GENDER RECOGNITION METHODS USEFUL IN MOBILE AUTHENTICATION APPLICATIONS

ZOFIA STAWSKA, PIOTR MILCZARSKI

Department of Computer Science,
Faculty of Physics and Applied Informatics,
University of Lodz, Poland

Soft biometrics methods that involve gender, age and ethnicity are still developed. Face recognition methods often rely on gender recognition. The same applies to the methods reconstructing the faces or building 2D or 3D models of the faces. In the paper, we conduct study on different set of gender recognition methods and their mobile applications. We show the advantages and disadvantages of that methods and future challenges to the researches. In the previous papers, we examined a range variety of skin detection methods that help to spot the face in the images or video stream. On acquiring faces, we focus on gender recognition that will allow us to create pattern to build 2D and 3D automatic faces models from the images. That will result also in face recognition and authentication, also.

Keywords: face identification, gender recognition, face authentication, biometrics

1. Introduction

In the soft biometrics face recognition and authentication is still very important problem. One of the examined feature that is gender. There are many gender classifications approaches and methods. They can be divided in feature-based and appearance-based methods [36]. They are presented in this paper. They use different approaches but they usually are trimmed on the same database. When they are compared on the same FERRET database [32] they show around 90 per
cent accuracy. When the database is different than the one used by the authors of the gender classification method, then the methods’ accuracy drops to 60-70%.

In the previous papers [1, 2] we concentrated on the skin colour classification. Finding skin colour pixel is important in face detection. On finding a face at the picture, e.g. [3], one can examine the face to extract their feature. After determining the gender of the face it will be easier to recognize and authenticate the person. From the other point apart from computer designed methods and applications there are more and more mobile apps that can do the same because of their hardware specification. In this article we wanted to check the suitability of gender as one of biometric features in the task of authentication based on the face.

The paper is organised as follows. In section 2, we describe different face recognition aspects. In the next section, gender classification methods by face are presented. There are shown different approaches and methods e.g. neural network, SVM, Adaboost etc. In section 4, there are briefly described face databases and datasets, e.g. FERET etc. Analysis and evaluation of gender classification methods is shown and discussed in section 5. Discussion about use of the presented methods in mobile applications contains section 6. Finally, we present some concluding remarks.

2. Face recognition aspects

The gender recognition might be one of the factors that fastens the person authentication because you limit the search for a person to 50% of the population. On the market the applications like How-Old.net allows you to upload your frontal picture then the service recognizes faces and analyses them to determine their age. The application gender accuracy is quite high but age accuracy depends on e.g. makeup, light, face size in the picture etc.

2.1. Skin colour and face detection

In the previous papers [1, 2] there were presented several approaches to the skin colour classification. To determine gender of the person the first step in automatic methods is finding skin colour pixel or area. Then the area is checked whether it belongs to a face. It can be done using e.g. template matching methods. Nowadays there is a method very popular and helpful in face detection that was presented by Viola and Jones [3]. On finding a face at the picture, e.g. [3], one can examine the face to extract their feature for future processing.

For example in Windows 10 facial recognition is a log-in option. Systems and services like those mentioned above are more and more spread but not perfect. That is why the problem is still vital.
2.2. Face recognition and authentication

In Windows 10 facial recognition is a log-in option. The same has appeared in mobile devices and operating systems. Systems and services like those mentioned above are more and more spread but not perfect and reliable. In the security systems recognizing and authentication of the face can strengthen the systems and data security. That is why the problem is still vital because the methods are used in a wide range of applications for access control, security and database indexing etc. [38].

The next problem is whether to classify the gender of the faces using 2D or 3D images. Most of the methods base on 2D images and their processing, e.g. [39]. There number of 3D processing methods used in gender recognition is minimal. There are a few derived methods for the gender recognition and facial asymmetry using 3D methods [37].

3. Gender classification methods by face

The automatic gender classification methods have been developed for over 20 years. The first methods were presented by Cottrell and Metcalfe and Golomb at al. in the 1990. They both used a multi-layer neural network as classification method and was tested on the manually aligned faces.

In the next years many authors tried different methods. They experiments with different network inputs: the set of geometrical features extracted from faces [4] or pixel-based input [5]. Also many different classification methods were used: neural network [7, 8], hyper basis function networks [4], radial basis function network [5], Gabor wavelets [6, 9], threshold Adaboost [10,13] LUT Adaboost [12], support vector machine (SVM) [11], linear discriminant analysis (LDA) [11], Self Organizing Map (SOM) [13], Bayesian classifier [11].

