. In any case we search for at least three such areas, which are not direct neighbors of each other. These areas are considered to be candidate mouths. First we select areas with the weighted correlation coefficient r_w above the value of 80 % of the maximal r_w , which was attained in the particular image. This boundary can be decreased if necessary.

Different mouth templates yield different candidate mouths. Then the eyes are searched for in such areas, which correspond to particular candidate mouths. These areas range from 20 to 55 pixels above candidate mouths and are considered to be candidate areas for eyes. We place each of the 6 eye templates and their reflections in the axial symmetry on every possible position in these candidate eyes and compute the weighted correlation coefficient between the template and the corresponding region of the same size. At the same time the area with the expected position of the eyes is divided to two parts along the virtual vertical line (Figure 2), which bisects the mouth. All eye templates and also their versions reflected in the axial symmetry are used in both parts of the candidate area for eyes. We search for the area with the largest weighted correlation coefficient with any of the templates in the left and right parts separately. This approach is repeated for several candidate mouths.

Moreover we use a condition on the mutual distance of both eyes in the limit between 25 and 42 pixels. These bounds are obtained empirically and real eyes fulfill the condition also in images with a slightly modified size or rotated by a small angle.

For a given candidate mouth we find a candidate area for the right and left eye. If

these do not fulfill the condition on the distance between eyes, we continue searching for such area, which has the largest weighted correlation coefficient with any eye template among all remaining areas in the candidate area for eyes. Such area with the midpoint in the candidate area is one candidate eye, while the other candidate eye has the largest weighted correlation coefficient with any of the templates and must have the midpoint in the opposite part of the candidate area. Again the condition on the eye distance is checked and if needed these steps are repeated, until two candidate eyes are found, which satisfy the condition. This method distinctively improves the results for rotated images (Chapter 3).

The basic idea is to add three weighted correlation coefficients corresponding to the mouth and both eyes. Let us start with one of pixels, which is the midpoint with a large weighted correlation coefficient with any mouth templates. Let us denote the largest value among these weighted correlation coefficients between the area and any of the templates as r_{w1} . We place all eye templates on every such possible position with the midpoint in the left part of the candidate area for eyes. The eyes are searched for only in the left part of this area and let us denote the largest weighted correlation coefficient over all possible eye templates by r_{w2} . In a similar manner we place all eye templates to the right part of the candidate area and denote the largest weighted correlation coefficient by r_{w3} .

Let us consider the coefficient

$$r_w^* = (r_{w1} + r_{w2} + r_{w3}) \cdot I(r_{w1} > 0) \cdot I(r_{w2} > 0) \cdot I(r_{w3} > 0), \tag{4}$$

where 1 denotes an indicator. (4) ignores negative values of the weighted correlation coefficient. The largest value of r_w^* is computed for different candidate mouths.

2.4 Robust correlation

We apply a robust version of the correlation coefficient to the joint localization of both eyes and the mouth in the images. The method is based on the robust estimators for the linear regression context (Chapter 2). Trimming away some pixels corresponds to the idea that some pixels are irrelevant and ignoring large portions of pixels may resemble patch-based approaches [26].

We compute the LWS-based correlation coefficient between the image and the template by transforming the matrices to vectors, computing the LWS regression of the image against the template and finally computing the weighted correlation coefficient with the weights determined by the LWS. For the LWS we consider linearly decreasing weights and the adaptive weights [21]. The LTS-based correlation coefficient is defined in an analogous way and we choose to trim away 20 % of the pixels ($h = 0.8n$).

Tab. 1. Percentages of correct results for the localization of eyes using 6 templates and the joint localization of mouth and eyes using 7 mouth templates and 6 eye templates. The weighted Pearson product-moment correlation coefficient is used with radial weights. The original images were also modified by making the size smaller by 10 % or by rotating the image by $+10^\circ$ or -10° .

