Journal of the Applied Mathematics, Statistics and Informatics (JAMSI), **3** (2007), No. 1

Locating the Mouth Using Weighted Templates

JÁN KALINA

Abstract

The aim is automatic location of mouths in two-dimensional grey-scale images of faces. We study theoretical properties of the weighted correlation coefficient, which outperforms the classical sample correlation coefficient as a similarity measure between the template and the image. Robustness to asymmetry of the mouth is studied theoretically. The practical part describes denoising the images and locating the mouth using templates and compares performance of different correlation measures.

Mathematics Subject Classification 2000: 62P99, 62G35
General Terms: Image analysis, Template matching, Robust statistics
Additional Key Words and Phrases: locating landmarks, weighted correlation, least weighted squares

1. INTRODUCTION

This paper studies some theoretical aspects of our practical work in the area of image analysis of faces and presents some results as well. The motivation for this work comes from the cooperation with the Institute of Human Genetics of the University Clinic in Essen, Germany.

The database of images from the Insitute contains 124 grey-scale images of faces of size 192×256 pixels was taken as a part of grants BO 1955/2-1 and WU 314/2-1 of the German Research Council (DFG). 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. This paper is complementary to [Kalina 2007*a*] who optimizes weights for the templates and locates the mouth and both eyes jointly.

This paper has the following structure. Section 2 describes the robustification or denoising, which we perform on every image as the first step of our image processing. Section 3 is devoted to a description of our locating the mouth using templates and an example with real data is presented in details. The weighted correlation coefficient turns out to outperform the classical sample correlation coefficient as a similarity measure between the template and the image. Therefore Section 4 derives some properties of the weighted correlation. Section 5 presents an application of different versions of a robust correlation coefficient to the example of locating the mouth. Finally Section 6 discusses robustness issues of the template matching.

This research comes from the Jaroslav Hájek center for theoretical and applied statistics, project LC06024 of the Ministry of education, youth and sports of the Czech Republic.

2. ROBUSTIFICATION OF THE IMAGE

Denoising or a robustification is very often the first natural step of any image processing, see [Kalina and Davies 2004] for a list of references on various transformations removing noise.

In this paper we use the following approach to denoise the images of faces. For each pixel the grey values from its circular neighbourhood are considered and a robust estimator of location is computed from them. The information about the coordinates is ignored. We have examined the least median of squares, least weighted squares and least trimmed squares estimators. All three are regression estimators and we present them in the special situation as estimators of the location (shift) parameter. Later they are used in applications in Section 5.

Let us consider the standard location model

$$Y_i = \beta + e_i, \quad i = 1, \ldots, n \tag{1}$$

with observed values $Y_1, \ldots, Y_n$, expectation β and i.i.d. random errors $e_1, \ldots, e_n$ with expectation zero. Let us consider (any) estimate $b \in \mathbb{R}$ of the parameter β. By

$$u_i(b) = Y_i - b, \quad i = 1, \ldots, n$$

we denote the residual corresponding to the i-th observation. Let us order squared residuals

$$u^2_{(1)}(b) \leq u^2_{(2)}(b) \leq \cdots \leq u^2_{(n)}(b).$$

The least median of squares (LMS) estimator in the model (1) is proven by [Rousseeuw and Leroy 1987] to be equivalent to the mean of the shortest half of the data.

The least weighted squares regression (LWS) was proposed by [Víšek 2001] as one of robust regression methods with a high breakdown point. The least weighted squares estimator of the location parameter β requires weights $w_1, \ldots, w_n$ to be chosen *a priori*, however they are assigned to particular observations only in an implicit way during the computation. The estimator is namely defined by

$$b_{LWS} = \arg\min_{b \in \mathbb{R}} \sum_{i=1}^{n} w_i u^2_{(i)}(b).$$

The least trimmed squares (LTS) estimator defined in [Rousseeuw and Leroy 1987] is a special case of the least weighted squares. An integer h (the trimming constant) satisfying $h > n/2$ must be chosen *a priori* which corresponds to the number of "good" data while completely ignoring the remaining $n-h$ observations. Then there are h weights equal to 1 and the remaining weights are zero. The LWS and LTS estimators of the location parameter allow the following interpretation.

LEMMA 2.1. *The least weighted squares estimator of the locations parameter corresponds to the weighted mean of the data with such permutation of the weights yielding the smallest weighted variance of the data.*

The least trimmed squares estimator of the locations parameter corresponds to the mean of the h-tuple of the data with the smallest variance (among all h-tuples).

Therefore the computation of the LTS estimator of the location parameter can be based on the following rule.

LEMMA 2.2. *Consider the total number of $n-h+1$ following groups containing h consecutive observations:*

$$Y_1,\dots,Y_h; \quad Y_2,\dots,Y_{h+1}; \quad \dots \quad ;Y_{n-h+1},\dots,Y_n.$$

The least trimmed squares of the location parameter is the mean of that h-tuple with the smallest variance.

There is no such a simple rule for computing the least weighted squares, because the computation complexity of the estimator is higher.

We have examined the denoising of images by the LMS or LTS with different trimming constants or LWS with different weights. Results obtained by the least median of squares are rather poor because they remove contrast and the resulting picture is rather grayish. Results of LWS or LTS are considerably better, their difference is however not big and we have not attempted to compare the estimators systematically, our conclusion is a simple recommendation that a robustification (with any method) should be performed. Later we explain that we use blurred templates for the mouth and the mouth areas of the robustified images resemble them visually, which gives an argument in favour of denoising the images.