3.1. Preprocessing

Before the face will be subject to classification some preprocessing may be necessary. The classifiers are usually sensitive to imperfection of the image for example bed illuminations. Good preparation of the face image can reduce obtaining process influence to the classification results.

In this phase we can normalize brightness and contrast using histogram equalization, improve the face image geometric features, reduce the image size (number of pixels).
3.2. Face features

Generally, in the gender recognition task, we can distinguish two methods of feature extraction: geometric based and appearance based.

The first method requires finding of the face characteristic points as a nose, mouth, eyes, ears or hair. These points are called fiducial points. Geometric relationships between this point (fiducial distances) are used as a feature vector in the classification process. Importance of these distances in the gender discrimination task is confirmed by the psychophysical studies. The number of using fiducial distances may be different, for example 18 in the [4] or 22 in the [15].

Geometric based methods allow to reduce the size of the feature vector using in the classification process, but some useful information can be thrown away. Further, the points extraction process must be very precise.

Appearance based methods work on the image pixels that were previously transformed on the local or global level. At the local level image can be divided into lower windows or specific face regions such as mouth, nose or eyes. This approach preserves natural geometric relationships which can be used as a naive features. It may be advantage because it is quite difficult to determine which geometrical features are important in the gender recognition task. However, we have large number of features (each pixel is treated as a feature). This method also does not provide a good explanation why an image is classified as a man or woman.

Pixel intensity values can be used directly as an input of the neural network or support vector machine (SVM). However, for better results, some preprocessing is necessary. The image is usually normalized to equalize lighting or geometrical variations and resized to lower number of image pixels. The size of image using as a classifier input can be various, for example Moghaddam and Yang [16] used 21x12 pixel images and Baluja and Rowley [20] 20x20 pixel images.

Using pixel intensity values we have a large number of features. Dimension reduction methods allow to restrain this problem using representation of an image in the reduced dimension space. Most popular method is Principal Component Analysis (PCA) [11, 17, 18]. Also another dimension reduction method was studied, for example Jain et al. [19] were using Independent Component Analysis (ICA).

Viola and Jones [3] proposed Haar-like features for fast face detection (Fig. 1). The sum of the pixels which lie within the white rectangles are subtracted from the sum of pixels in the grey rectangles to obtain the value of a rectangle feature. Adaboost is used to select the features. Also Shakhnarovich et al. [10] used this method for fast gender classification of videos in real-time.
Local Binary Patterns (LBP) are features that are calculated from pixel intensities in a pixel neighborhood. The basic idea is that as many binary values are created as there are pixels in the neighborhood of the center pixel [29, 30]. At the end these are concatenated to one binary value. Originally LBP was defined for 3x3 pixel neighborhood but later it was extended to different neighborhoods. Alexandre et al. [25] combined LBP with intensity and histogram of edge directions. Other variants inspired by LBP have been proposed for example by Zheng et al. [27].

Lowe [31] proposed Scale Invariant Feature Transform (SIFT) - features which are invariant to the scaling, translation and rotation transformations in the image domain and further partly invariant to illuminations changes. The advantage of this kind of method is the preprocessing step is not necessary. Wang et al. [24] proposed combination the SIFT with Gabor features.

3.3. Classification methods

**Neural network** is one of the most popular classification method used in many different domain. Also in gender recognition case many authors proposed this method as a classifier. Before use the image as an input of the neural network it is scaled to the chosen size. The size of image determines the number of network input nodes, so we would like to have the image as small as possible but its size must be enough to keep the essential information about face features. The least used size of face image was 8x8 pixels [14]. The neural network used for the gender classification has one output node.

The neural network is trained in round, by presentation labelled face example images as an network input. The connections weights are changed after each example presentation closer output to the expected result. After each round the validation image set is used for calculate the output error.

**Support Vector Machine (SVM)** assume that samples of different classes can be separated in the higher dimensional space using the transformed features. In the task of gender recognition by face, features can be image pixel intensities. For the transformation a kernel function is needed. Many authors proposed different kernel functions as radial basis function or polynomial kernels. It is possible to combine SVM with different kinds of features, we can use face image pixels intensity as well as local binary pattern.