Templates	Original image	Smaller image	Rotated image
Eyes, r	1.00	0.80	0.50
Eyes, r_w	1.00	0.97	0.86
Eyes and mouth, r	1.00	0.99	1.00
Eyes and mouth, r_w	1.00	1.00	1.00
Eyes and mouth (relaxed), r	1.00	0.99	0.92
Eyes and mouth (relaxed), r_w	1.00	0.99	0.87

3. Results

3.1 Locating eyes

The set of 6 eye templates (Chapter 2.2) allows both eyes to be localized correctly in 100 % of images of the database, using the product-moment correlation coefficient as the similarity measure between the image and the template. This means that the output of the localization corresponds to true eyes in all 100 % of images of the database. The 100 % correct results are obtained also with the weighted correlation coefficient with radial weights defined to be inversely proportional to the distance of each pixel from the midpoint of the template and yield better results in terms of robustness properties (Chapter 4).

Table 1 presents the results of locating both eyes separately by 6 eye templates and their mirror reflections. Further it gives results described later. Results obtained with r and r_w with radial weights are compared.

3.2 Locating eyes and the mouth

The method (4) for the joint localization of both eyes and the mouth (Chapter 2.3) localizes the mouth and both eyes correctly in 100 % images of our database. Table 1 presents the results of this method obtained with r and with radial weights and also results obtained with a relaxed version, not using the limit on the distance of both eyes. These results support the necessity of such condition. Further Table 1 contains results obtained for images with a modified size or rotation, which will be described in Chapter 3.

3.3 Robust correlation

We use the database with 212 images and the 7 mouth templates and 6 eye templates together with their mirror reflections. Several of the robust correlation methods of Chapter 2.4 yield 100 % correct results in the localization of both eyes and the mouth jointly. These correct results are obtained with the LTS-based correlation coefficient with $h = 0.8n$, the LWS-based correlation coefficient with linear weights and also the LWS-based correlation coefficient with adaptive weights. Typically the outliers trimmed by the LTS or down-weighted by the LWS are located in the neighborhood of eyes, at the boundary of the rectangular templates.

For comparison we consider a robust correlation coefficient based on trimming principal variables $u = x + y$ and $v = x - y$, proposed by [27]. This fails in our application in locating the mouth. The reason can be explained on the correlation coefficient between the mouth template and the mouth. The outliers in the variable u are namely typically in the lips, while the outliers in the variable v correspond to the cheeks. For the correlation coefficient between the mouth template and the non-mouth the outliers are typically present in the cheeks. There is no clear correspondence between the outliers in the principal variables and the original data, which does not allow the mouth to be classified correctly.

To summarize, we apply robust versions of the correlation coefficient based on robust regression with a high breakdown point. These give 100 % correct results in a joint localization of both eyes and the mouth, similarly with the weighted correlation coefficient with radial weights.

3.4 Robustness of the methods

Tab. 2. Performance of the weighted Pearson product-moment correlation coefficient with radial weights in locating both eyes and the mouth jointly. Percentages of correct results.

The method is applied on original images, images rotated by any degree and images slightly modified by noise, occlusion or asymmetry as described in Chapter 3.

Images	r_w with radial weights
Original	1.00
Rotated by any degree	1.00
Noise	1.00
Occlusion	1.00
Asymmetry of the face	1.00

We present validation steps to verify the performance of our methods for the joint localization of the mouth and both eyes. We verify the method (4) on a validation set, further we inspect the properties under rotation by any degree and finally we study the robustness empirically by modifying the original images by introducing additional noise, small occlusion or small asymmetry; Table 2 summarizes our modifications of the original images.

This method does not use any parameters learned over the database of images. Specific properties of this particular database influence the choice of the templates and the expected distances between both eyes and between the mouth and eyes. Nevertheless we verify the performance of the method on a validation set of images. We have photographed 30 randomly selected students at the University of Duisburg-Essen with a compact digital camera. Our conditions were standardized to obtain images of faces with the same illumination, distance from the camera, possibly rotated in the plane by small angles, without facial expressions. We transformed the color images from their size 2048*1536 pixels to grey scale images of size 266*200 pixels so that the size of the faces corresponds to the size of the faces in the original database. We used the method of Chapter 2.3 to localize the mouth and both eyes in the images. The method gives 100 % correct results for these images with our set of 7 mouth templates and 6 eye templates. Nevertheless we consider our method sensitive to the size of the images and therefore we do not verify the methods on other databases of images.