We also performed the robustification by fitting a linear regression fit similarly with [Gather et al. 2006] and the results are not very different from using robust estimators of location. The residual images obtained by subtracting the robustified image from the original one still visually resemble the original image, so the noise is not perfectly extracted from the signal.

3. LOCATING THE MOUTH

Basics

Template matching is a tailor made method for object detection which uses the information about the ideal shape. References on template matching used in the area of image analysis of faces are summarized by [Yang et al. 2002] and more recent papers by [Kalina 2007*a*]. [Viola and Jones 2004] use a similar combination of several very simple features, which yields reliable as well as robust results in classifying parts of the face.

We need to distinguish precisely if the mouth is located correctly. Let a given template be compared with an area of an image of the same size as the template itself. We consider the part of the image to belong to the *mouth area*, if its midpoint has a distance from the manually located midpoint of the mouth smaller than four pixels. All other areas with the midpoint outside of the mouth area, will be called *non-mouths*. We remark that the mouth area contains the lips, but never reaches the nostrils nor the chin.

Measures of correlation

The literature mentions only the classical sample correlation coefficient as the only similarity measure between the template and the image. In our work we use also

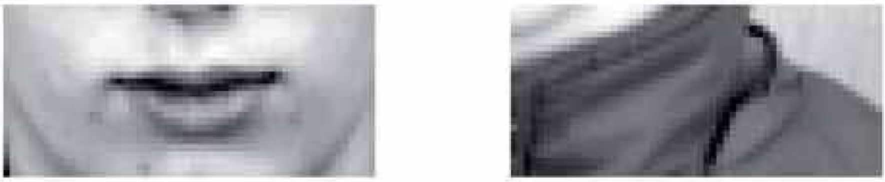

Fig. 1. The mouth and a suspicious nonmouth.

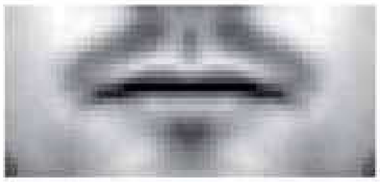

Fig. 2. A template for the mouth.

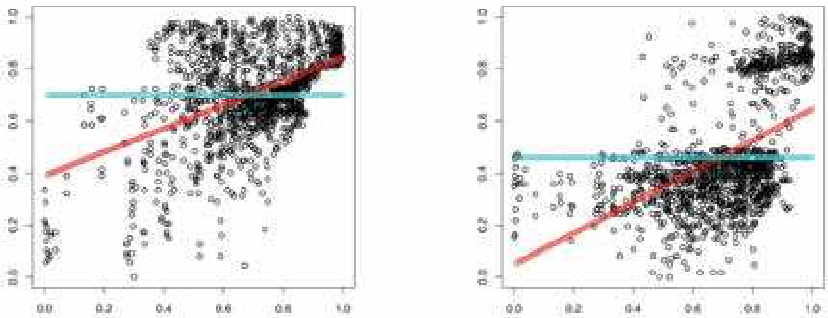

Fig. 3. Left: mouth against the mouth template. Right: nonmouth against the same template. Least squares regression (increasing trend) and arithmetic mean (horizontal).

the weighted correlation coefficient, which is a weighted analogy of the sample correlation coefficient with some specified weights.

The (sample) weighted correlation coefficient for data vectors

$$\boldsymbol{x} = (x_1, \ldots, x_n)^T \quad \text{and} \quad \boldsymbol{y} = (y_1, \ldots, y_n)^T$$

and non-negative weights

$$\boldsymbol{w} = (w_1, w_2, \ldots, w_n)^T \quad \text{satisfying} \quad \sum_{i=1}^{n} w_i = 1 \tag{2}$$

is defined by the formula

$$r_W(\boldsymbol{x},\ \boldsymbol{y};\ \boldsymbol{w}) = \frac{\sum_{i=1}^{n} w_i(x_i - \bar{x}_W)(y_i - \bar{y}_W)}{\sqrt{\sum_{i=1}^{n}[w_i(x_i - \bar{x}_W)^2] \sum_{j=1}^{n}[w_j(y_j - \bar{y}_W)^2]}},$$

where $\bar{x}_W = \sum_{i=1}^{n} w_i x_i$ and $\bar{y}_W = \sum_{i=1}^{n} w_i y_i$ are weighted means.

Table I. Detailed results of the regression of the mouth against the template and of the nonmouth against the template.

		Least squares	LTS $h = 0.92n$	Max r over 92 % of pixels
Mouth	b_0	0.391	0.389	0.218
	b_1	0.452	0.453	0.661
	SSE	35.4	19.9	21.6
	SST	45.8	28.3	40.6
	r	0.476	0.545	0.684
Nonmouth	b_0	0.049	0.018	-0.160
	b_1	0.596	0.606	0.878
	SSE	48.8	32.5	35.2
	SST	66.8	47.9	60.8
	r	0.519	0.566	0.649

Example

We present an example in which the sample correlation fails in correct locating the mouth. Figure 1 shows a mouth of one lady and a collar of her shirt. Both are matrices of grey values of the dimension 26×56 pixels. The aim is to classify the mouth correctly using the bearded template from Figure 2, in other words to discriminate between the mouth and the collar. In general locating the mouth is a task to classify each part of the image as a mouth or nonmouth.

