ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 216, 2008

Jan Kalina[*]

WEIGHTING IN THE TEMPLATE MATCHING

ABSTRACT. In the area of image analysis of templates, we have proposed and implemented a method for locating landmarks in 2D-images of faces. It uses the weighted correlation coefficient as a similarity measure between the template and the image. The weights are selected in the optimal way to improve the discrimination of parts of the image which correspond to the template from those which do not. The method does not use specific properties of faces.

Key words: Image analysis, locating landmarks, correlation coefficient, robust non-parametric discrimination.

I. INTRODUCTION AND MOTIVATION

The primary motivation for this work is the automatic location of landmarks in images of human faces. We are working with a database of images which come from the Institute of Human Genetics, University Clinic, Essen, Germany (projects BO 1955/2–1 and WU 314/2–1 of the German Research Council). This database contains 124 grey value images of the size 192 times 256 pixels, each image from a different person. The faces have about the same size but are rotated in the plane by small angles. Present software is highly sensitive to small rotations.

Each picture is a matrix with the size 192 times 256 pixels. A grey value in the interval [0,1] corresponds to each pixel, where low values are black and large values white. Pictures are taken under the same conditions, with the person sitting straight against the camera looking straight at it. The Institute 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).

[*] MSc, Jaroslav Hájek Center for Theoretical and Applied Statistics, Charles University, Prague, Czech Republic

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 paper by Loos *et al.* (2003). There are images of 55 patients considered and each of them can be classified to one of five groups according to a genetic malformation deforming the face. The correct classification rate of the syndromes by an automatic procedure based only on the image of the face was 84 % which is considered remarkably successful. A new study at the Institute works with a larger number of patients to one of 10 different syndromes and on average there are about 12 individuals present in each group. The aim is again to recognize the syndrome in each person. For different syndromes the success rate lies between 75 % and 80 %.

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 forty landmarks in each face. One possibility is the manual selection, which can be in spite of its accuracy criticized as subjective and not scientific. When the landmarks are located repeatedly even by the same person, the results can be different. As the second approach the institute uses an automatic method, namely a software implemented by a commercial company, cooperating with the Institute of Neuroinformatics of the Ruhr University in Bochum. This software based on the algorithm Würtz (1997) and partially also on Wiskott *et al.* (1997) 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.

The algorithm makes however more than just a comparison of the jets of the new image with the jets of each of the training images separately. It combines the jets of the new image with *any* combination of jets from different training images. Then the best similarity can be obtained with the mouth from one person, nostrils from another person and so on. Once these local experts are selected, some transformations are possible, for example local shift of particular points or scale transformations. The everyday experience of the genetics researchers with the available software is however unsatisfactory because of its extremely high sensitivity 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. From the practical point of view it is actually desirable to search for landmarks rather than the mouth and eyes, but a natural first step is to find the mouth and eyes themselves. The information about their position in a given image simplifies the future task of locating the landmarks, which are prominent points (not only) of the mouth and eyes.

The practical performance of the methods is mostly desirable. Therefore we apply the methods not only to standardized images of the given database, but also to images with a different size or rotation of faces. In the text we consider only the rotation in a plane with the whole face well visible from the front. Another important aspect of the methods for locating landmarks in faces is robustness to noise in images.

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. In the literature we have not found references on a sophisticated construction of templates nor their optimization.

In this text in Chapter 2 different measures of fit between the template and the image are described. Chapter 3 is devoted to the optimization of the template and Chapter 4 comments results in the search for the mouth and eyes. This paper is supported by the Jaroslav Hájek Center for Theoretical and Applied Statistics, Czech Republic. Some of the results of Chapters 3 and 4 are based on my Ph.D. thesis written under the advisor Prof. Dr. P. L. Davies, University of Duisburg-Essen. The results go further than those presented in Kalina (2006).

I am thankful to the Institute of Human Genetics of the University Clinic Essen for the access to their database of images. These were taken as a part of the grants BO 1955/2–1 and WU 314/2–1 of the DFG.

II. MEASURES OF FIT

In the image analysis literature we have found the correlation coefficient to be the only measure of similarity between the template and an area of the picture. Such area is considered to be suspicious which has the largest value of the sample correlation with the template. We have examined also other measures, for example Spearman's rank correlation coefficient.