Theoretical properties of template matching concerning robustness with respect to rotation, occlusion or asymmetry are studied by [3]. It follows immediately that (4) is robust in the same situations. On the other hand the sample influence function [28] is not limited and (4) turns out to be vulnerable to highly influential weights.

We start by examining the method for the rotated images, while we retain non-rotated templates. Localizing the eyes separately or the mouth alone does not yield successful results in images rotated by $\pm 10^\circ$; not even are the eyes or mouth localized correctly in non-rotated images. Further we compute the joint localization of both eyes and the mouth for images rotated up to $\pm 10^\circ$. Using 6 eye templates, their reflections in the axial symmetry and 7 mouth templates together with radial weights for all mouth and eye templates, the mouth and both eyes are correctly localized in 100 % of the images.

Nevertheless rotating the image by $\pm 20^\circ$ makes the method collapse in about 50 % of the images.

If the face is rotated by any of angles 10, 20, ..., 350 degrees, the sum of three weighted correlation coefficients is smaller than in a non-rotated face with horizontal eyes. A good strategy is to rotate the face by several different angles and to find the largest value of (4) for each rotation. In this way we localize the mouth and both eyes correctly independently on the initial rotation of the given image. To be specific, we rotated each images by angles 0, 10, 20, ..., 180, ..., 350 degrees. The largest value of r_w^* in each of the 212 images is obtained exactly for such position, when the eyes are in a horizontal position. That leads to the correct localization of the mouth and both eyes in 100 % of the images. The 100 % correct results for images rotated by any degree are obtained also for the LWS-based correlation coefficient with linearly decreasing weights and with the adaptive weights.

The method for localizing the mouth and both eyes contains a double protection against a possible rotation of the face. Firstly the templates are robust to a rotation up to $\pm 10^\circ$. Secondly the coefficient r_w^* attains the largest value exactly for a non-rotated face. Therefore the correct rotation of the face is automatically detected in each face with an arbitrary rotation up to $\pm 180^\circ$ degrees. Table 1 presents results obtained with r and r_w with radial weights for images rotated by $\pm 10^\circ$ or reduced in size by 10 %.

Finally we inspect robustness properties to other non-standard situations. Pixel-independent noise with Gaussian distribution with zero expectation does not harm the results with variance up to $\sigma = (0.11)^2$. The mouth and both eyes are correctly localized also in images rotated by any angle (Figure 3) and spoiled by the noise with zero expectation and variance σ^2 up to 0.01.

We occluded the mouth in every image by a small plaster to examine the local sensitivity of the method. Grey values in a rectangle of size 3*5 pixels are set to 1. Every mouth in the database is modified in



Fig. 3. Localization of the mouth and both eyes. For a particular candidate mouth the eyes are searched for in a relevant candidate area.

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. An example of such occluded mouth is shown in Figure 5. The mouth and both eyes are localized in 100 % of images with such occluded mouths.

To study the effect of asymmetry of the mouth, we increase grey values in the right half of every mouth by a constant, say ϵ . There is no monotone relationship between the value of ϵ and the value of the separation.

The joint search for the mouth and both eyes gives 100 % correct results with ϵ up to 0.15, which is already quite a severe alteration of the original mouth. Here equal and radial weights fail for locating the mouth itself by 7 symmetric mouth templates.

To summarize, the joint localization of mouth and eyes with 7 mouth templates and 6 eye templates has robust properties with respect to small rotation, occlusion and non-symmetry of the image. Robust modifications of the correlation coefficient are applied to the same problem in the next chapter.



Fig. 4. The rotated image additionally modified by noise.

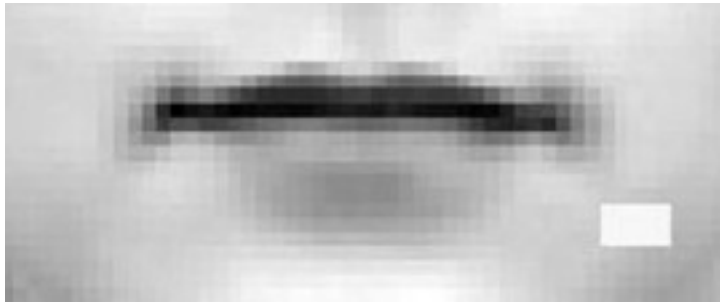


Fig. 5. Mouth modified by a small occlusion (white plaster).