The sample correlation between the mouth and the bearded template attains in this example only 0.476, while the sample correlation between the nonmouth and the template is larger, namely 0.519. Therefore the nonmouth is more suspicious to be the mouth than the true mouth itself. Although the bearded template is not the best possibility to search for a nonbearded mouth, a more detailed study shows that the main problem is using the sample correlation as the similarity measure.

The next explanation is based on Figure 3. The left plot contains grey values of the bearded template on the horizontal axis and the mouth from Figure 1 on the vertical axis. Actually both variables are standardized to fill the interval $[0, 1]$ which does not however influence the correlation between them. First of all the image shows that it is reasonable to use the linear model for the relationship between the mouth and the mouth template.

The sample correlation coefficient is known to be (up to the sign) equivalent to the coefficient of determination in the linear regression. The classical sample correlation coefficient compares the variability explained by the linear model with the total variability in the linear regression.

In the left image of Figure 3 the increasing line corresponding to the least squares regression fit of the mouth against the template gives a reasonably good fit. The horizontal line corresponds to the mean of grey values of the mouth and represents its more or less typical value. The sample correlation is based on comparison of the linear model (increasing line) with the submodel (horizontal line).

In the right image of Figure 3 there are grey values of the nonmouth plotted against those of the template. The pixels in the top right corner corresponding to the top corners of the nonmouth have a strong leverage effect. Also the pixels of the lips in the left bottom corner of the scatter plot have a slight leverage effect.

These two leverage effects cannot push the line upwards because the linear regression line must go through the center of gravity of the data, which is the intersection point of the two lines. The leverage effect therefore tries to rotate the increasing line, but the two effects work in contradictory directions. The horizontal line in the right image is not a suitable submodel. The mean is not a typical value of the response and for such mixture of two clusters a suitable submodel can be hardly found.

A better insight to this example can be obtained in Table I. The left column considers linear regression of the mouth from Figure 1 against the bearded template and gives numerical values of the sums of squares together with the coefficient of determination and the sample correlation between the mouth and the template. The left column of the bottom part contains analogous values for the nonmouth from Figure 1 compared with the same template. Here b_0 denotes the intercept and b_1 the slope of the linear regression. Further SSE and SST denote the error sum of squares and total sum of squares respectively and finally r stands for the sample correlation.

To summarize the results of least squares, there is a larger variability of the response and also a larger regression sum of squares in the right image of Figure 3 than in the left image. The coefficient of determination is larger in the right image and this is based on comparing the quality of the linear model with the quality of the submodel. Other columns of Table I devoted to robust correlation are described in Section 5.

The weighted correlation with radial weights between the mouth and the template equals 0.66, while between the nonmouth and the template 0.38. Larger weights are namely assigned to the lips of the template, which makes the linear trend in the left scatter plot of Figure 3 steeper and in the right scatter plot more horizontal, improving the discrimination between the mouth and the nonmouth.

Results

Now we use different templates to locate the mouth in the 124 available images. First the denoising is used with the method of least weighted squares with linear weights as described in Section 2.

Our work does not use heuristical simplifications assuming for example that the mouth has a usual position in the middle of the image; the aim is a general method for locating the mouth based entirely on templates.

We consider one mouth template of the size 26×56 pixels containing dark lips with a lighter moustache coming from [Kalina 2007*b*] as the mean of four mouths from different bearded men. It locates the mouth correctly in 91 % of the images of the database when the sample correlation is used as a similarity measure between the template and the image. The performance with Spearman's rank correlation attains only 80 %, while with the weighted correlation and radial weights 100 %, which outperforms the results of all the remaining mouth templates. The performance with the least trimmed squares trimming away 20 % of the data is 99 % and least weighted squares with linear weights classify the mouth in 100 % of images of the database. The explanation of the success of the bearded template is the following. Reliable templates should resemble real mouths and on the other hand should be as different from all non-mouths as possible. The bearded template combines these two contradictory requirements well.

A combination of two templates has a worse performance in comparison with the best template itself, because some of non-bearded pictures become problematic. It happens that every of the non-bearded templates is highly correlated with some of such areas as the chin, hair or eyebrows and this correlation is larger than the correlation between the bearded template and the mouth. Therefore when two templates are used jointly, namely when the best bearded template is used together with another non-bearded one. the method inherits the performance of the weaker template. This is remarkable also for robust versions of the correlation coefficient.

We created seven mouth templates of different sizes as averages of real mouths so that highly correlated mouths are used to compute the average. Such rectangular area of the image with the largest correlation over *all* mouth templates is classified to be the suspicious area and the output of the image processing. In other words we search for the maximum of the maximal values over different templates and over the whole image. These templates together locate the mouth in 100 % of cases with the sample correlation and also in 100 % of cases with the weighted correlation with radial weights, but only in 94 % with Spearman's rank correlation and less than 90 % for various versions of the robust correlation coefficient.

4. SOME PROPERTIES OF THE WEIGHTED CORRELATION

The aim of this section is to study properties of weighted correlation and its relationship to weighted regression. Firstly we study the weighted regression in general and later we proceed to a special case with only one independent variable and to weighted correlation.