It seems that SVM is recently the mostly used classification method for the face gender recognition [11, 16, 17, 18, 19, 25, 26].
The idea of Adaboost algorithm is that group of the weak classifiers can create together a strong classifier. The weak classifiers can be anything, they must only allow to classify appropriate data examples. Also the kind of using features is whichever. Most popular weak classifiers are threshold, mean and LUT.

In the threshold weak classifiers the value calculated for an image with the single classifier is compare to the threshold. Each classifier has its own threshold value. The optimal threshold is finding in the training process. The mean weak classifiers have a threshold by first calculating mean feature values separately for the male and female training example.

LUT Adaboost [12] differs from others weak classifiers so that instead of a threshold it places the calculated feature value to a bin. The number of bins is determined before the training.

4. Face databases and datasets

For evaluation the accuracy of the gender recognition we need face database. There are several publicly available databases that have been used for experiments. Sometimes researchers combine several databases to obtain larger set of images or take a subset of the database to exclude inadequate images. Publicly available datasets examples are AR, FERET, BioID, CAS-PEAL-R1, MORPH-2, LFW. Databases FERET and CAS-PEAL-R1 deserve special attention.

FERET [32] is one of the most popular evaluation dataset in the gender recognition domain. It contains above 14000 images of almost 1200 persons. The faces have a variety pose and varying facial expressions. The images have also some variation in illumination.

CAS-PEAL [33] is a large scale Chinese face database. It contains above 99000 images of 1040 persons (595 males and 445 females). They have varying pose, expression, accessories and lighting. A subset of the database CAS-PEAL-R1 which contains 30900 images is available for research. The images were taken in a controlled environment.

In contrast to FERET and CAS-PEAL, LFW [34] dataset images were required in uncontrolled environment which allow to study of face recognition in more natural condition. It contains above 13000 images.

The MORPH [35] database was collected in real-world conditions. The dataset also contains metadata in the form of age, gender, race, height, weight, and eye coordinates.

Not all databases have information about gender. If the images are not annotated with this information, researchers must do it manually.

It appears that recently the MORPH-2 and LFW are the best datasets for face gender classification (first for controlled and second one for uncontrolled).

There are also private databases as BCMI [26].
Table 1 presents authors reported results of the selected works on face gender recognition. It contains information about features extraction method and classifier used, and also training and test database used by author. In the last column we have average total classification rate.

<table>
<thead>
<tr>
<th>Author</th>
<th>Feature extraction</th>
<th>Classifier</th>
<th>Training data</th>
<th>Test data</th>
<th>Classification rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moghaddam (2002) [16]</td>
<td>Pixel values</td>
<td>SVM-RBF</td>
<td>FERET</td>
<td>cross validation</td>
<td>96.62</td>
</tr>
<tr>
<td>Castrillon (2003) [17]</td>
<td>PCA</td>
<td>SVM+ temporal fusion</td>
<td>Video frames</td>
<td></td>
<td>98.57</td>
</tr>
<tr>
<td>Buchala (2005) [18]</td>
<td>PCA</td>
<td>SVM –RBF</td>
<td>Mix (FERET, AR, BioID)</td>
<td>cross validation</td>
<td>92.25</td>
</tr>
<tr>
<td>Jain (2005) [19]</td>
<td>ICA</td>
<td>SVM</td>
<td>FERET</td>
<td>FERET</td>
<td>95.67</td>
</tr>
<tr>
<td>Fok (2006) [21]</td>
<td>Pixel values</td>
<td>Convolutional neural net.</td>
<td>FERET</td>
<td>cross validation</td>
<td>97.2</td>
</tr>
<tr>
<td>Aghajanian (2009) [22]</td>
<td>Patchbased</td>
<td>Bayesian</td>
<td>Web images</td>
<td>Web images</td>
<td>89</td>
</tr>
<tr>
<td>Demirkus (2010) [23]</td>
<td>SIFT</td>
<td>Bayesian</td>
<td>FERET</td>
<td>Video seq.</td>
<td>90</td>
</tr>
<tr>
<td>Li (2011) [26]</td>
<td>LBP+ hair &amp; clothing features</td>
<td>SVM</td>
<td>FERET</td>
<td>FERET</td>
<td>95.8</td>
</tr>
<tr>
<td>Zheng (2011) [27]</td>
<td>LGBPPLDA</td>
<td>SVMAC</td>
<td>FERET</td>
<td>CAS-PEAL</td>
<td>99.1, 99.8</td>
</tr>
<tr>
<td>Shan (2012) [28]</td>
<td>LBP</td>
<td>SVM-RBF</td>
<td>LFW</td>
<td>cross validation</td>
<td>94.8</td>
</tr>
</tbody>
</table>

Source: own elaboration based on results reported by authors
The authors of the selected works use different feature extraction and classification methods. It can be seen that most popular classifier is SVM. All feature extraction algorithms are the appearance based methods.