4. Discussion and conclusions

We have proposed a method for the joint localization of the mouth and both eyes in images of faces. The methods are tailor made for the given images of faces with expected usage of the methods in genetic applications. Templates are conceptually simple and clearly interpretable. The method is robust with respect to noise in the images, occlusion or asymmetry of the faces. The results are correct for any initial rotation of the face. The disadvantage of our approach for a general usage is the sensitivity to the size of the images.

Our attempt to apply robust correlation measures to the same task of localizing objects in the image yields promising results and we can recommend robust analogies of the correlation coefficient based on the least trimmed squares and least weighted squares regression for practical usage. Nevertheless the computation of these estimators is very tedious also with approximative algorithms.

The initial denoising was motivated by the necessity to obtain a method robust to noise in the images; its effect may be however small, especially when robust correlation measures with a high breakdown point are used to measure the similarity between the image and the template. While theoretical robustness properties of the robust methods remain for future research, the methods described in this paper can be recommended for practical usage in genetic [17], [18] or anthropologic research [3].

In contrary to standard approaches we examined correlation analysis methods for locating objects in raw images, without a

prior reduction of dimension and feature extraction. Standard approaches use the normalized position or invariant descriptors in order to remove or diminish the effect of rotation or even the change of scale of the images. Nevertheless we contradict a popular belief that statistical methods cannot handle the task of analyzing raw images. Promising results are obtained with the weighted Pearson product-moment correlation coefficient and also robust versions of the Pearson product-moment correlation coefficient. There remains an open problem of computing optimal weights for the weighted correlation coefficient. These should increase the discrimination between those parts of the image corresponding to the template and those which do not. Such problem must be solved again in a robust way, not allowing to obtain highly influential weights by regularizing the problem by introducing a certain upper bound for the optimal weights. An interesting task would be to compare standard classification methods of multivariate statistical analysis and their performance in locating landmarks in the given database of images.

Acknowledgements

This work is supported by Center of Biomedical Informatics, project 1M06014 of the Ministry of Education, Youth and Sports of the Czech Republic. The author is thankful to two anonymous referees for valuable comments and tips for improving the paper.

References

- [1] Yang M.-H., Kriegman D.J., Ahuja N. (2002): Detecting faces in images: A survey. *IEEE Trans. Pattern Anal. and Machine Intel.* 24, No. 1, 34-58.

- [2] Viola P., Jones M.J. (2004): Robust real-time face detection. *Int. Journal of Computer Vision* 57, 137-154.
- [3] Kalina J. (2010): Locating landmarks using templates. *Nonparametrics and Robustness in a Broader Perspectives. A Festschrift in Honor of Professor Jana Jurečková. IMS Collections No. 7.* Accepted, in print.
- [4] Dobeš M., Machala L., Tichavský P., Pospíšil J. (2004): Human eye iris recognition using the mutual information, *Optik* 115, 399-404.
- [5] Lin Z., Davis L.S., Doermann D.S., DeMenthon D. (2007): Hierarchical part-template matching for human detection and segmentation, in *Proceedings of the Eleventh IEEE International Conference on Computer Vision ICCV 2007*, IEEE Computer Society, Washington, 2007, 1-8.
- [6] Wolf L., Huang X., Martin I., Metaxas D. (2006): Patch-Based Texture Edges and Segmentation. In *Leonardis A., Bischof H., Pinz A. (Eds.): Computer Vision - ECCV 2006*, 9th European Conference on Computer Vision, Graz, *Proceedings, Part II. Lecture Notes in Computer Science 3952.*
- [7] Papageorgiou C.P., Oren M., Poggio T. (1998): A general framework for object detection, in *Proceedings of the Sixth IEEE International Conference on Computer Vision ICCV 1998*, IEEE Computer Society, Washington, 555-562.
- [8] Dalal N., Triggs B. (2005): Histograms of oriented gradients for human detection, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR 2005*, IEEE Computer Society, Washington, 2005, pp. 886-893.
- [9] Torralba A., Murphy K.P., Freeman W.T. (2007): Sharing visual features for multiclass and multiview object detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 5, 854-869.
- [10] Tuzel O., Porikli F., Meer P. (2007): Human detection via classification on Riemannian manifolds, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR 2007*, IEEE Computer Society, Washington.
- [11] Böhringer S., Vollmar T., Tasse C., Würtz R.P., Gillessen-Kaesbach G., Horsthemke B., Wiczorek D. (2006): Syndrome identification based on 2D analysis software. *Eur. J. Hum. Genet.* 14, 1082-1089.
- [12] Loos H.S., Wiczorek D., Würtz R.P., Malsburg von der C., Horsthemke B. (2003): Computer-based recognition of dysmorphic faces. *Eur. J. Hum. Genet.* 11, 555-560.
- [13] Würtz R.P. (1997): Object recognition robust under translations, deformations, and changes in background. *IEEE Trans. Pattern Anal. and Machine Intel.* 19, No. 7, 769-775.