Let us consider the weighted regression model

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + e_i, \quad i = 1, \ldots, n, \tag{3}$$

where the i.i.d. random errors $\boldsymbol{e} = (e_1, \ldots, e_n)^T$ fulfill

$$\mathsf{E}\boldsymbol{e} = \mathbf{0} \quad \text{and} \quad \mathsf{var}\,\boldsymbol{e} = \sigma^2 \boldsymbol{W}^{-1} = \sigma^2 \mathsf{diag}\left\{\frac{1}{w_1}, \ldots, \frac{1}{w_n}\right\}. \tag{4}$$

The response is a random variable and capital letters $Y_1, \ldots, Y_n$ should be used, but we prefer using the same notation throughout the section in both contexts of regression and correlation. (4) is not fulfilled if any of the weights is zero, but then the observation is ignored completely and all the following statements are correct as well.

Let us rewrite the model in the matrix notation as $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$, use the (generalized) least squares estimator $\boldsymbol{b}_W = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y}$ and denote fitted values by

$$\hat{\boldsymbol{y}}_W = (\hat{y}_{W1}, \ldots, \hat{y}_{Wn})^T = \boldsymbol{X}\boldsymbol{b}_W.$$

Let us use the notation $\mathbf{1} = (1, \ldots, 1)^T$ for the vector of length n.

LEMMA 4.1. *In the weighted regression model (3) with weights (2) let the residuals by defined by*

$$\boldsymbol{u} = (u_1, \ldots, u_n)^T = \boldsymbol{y} - \hat{\boldsymbol{y}}_W. \tag{5}$$

Then it holds

$$r_W(\boldsymbol{x},\ \boldsymbol{u};\ \boldsymbol{w}) = 0.$$

Proof: The numerator of the desired weighted correlation coefficient equals

$$\sum_{i=1}^n w_i x_i u_i - \left(\sum_{i=1}^n w_i x_i\right)\left(\sum_{j=1}^n w_j u_j\right).$$

The first sum is zero because $\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{u} = 0$. The weighted sum of residuals is also zero, which follows from the set of normal equations for the weighted regression model with the intercept, so the whole weighted correlation is zero.

LEMMA 4.2. *In the weighted regression model (3) with weights (2) the following statements are true:*

- $\hat{\boldsymbol{y}}_W^T\boldsymbol{W}\boldsymbol{u} = 0$;
- $\boldsymbol{1}^T\boldsymbol{W}\boldsymbol{u} = 0$;
- $(\hat{\boldsymbol{y}}_W - \bar{y}_W\boldsymbol{1})^T\boldsymbol{W}(\boldsymbol{y} - \bar{y}_W\boldsymbol{1}) = (\hat{\boldsymbol{y}}_W - \bar{y}_W\boldsymbol{1})^T\boldsymbol{W}(\hat{\boldsymbol{y}}_W - \bar{y}_W\boldsymbol{1})$.

Proof: The first part

$$\hat{\boldsymbol{y}}_W^T\boldsymbol{W}\boldsymbol{u} = \boldsymbol{b}_W^T\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{u} = 0$$

follows from the normal equations. The second part

$$\boldsymbol{1}^T\boldsymbol{W}\boldsymbol{u} = \boldsymbol{1}^T\boldsymbol{W}\boldsymbol{y} - \boldsymbol{1}^T\boldsymbol{W}\hat{\boldsymbol{y}}_W = 0$$

follows from the normal equations only when the model includes an intercept. Finally these two parts will be now used to prove that

$$(\hat{\boldsymbol{y}}_W - \bar{y}_W\boldsymbol{1})^T\boldsymbol{W}(\boldsymbol{y} - \bar{y}_W\boldsymbol{1}) - (\hat{\boldsymbol{y}}_W - \bar{y}_W\boldsymbol{1})^T\boldsymbol{W}(\hat{\boldsymbol{y}}_W - \bar{y}_W\boldsymbol{1}) =$$

$$= \hat{\boldsymbol{y}}_W^T\boldsymbol{W}\boldsymbol{y} - \boldsymbol{1}^T\bar{y}_W\boldsymbol{W}\boldsymbol{y} - \hat{\boldsymbol{y}}_W^T\boldsymbol{W}\hat{\boldsymbol{y}}_W - \boldsymbol{1}^T\bar{y}_W\boldsymbol{W}\hat{\boldsymbol{y}}_W =$$

$$= (\hat{\boldsymbol{y}}_W - \bar{y}_W\boldsymbol{1})^T\boldsymbol{W}\boldsymbol{y} - (\hat{\boldsymbol{y}}_W - \bar{y}_W\boldsymbol{1})^T\boldsymbol{W}\hat{\boldsymbol{y}}_W =$$

$$= (\hat{\boldsymbol{y}}_W - \bar{y}_W\boldsymbol{1})^T\boldsymbol{W}(\boldsymbol{y} - \hat{\boldsymbol{y}}_W) = \hat{\boldsymbol{y}}_W^T\boldsymbol{W}\boldsymbol{u} - \bar{y}_W\boldsymbol{1}^T\boldsymbol{W}\boldsymbol{u} = 0$$

and the proof is completed.