The FERET database is the most often used as a training set, but researchers are using also different datasets as a datasets composed of the web images.

The classification rate is the ratio of correctly classified test examples to the total number of test examples. Most popular method of testing results is cross validation, but many authors reports also that they train and test results on the different datasets.

Comparison between the methods is very difficult because researchers use different datasets and parameters for evaluation their methods. Even if authors use the same FERET database, they can select the different subset of images from this database. For example, some researchers used only frontal face images and others may also use non frontal face images.

The best results were reported by Zheng et al. [27]. For FERET database they obtained 99.1% and for CAS-PEAL database even better result – 99.8%, but the authors were selected only frontal face images from the datasets. For the images taken in uncontrolled environment the results are a little worst The best result in this case was 94.8% obtained by Shan et al. [28] using LFW dataset.

Mäkinen and Raisamo [36] carried out the comparison study of the selected gender recognition methods. Each method was tested in the same condition – starting with the same way of the face recognition and testing on the same datasets.

Table 2 presents the results for the FERET and the Web testing databases. In the first case separate set of the FERET images was used for training, in the second case all FERET dataset was used for training.

<table>
<thead>
<tr>
<th>Method</th>
<th>Classification rate (FERET images)</th>
<th>Classification rate (Web images)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>without hair (%)</td>
<td>with hair (%)</td>
</tr>
<tr>
<td>Neural network</td>
<td>92.22</td>
<td>90</td>
</tr>
<tr>
<td>SVM</td>
<td>88.89</td>
<td>82</td>
</tr>
<tr>
<td>Threshold Adaboost</td>
<td>86.67</td>
<td>90</td>
</tr>
<tr>
<td>LUT Adaboost</td>
<td>88.89</td>
<td>93.33</td>
</tr>
<tr>
<td>Mean Adaboost</td>
<td>88.33</td>
<td>90</td>
</tr>
<tr>
<td>LBP + SVM</td>
<td>80.56</td>
<td>92</td>
</tr>
</tbody>
</table>

*Source: Mäkinen E., Raisamo R. [36]*
It can be seen that the results obtained for the FERET testing set are significantly better than the results obtained for the web images. The authors suppose that the reason may be greater similarity of the training and testing sets when they used FERET set in both cases. Classifier adapts better to the recognition examples in this situation. Further, when the web images are used as a testing set, they have more quality variations, so they are more difficult to classify.

6. Discussion – use in mobile applications

The time that is needed to detect faces in an image depends on the methods that are used. Nonetheless, the time is quite big in comparison with the time taken in gender classification. That is why in practice, gender classification can be real-time with all the described methods [36]. This is generally true when using a standard PC or high-end smartphones. However, there are huge differences in training times between the methods. In search for the best and most efficient method or set of methods used for recognition and/or authentication in mobile devices one has to take into account the architecture of the system.

If it is service based system then mobile application can send acquired picture to the system service run on powerful hardware. The whole procedure can be done in the system and mobile application can receive results only. That will shorten the time. Of course the quality and resolution of the picture as well as the quality of the Internet connection will have a stress on it.

If all the calculations are done of client/mobile application side, then the face detection, gender recognition and authentication methods can consume a lot of time depending on the smartphone/tablet specification.

Both methods have their dis- and advantages in e.g. robustness, reliability or security measures.

7. Conclusions

In the paper, we focused on different set of gender recognition methods and their possible mobile applications. We show the advantages and disadvantages of that methods and future challenges to the researches.

Comparison between the methods is very difficult because researchers use different datasets and parameters for evaluation their methods. Taking into account that the authors use the same FERET database, they can select the different subset of images from this database, e.g. some researchers used only frontal face images and others may also use non frontal face images.

From the Table 2 it can be clearly seen that the results obtained for the FERET testing set are significantly better than the results obtained for the web
images. It might be a result of uniformity of FERET database. If the web images are used as a testing set, they have more quality variations, so they are more difficult to classify.

Because we use in mobile solutions images taken in unspecified conditions, then our results will be closer to the results obtained while using web set images.

REFERENCES