Some of the possible robust measures of similarity between images are defined in the context of robust regression. The picture is modelled as a response of the template. For robust analogies of the correlation coefficient we have used its trimmed and weighted versions summarized in Kalina (2006), which are correlation counterparts of least trimmed squares (LTS) and least weighted squares (LWS) regression. These suffer from a considerable reduction of the speed even if a modification of the fast algorithm for least weighted squares regression of Kalina (2003) is applied. Moreover the results are not satisfactory. The reason is that an eye consists of both black and white ares. Replacing the gray values by ranks removes the contrast between both groups and very robust approaches completely discard one of the groups. A non-robust approach is therefore desirable and the correlation coefficient can be recommended. Furthermore we have observed that robust estimates of the slope or robustified sum of squares, for example the trimmed sum of squares in the LTS context or the weighted sum of squares for the LWS, do not perform well, seemingly because they are not invariant to linear transformations of the data.

We have observed that the weighted correlation coefficient with different choices of weights performs well in locating the mouth and eyes using templates. The weighted correlation coefficient is invariant to a linear transformation of the data and represents a natural weighted analogy of the classical sample correlation coefficient. It is also equivalent (up to the sign) to the weighted coefficient of determination in the weighted regression.

III. OPTIMIZATION

We explain the optimization on the example of locating the mouth. The idea is to retain one particular template and change the weights. This is a discrimination problem in which we aim at improving the separation between mouths and non-mouths in the training set of 124 images in which the mouths are already identified. The solution can be then used in the classification context to locate the mouth in new images.

With a given template and given weights, the separation between a particular mouth and non-mouth can be measured by the ratio of the weighted correlation between the mouth and the template and the weighted correlation between the non-mouth and the template. The Fisher's transform is used on the weighted correlation coefficient to further improve the separation. We use the minimax approach to improve the worst case over the whole database. The algorithm giving an approximation to the highly non-linear problem retains a given template and starts with initial weights. The best position of every mouth is found as the position with the highest weighted correlation with the mouth template. In every image we also find the non-mouth with the worst separation from the best mouth. The algorithm optimizes the linear approximation to the worst separation and is supplemented by a rough algorithm modifying the weights in pixels selected at random, which allows for a setting out from a local extreme and further improves the worst separation.

Another possibility is to retain the weights and optimize the template in the same way. Then it turns out that optimizing the linear approximation to the separation function converges better to the solution, namely the rougher approach modifying the template in randomly selected pixels does not bring a further improvement of the results. The best results are obtained starting with equal weights and optimizing the template with a following optimization of the weights for this fixed template.

IV. LOCATING THE MOUTH

The weights are optimized over the training set of 124 images using one mouth template with a light moustache. This enables us to localize the mouth well in both bearded and non-bearded faces. We start with radial initial weights which are inversely proportional to the distance of each pixel from the midpoint.

The solution has the tendency to be degenerated and a large portion of the mass of the weights is concentrated in a small number of pixels. The (unconstrained) optimization of the weights is using special properties of the data such as perfect symmetry of each mouth. Although this cannot be identified in the training set, the optimal weights are sensitive to atypical features and we prefer the constrained solution for locating the mouth in new images. Therefore we have placed constraints} on the optimization toregularize the problem and to find a robust solution without highly influential pixels. The method gives very reliable results in locating the mouth over the whole database of images. Moreover the method turns out to be robust with respect to small changes of the size and rotation of the face, noise in images, non-symmetry or local deformations of the mouth.

This is a general method not using specific properties of the mouth. While the symmetry of the mouth template and corresponding weights is assumed, in the search for eyes we do not assume such symmetry. We search for each eye separately and these are located reliably in the whole database of images. The difficulty in locating eyes consists in their smaller size and also the discrimination of one eye against the other.

REFERENCES

Kalina J. (2003): Autocorrelated disturbances of robust regression. In Fournier B. et al.(eds.): *Proceedings European Young Statisticians Meeting (EYSM 2003)*, Ovronnaz, Switzerland, 65–72.

Kalina J. (2006): Robust correlation applied to locating landmarks. Submitted to *Proceedings Olomouc days of applied mathematics (ODAM 2006)*, Olomouc, Czech Republic.

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

Wiskott L., Fellous J.-M., Krüger N., Malsburg von der C. (1997): Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Anal. and Machine Intel.* **19**, No. 7, 775–779.

Würtz R.P. (1997): Object recognition robust under translations, deformations, and changes in background. *IEEE Trans. Pattern Anal. and Machine Intel.* **19**, No. 1, 34–58.

Jan Kalina

WAGI DOPASOWYWANIA SZABLONÓW

Zaproponowaliśmy i zastosowaliśmy metodę lokalizacji punktów orientacyjnych obrazów dwuwymiarowych. Podstawą jej jest użycie ważonego współczynnika korelacji pomiędzy szablonem a obrazem. Wagi są wybierane w sposób optymalny dla poprawy rozróżnienia części obrazu, które odpowiadają szablonowi od tych które nie odpowiadają. Metoda nie odwołuje się do specyficznych własności twarzy.