Now we examine the relationship between the weighted correlation coefficient and weighted regression. The following definitions and relations will be needed. The residual and total sums of squares in the weighted regression model allow the decompositions

$$\mathsf{SSE}_W = (\boldsymbol{y} - \hat{\boldsymbol{y}}_W)^T\boldsymbol{W}(\boldsymbol{y} - \hat{\boldsymbol{y}}_W) = \boldsymbol{y}^T\boldsymbol{W}\boldsymbol{y} - \hat{\boldsymbol{y}}_W^T\boldsymbol{W}\hat{\boldsymbol{y}}_W$$

and

$$\mathsf{SST}_W = (\boldsymbol{y} - \bar{y}_W\boldsymbol{1})^T\boldsymbol{W}(\boldsymbol{y} - \bar{y}_W\boldsymbol{1}) = \boldsymbol{y}^T\boldsymbol{W}\boldsymbol{y} - \boldsymbol{1}^T\bar{y}_W\boldsymbol{W}\bar{y}_W\boldsymbol{1}. \tag{6}$$

Based on this the regression sum of squares $\mathsf{SSA}_W = \mathsf{SST}_W - \mathsf{SSE}_W$ can be expressed the in the form

$$\mathsf{SSA}_W = \hat{\boldsymbol{y}}_W^T \boldsymbol{W} \hat{\boldsymbol{y}}_W - \mathbf{1}^T \bar{y}_W \boldsymbol{W} \bar{y}_W \mathbf{1} =$$

$$= (\hat{\boldsymbol{y}}_W - \bar{y}_W \mathbf{1})^T \boldsymbol{W} (\hat{\boldsymbol{y}}_W - \bar{y}_W \mathbf{1}).$$

LEMMA 4.3. *The weighted coefficient of determination* $R_W^2(\boldsymbol{x},\ \boldsymbol{y};\ \boldsymbol{w})$ *in the weighted regression (3) is equal to the square of the weighted correlation coefficient* $r_W^2(\boldsymbol{y},\ \hat{\boldsymbol{y}}_W;\ \boldsymbol{w})$.

Proof: Using the standard definition of the weighted coefficient of determination and simple steps from the previous lemma we obtain

$$R_W^2(\boldsymbol{x},\ \boldsymbol{y};\ \boldsymbol{w}) = 1 - \frac{\mathsf{SSE}_W}{\mathsf{SST}_W} = \frac{\mathsf{SSA}_W}{\mathsf{SST}_W} = \frac{(\hat{\boldsymbol{y}}_W - \bar{y}_W \mathbf{1})^T \boldsymbol{W} (\hat{\boldsymbol{y}}_W - \bar{y}_W \mathbf{1})}{(\boldsymbol{y} - \bar{y}_W \mathbf{1})^T \boldsymbol{W} (\boldsymbol{y} - \bar{y}_W \mathbf{1})} =$$

$$= \frac{\left[(\hat{\boldsymbol{y}}_W - \bar{y}_W \mathbf{1})^T \boldsymbol{W} (\hat{\boldsymbol{y}}_W - \bar{y}_W \mathbf{1})\right]^2}{(\boldsymbol{y} - \bar{y}_W \mathbf{1})^T \boldsymbol{W} (\boldsymbol{y} - \bar{y}_W \mathbf{1}) \cdot (\hat{\boldsymbol{y}}_W - \bar{y}_W \mathbf{1})^T \boldsymbol{W} (\hat{\boldsymbol{y}}_W - \bar{y}_W \mathbf{1})} =$$

$$= \frac{\left[(\hat{\boldsymbol{y}}_W - \bar{y}_W \mathbf{1})^T \boldsymbol{W} (\boldsymbol{y} - \bar{y}_W \mathbf{1})\right]^2}{(\boldsymbol{y} - \bar{y}_W \mathbf{1})^T \boldsymbol{W} (\boldsymbol{y} - \bar{y}_W \mathbf{1}) \cdot (\hat{\boldsymbol{y}}_W - \bar{y}_W \mathbf{1})^T \boldsymbol{W} (\hat{\boldsymbol{y}}_W - \bar{y}_W \mathbf{1})} =$$

$$= \frac{\left[\sum_{i=1}^n w_i (y_i - \bar{y}_W)(\hat{y}_{Wi} - \bar{y}_W)\right]^2}{\sum_{i=1}^n w_i (y_i - \bar{y}_W)^2 \cdot \sum_{i=1}^n w_i (\hat{y}_{Wi} - \bar{y}_W)^2} = r^2(\boldsymbol{y},\ \hat{\boldsymbol{y}}_W;\ \boldsymbol{w}),$$

which was to be proven.

In the rest of this section we consider a special form of the weighted regression model (3) with one independent variable

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, \ldots, n. \tag{7}$$

LEMMA 4.4. *The weighted coefficient of determination* $R_W^2(\boldsymbol{x},\ \boldsymbol{y};\ \boldsymbol{w})$ *in the weighted regression with one independent variable (7) is equal to the square of the weighted correlation coefficient between the response and the independent variable, namely* $R_W^2(\boldsymbol{x},\ \boldsymbol{y};\ \boldsymbol{w}) = r_W^2(\boldsymbol{x},\ \boldsymbol{y};\ \boldsymbol{w})$.