- [14] Pitas I., Venetsanopoulos A.N. (1990): Nonlinear digital filters. Kluwer, Dordrecht.
- [15] Jurečková J., Pícek J. (2006): Robust statistical methods with R. Chapman & Hall/CRC, Boca Raton.
- [16] Arya K.V., Gupta P., Kalra P.K., Mitra P. (2007): Image registration using robust M-estimators, *Pattern Recognition Letters* 28, 1957-1968.
- [17] Dunning M.J., Smith M.L., Ritchie M.E., Tavaré S. (2007): beadarray: R classes and methods for Illumina bead-based data, *Bioinformatics* 23, 2183-2184.
- [18] Kalina J. (2010): Robust image analysis in the evaluation of gene expression studies. *ERCIM News, European Research Consortium for Informatics and Mathematics*, No. 82, p. 52.
- [19] Víšek J.Á. (2001): Regression with high breakdown point. In Antoch J., Dohnal G. (Eds.): *Proceedings of ROBUST 2000, Summer School of JČMF, JČMF and Czech statistical society*, 324-356.
- [20] Kalina J. (2007): Locating the mouth using weighted templates. *Journal of applied mathematics, statistics and informatics* 3, No. 1, 111-125.
- [21] Čížek P. (2008): Efficient robust estimation of time-series regression models. *Appl. Math.* 53, No. 3, 267-279.
- [22] Rousseeuw P.J., van Driessen K. (1999): A fast algorithm for the minimum covariance determinant estimator, *Technometrics* 41, 212-223.
- [23] Rousseeuw P.J., Leroy A.M. (1987): Robust regression and outlier detection. Wiley, New York.
- [24] Graf H.P., Cosatto E., Gibbon D., Kocheisen M., Petajan E. (1996): Multi-modal system for locating heads and faces. *Second IEEE International Conference on Automatic Face and Gesture Recognition FG 1996*, 88-93.
- [25] James M. (1987): *Pattern recognition*. BSP Professional books, Oxford.
- [26] Seshadri K., Savvides M. (2009): Robust modified active shape model for automatic facial landmark annotation of frontal faces, *Proceedings of the 3rd IEEE international conference on Biometrics: Theory, applications and systems*, IEEE Press, Piscataway, 319-326.
- [27] Shevlyakov G.L., Vilchevski N.O. (2001): *Robustness in data analysis: criteria and methods*. VSP, Utrecht.
- [28] Hampel F.R., Ronchetti E.M., Rousseeuw P.J., Stahel W.A. (1986): *Robust statistics, The approach based on influence functions*. Wiley, New York.

Contact

RNDr. Jan Kalina, Ph.D.

Institute of Computer Science AS CR

Centre of Biomedical Informatics

Pod Vodárenskou věží 2

182 07 Prague 8

Czech Republic

e-mail: kalina@euromise.cz