Weighted correlation with a template is therefore equivalent to fitting weighted regression. The proof is an immediate consequence of Lemma 4.3 and the regression-invariance of the weighted correlation coefficient.

LEMMA 4.5. *In the weighted regression model (7) with weights (2) and residuals (5) it holds that*

$$r_W(\boldsymbol{y},\ \boldsymbol{u};\ \boldsymbol{w}) = \sqrt{1 - r_W^2(\boldsymbol{x},\ \boldsymbol{y};\ \boldsymbol{w})}.$$

Proof: From Lemma 4.4 we get

$$r_W^2(\boldsymbol{x},\ \boldsymbol{y};\ \boldsymbol{w}) + r_W^2(\boldsymbol{u},\ \boldsymbol{y};\ \boldsymbol{w}) = R_W^2(\boldsymbol{x},\ \boldsymbol{y};\ \boldsymbol{w}) + R_W^2(\boldsymbol{u},\ \boldsymbol{y};\ \boldsymbol{w}). \tag{8}$$

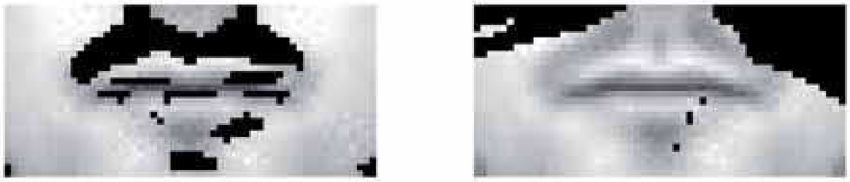

Fig. 4. 20 % of the data determined as outliers by least trimmed squares in the mouth (left) and nonmouth (right) of Figure 1.

Now

$$R^2_W(\boldsymbol{x},\ \boldsymbol{y};\ \boldsymbol{w}) = 1 - \frac{\mathsf{SSE}_W(\boldsymbol{x} \sim \boldsymbol{y};\ \boldsymbol{w})}{\mathsf{SST}_W}, \tag{9}$$

where SST_W is expressed in (6) and $\mathsf{SSE}_W(\boldsymbol{x} \sim \boldsymbol{y};\ \boldsymbol{w})$ is the weighted residual sum of squares in the linear regression of $\boldsymbol{x}$ against $\boldsymbol{y}$ with weights $\boldsymbol{w}$. In other words we take $\boldsymbol{x}$ as the response of $\boldsymbol{y}$. An analogous expression is

$$R^2_W(\boldsymbol{u},\ \boldsymbol{y};\ \boldsymbol{w}) = 1 - \frac{\mathsf{SSE}_W(\boldsymbol{u} \sim \boldsymbol{y};\ \boldsymbol{w})}{\mathsf{SST}_W}. \tag{10}$$

The numerators of (9) and (10) consider two regression models, each with $\boldsymbol{y}$ as explanatory variable. Lemma 4.1 gives the orthogonality between $\boldsymbol{x}$ and $\boldsymbol{u}$ in the weighted sense and therefore (8) equals 1, which is equivalent to the statement of the lemma.

Lemma 4.5 explains what happens with residuals in the template matching. The weighted correlation between every mouth and the template is desirable to be large. The residuals between the mouth and the residuals of the weighted regression between the mouth and the template is zero. The weighted correlation between the residuals and the template is already uniquely determined and it is desirable to be small.

On the other hand the residuals of the fit of a non-mouth against the template should have a large weighted correlation with the template, as the non-mouth should have a small weighted correlation with the template.

5. ROBUST CORRELATION

This section presents results of locating the mouth using different robust modifications of the correlation coefficient. The sensitivity of the sample correlation coefficient to outliers is well known and from the theoretical point of view analyzed for example by [Shevlyakov and Vilchevski 2001]. The weighted correlation however is sensitive to outliers, bad leverage points as well as influential data points with large weights. Therefore we also work with robust versions of the correlation coefficient.

One possibility is to compute the least trimmed squares estimator of the regression line with a given trimming constant (let us say 20 % of the data) and then to compute the correlation coefficient only with the "good" 80 % of the data ignoring the outliers. This corresponds to assigning weights equal to zero or one in order to minimize the weighted residual sum of squares. *Another* possibility is

Table II. Different measures of robust correlation in the example of locating the mouth.

		Robust correlation between the template and the mouth	Robust correlation between the template and the nonmouth
Least squares		0.476	0.519
LTS	$h = 0.92n$	0.545	0.553
	$h = 0.8n$	0.778	0.243
	$h = 0.7n$	0.873	0.196
	$h = 0.6n$	0.924	0.414
Maximal r over			
	92 % of pixels	0.665	0.649
	80 % of pixels	0.817	0.760
	70 % of pixels	0.886	0.830
	60 % of pixels	0.926	0.889
LWS with weights			
	linear	0.725	0.470
	radial	0.726	0.470
Maximal r_W over permutations of weights			
	linear	0.815	0.810
	radial	0.685	0.654

to search for such permutations of the zero or one weights which maximizes the weighted coefficient of determination. A weighted analogy of this approach starts with the least weighted squares regression estimator with certain weights; the output is the permutation of the weights and finally the weighted correlation coefficient is computed with these weights.

Both robust approaches to the correlation give very similar results in most of the cases. However Table II presents an example yielding very different results of different robust correlation measures. The similarity is measured between the mouth from Figure 1 and the template from Figure 2 and between the nonmouth from Figure 1 and the same template. The robust coefficient was computed with a modification of the algorithm of [Víšek 2001]

The first section of Table II is devoted to least squares, the next sections to the least trimmed squares with different values of the trimming parameter. We can see that as the number of trimmed data increases the correlation computed from the remaining data values can be larger or on the other hand smaller compared to the original correlation. The worst outliers in the relationship between the mouth and the template correspond mainly to the moustache of the template or to the lips. Outliers of the fit between the nonmouth and the template correspond to the top corners of the nonmouth. The next section of Table II is computed with the correlation coefficient which is maximal when trimming away the corresponding given percentage of data (considered to be outliers).

A weighted analogy of these two trimmed approaches is possible. Both start with weights which are given but their permutation are determined only in an implicit way. One lets the least weighted squares regression to determine the permutation and computes the weighted correlation using all the data and these weights. The other approach searches for such permutation of weights which maximizes the weighted correlation coefficient. Results are given again in Table II again for two choices of weights. One of them are linear and the other radial with largest weights corresponding to the midpoint. For both choices the smallest weights correspond to the moustache in the fit of the mouth against the template and to the corners of the nonmouth in its fit against the template.

Results obtained with trimming away 20 % of the data values and keeping the remaining 80 % of pixels are shown in Figure 4. The outliers determined by the least trimmed squares are shown black. In the fit of the mouth against the template, the outliers correspond mainly to the moustache. In the fit of the nonmouth against the template, they correspond to the top corners, in other words to the light corners of the nonmouth. The results are very similar with both described approaches to the trimmed correlation.

An interesting result is obtained with trimming away 8 % of the data values and keeping the remaining 92 % pixels. That is one of rare cases with very different results of the two described trimmed versions of the correlation coefficient. And now we discuss the results which are shown in Table I.

In the fit of the mouth against the template, the least trimmed squares does not give very different results from least squares because only a small percentage of data is trimmed away. These are scattered more or less all over the rectangle. The maximal correlation over a subset of 92 % of the data trims away the area corresponding to the moustache of the template. The pixels of the lips become more influential, namely their leverage effect becomes stronger and the regression line steeper. The correlation is much larger compared to that obtained from all the data points.

In the relationship between the nonmouth and the template, the outliers determined by the least trimmed squares are also scattered over the rectangle, but nevertheless their large part belongs to the top corners. The maximal correlation over all subsets containing 92 % of the data determines very different weights. The lips get the smallest weights while the top corners get larger ones. Then the regression line is very steep and the trimmed correlation again much larger than that computed from all the data points.

To summarize, the mouth as well as nonmouth have the robust correlation with the template much larger than the sample correlation. While the least trimmed squares are not so different from least squares and the nonmouth is more correlated with the template than the mouth, the maximal correlation over a subset of 92 % of the data gives opposite results. The difference between the mouth and nonmouth is however not so big, as the trimmed correlation between the mouth and the template is 0.684 and between the nonmouth and the template 0.649.

An opposite classification result is obtained with the other approach to the trimmed correlation. There the maximal r is found over all permutations of weights equal to 1 or 0 such that exactly 8 % of the pixels obtain a zero weight. The trimmed correlation between the mouth and the template 0.684 is larger than the trimmed correlation between the nonmouth and the template 0.649.

We can conclude that different robust estimators of the correlation can give very different results. In one particular example we see that the correlation based on robust regression with a larger percentage of outliers improves the discrimination between the mouth and nonmouth dramatically compared with the classical correlation or with the trimmed correlation of [6]. In principle the idea of using a robustified version of the correlation seems to be reasonable and well interpretable.

6. ROBUSTNESS OF THE RESULTS

An important aspect of the methods of image analysis of faces is the sensitivity to violations of the standardized conditions. In this paper we study theoretically the sensitivity to asymmetry of the mouth. Another problem is that the correlation coefficient and its modifications are valid for independent identically distributed random variables. While the autocorrelation structure of our data is discussed by [Kalina 2007a], the robustness of the methods to noise (corresponding to heteroscedastic errors) is currently under our research.

The grey values of the mouth will be denoted by $\boldsymbol{x} = (x_1, \ldots, x_n)^T$ ignoring the coordinates of each pixel and the weights in corresponding pixels by (2). The weighted mean in the original mouth $\bar{x}_W$ and the weighted variance $S_W^2(\boldsymbol{x};\, \boldsymbol{w})$ are defined by

$$\bar{x}_W = \sum_{i=1}^{n} w_i x_i \quad \text{and} \quad S_W^2(\boldsymbol{x};\, \boldsymbol{w}) = \sum_{i=1}^{n} w_i (x_i - \bar{x}_W)^2.$$

The next lemma describes the weighted correlation between the template and an asymmetric mouth of a special structure.

LEMMA 6.1. *Let $\boldsymbol{t}$ be a template with an even number of columns symmetric along the vertical axis. We assume that the weights (2) and a mouth $\boldsymbol{x}$ have the same size as the template and are symmetric along the vertical axis. Let us perform the following asymmetric modification $\boldsymbol{x}^*$ of the mouth $\boldsymbol{x}$. Grey values on one side of the axis are equal to those of $\boldsymbol{x}$ and the remaining are increased by ε compared to those from $\boldsymbol{x}$. Then the weighted correlation between the template and the modified mouth $\boldsymbol{x}^*$ can be expressed by*

$$r_W(\boldsymbol{x}^*,\, \boldsymbol{t};\, \boldsymbol{w}) = r_W(\boldsymbol{x},\, \boldsymbol{t};\, \boldsymbol{w}) \frac{S_W(\boldsymbol{x};\, \boldsymbol{w})}{\sqrt{S_W^2(\boldsymbol{x};\, \boldsymbol{w}) + \frac{\varepsilon^2}{4}}}.$$

Proof: In the modified mouth $\boldsymbol{x}^*$ the weighted mean and weighted variance can be easily shown to be equal to

$$\bar{x}_W^* = \bar{x}_W + \frac{\varepsilon}{2} \quad \text{and} \quad S_W^2(\boldsymbol{x}^*;\, \boldsymbol{w}) = S_W^2(\boldsymbol{x};\, \boldsymbol{w}) + \frac{\varepsilon^2}{4}.$$

Clearly both values increase compared to the mean of the grey values and weighted variance of the original mouth.

Let us denote the weighted mean of grey values of the template $\boldsymbol{t}$ by $\bar{t}_W$ and the weighted covariance between the mouth and the template by $S_W(\boldsymbol{x},\, \boldsymbol{t};\, \boldsymbol{w})$.

This is defined by

$$S_W(\boldsymbol{x},\, \boldsymbol{t};\, \boldsymbol{w}) = \sum_{i=1}^{n} w_i(t_i - \bar{t}_W)(x_i - \bar{x}_W).$$

Assuming the symmetry of the mouth, template and the weights, the weighted covariance between the template and the modified mouth equals the weighted covariance between the template and the original mouth $S_W(\boldsymbol{x},\, \boldsymbol{t};\, \boldsymbol{w})$. The conclusion is the desired formula

$$r_W(\boldsymbol{x}^*,\, \boldsymbol{t};\, \boldsymbol{w}) = \frac{S_W(\boldsymbol{x}^*,\, \boldsymbol{t};\, \boldsymbol{w})}{S_W(\boldsymbol{x}^*;\, \boldsymbol{w})S_W(\boldsymbol{t};\, \boldsymbol{w})} = r_W(\boldsymbol{x},\, \boldsymbol{t};\, \boldsymbol{w})\frac{S_W(\boldsymbol{x};\, \boldsymbol{w})}{S_W(\boldsymbol{x}^*;\, \boldsymbol{w})} =$$

$$= r_W(\boldsymbol{x},\, \boldsymbol{t};\, \boldsymbol{w})\frac{S_W(\boldsymbol{x};\, \boldsymbol{w})}{\sqrt{S_W^2(\boldsymbol{x};\, \boldsymbol{w}) + \frac{\varepsilon^2}{4}}},$$

which concludes the proof. Here $\boldsymbol{x}$ can be any symmetric part of the image, not necessarily a mouth.

Lemma 6.1 has the following interpretation. There is no contradiction between performance and robustness. The best weights yielding a large weighted correlation between the mouth and the template retain this weighted correlation large also for asymmetric mouths.

We remark that the template matching cannot be expected to be reasonably robust with respect to changes of the size of the head or rotation by a small angle. Therefore we propose to search for the mouth and both eyes jointly and to carry out the research of the robustness properties of such approach.

REFERENCES

(1) GATHER, U., SCHETTLINGER, K., FRIED, R. 2006. Online signal extraction by robust linear regression. *Technical report 23/2006,* SFB 475, Dortmund.

(2) KALINA, J. 2007*(a)*. Locating landmarks using templates. Doctoral dissertation (Dr.rer.nat.), University of Duisburg-Essen.

(3) KALINA, J. 2007*(b)*. Robust methods for template matching. *Folia Fac. Sci. Nat. Univ. Masaryk. Brunensis, Mathematica*, 9 pp. Accepted.

(4) KALINA, J., AND DAVIES, P.L. 2004. Locating eyes. In *Robust 2004, Proceedings of the 13-th summer school JČMF*, Třešť, June 2004, J. ANTOCH AND G. DOHNAL, Eds. JČMF, Prague, 169 – 176.

(5) ROUSSEEUW, P.J., AND LEROY A.M. 1987. Robust regression and outlier detection. Wiley, New York.

(6) SHEVLYAKOV, G.L., AND VILCHEVSKI, N.O. 2001. Robustness in data analysis: criteria and methods. VSP, Utrecht.

(7) VIOLA, P., AND JONES, M.J. 2004. Robust real-time face detection. *Int. Journal of Comp. Vision* **57**, Nr. 2, 137 – 154.

(8) VÍŠEK, J.Á. 2001. Regression with high breakdown point. In *Robust 2000, Proceedings of the 11-th summer school JČMF*, Nečtiny, September 2000, J. ANTOCH AND G. DOHNAL, Eds. JČMF and Czech statistical society, Prague, 324 – 356.

(9) 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.

Locating the Mouth Using Weighted Templates

Ján Kalina,
KPMS MFF UK,
Sokolovská 83, 186 75 Praha 8,
Czech Republic
e-mail: kalina@math.sk

Received August 2007