Chapter 1

INTRODUCTION

The face is one of the most common biometric characteristics used by humans to make a personal recognition people have a considerable ability to identify, extract and interpret information related to human face. Due to the nearly exponential growth of the rate of face images uploaded to the internet over the last decade, the development of automatic facial analysis techniques is becoming increasingly necessary with the purpose of improving the accuracy of systems that emule the visual abilities of the human being.

Nowadays, face recognition systems are used in different domains, such as access control, mobile appli- cations, video surveillance, and government related applications. In particular, these systems have become a valuable tool used by criminal investigators in the forensic field. In this scenario, the low resolution of images or videos may impair a reliable recognition of individuals, because many details may be missing. However, the cate- gorization of elementary properties such as age and gender, may still be recovered with sufficient reliability or probability of success.

Face analysis and recognition systems have shown to be a valuable tool for forensic examiners. Particularly, the automatic estimation of age and gender from face images, can be useful in a wide range of forensic applications. In this work we propose to use a local appearance descriptor in a component based way, to classify age and gender from face images. We subdivide a face image into regions of interest based on automatically detected landmarks, and represent them by using Histograms of Oriented Gradient (HOG). The representations obtained from different face regions are inserted to binary classifiers to estimate the age and gender of the person in the image. Experimental analysis show the good results of this component based approach, and its additional benefits when face images are affected by occlusions.

The fusion of these “soft attributes” with biometric recognition systems has shown to improve the overall recognition when confronting high variability conditions. Besides, in the case of the analysis of video data taken from security cameras in a crime scene, manually parsing hours of videos may be cumbersome and lead to mistakes. Therefore, an automatic system spotting the portions of the video where male/female subjects of a certain age are seen, can be useful to perform a preliminary parsing of the video.

Within this context, we motivate this work as a support to the forensic examiner or the police officer wishing to identify a suspect of known age and gender. It should be noted that, as the automatic system only relies on the face information, it can not be fooled by contextual (and maybe purposively chosen to be misleading) information such as hairdressing, gait or outfit.

Age and gender are two demographic attributes that influence the morphology and appearance of a face. Facial age estimation refers to the automatic labeling of age groups or the specific ages of individuals based on features extracted from the face. With the progress of age, the appearance of human faces exhibits remarkable changes related to its contour, the shape of facial features (eyes, nose, mouth, etc.) and its distribution, the skin pigmentation and the emergence of wrinkles, among others.

The aging process is influenced by external (health, living style, smoking, etc.) and internal (genetics, gender) factors, which makes age estimation difficult for humans, and even more difficult for machines. Gender classification, on the other hand, is the process of automatically assigning one of the two sex labels (male/female) to a facial image. Studies have shown that human observers can easily differentiate between a male and female, achieving an accuracy above 95% based on looking at the face. However, the accuracy rate reduces to just above chance when considering child faces once again, what represents a challenge to some extent for humans, becomes a much greater challenge for computer vision systems. In addition to these intrinsic problems, automatic age and gender classification are ag- gravated, like other branches of facial analysis, by a set of factors that include lighting conditions, facial expressions and pose variations, usually found on forensic scenarios.

1.1 Background

There has been a growing interest in automatic age and gender classification, as it has become relevant to an increasing amount of applications such as human computer interaction, surveillance, biometrics, intelligent marketing and many more. Facial age and gender from the face image of a person is one such significant demographic attribute. It presents a review of automatic facial gender classification and age estimation framework in computer vision. While highlighting the challenges involved during classification of images captured under unconstrained conditions or may be the laborious process of gathering the face images for age estimation, as aging is the uncontrolled and slow process. A comprehensive survey for facial feature extraction methods and face databases for gender and age estimation studied in the past couple of decades is mentioned. Evaluation and result based performance achieved for various face images from different databases has been explained.

In this work we aim at using a local appearance de- scriptor to classify both, age and gender, from face im- ages. Taking into account that different parts of the face can contribute differently to those attributes classification, we propose to apply a component based face representation methodology, and determine the most significative regions for each purpose. Binary classifiers are trained for each facial components and some combinations of them to obtain a prediction for gender and age. Under our proposed component based approach, the main contributions of this work are: (1) we redefine the face regions in order to make them more robust under pose and face expressions variations, (2) we introduce a new representation for each region, that is used to estimate both age and gender and, (3) we analyze the benefits of using this approach when the face images are affected by occlusions (e.g. cap, eyeglasses or scarf), which is very common on forensic scenarios.

1.2 Existing System and their Drawbacks

Besides the huge amount of literature existing on both age and gender classification from face images, there is a hot research topic in face image analysis related to facial attributes classification. While a large list of attributes are classified on those works, focus here on age and gender, which are intrinsic factors and can not be easily modified as other attributes such as hair color or the use of glasses for example. Hence, they can be more effectively used by forensic examiners for reducing a list of suspects or make a search in a given video, where other attributes can be unknown or faked. Moreover, found that specifically age estimation (child, young, adult, etc.) are among the worst attributes classified on those works.

The work of Kumar et, where a total of 10 regions are used to classify 73 attributes, aiming at strengthens the face recognition process. Low level features such as pixel intensities, edge magnitude and orientation are used to represent the facial components and by using a computationally expensive procedure, different classifiers (for all possible combinations of features with three normalization techniques and three aggregation approaches) are trained and evaluated to automatically selects the best combination for a given attribute. In this work, we propose to use Histograms of Oriented Gradients (HOG), which have shown to be a powerful local appearance descriptor.

Besides, redefine the facial components to be used in order to make them more adequate for our specific problem. Another difference w.r.t that work is that makes a deeper analysis of the impact of using different face regions for predicting age and gender, which allows us to explore this task for occluded face images, which are very common in forensic scenarios. There are some works that analyze the influence of face aging in facial components, and although they have found that the changes in facial appearance related to age and gender differ from one region to another, to the best of our knowledge this has not been used for the age estimation purpose. Moreover, the possibilities of estimate both age and gender using the same descriptors can be beneficial for practical applications.

1.3 Proposed Approach

Visual features from human faces provides one of the more important sources of information for age and gender estimation. A typical attribute classification framework is described below. For an input face image, the first step is to perform face detection. Then, the detected face is aligned based on its eyes coordinates with the purpose of correcting possible transformations such as translation, rotation and scaling. Facial features extracted from multiple aligned face images are used, along with their corresponding attribute labels, from pretrained models. Finally, the learned classification function can be used to recognize the target attribute for a test face image. Main contribution in this work lies on the feature extraction process and the subsequent representation of the face image.

There is (a) one region for the whole face, and (b) seven regions corresponding to functional parts of the face: forehead, eyebrows, eyes, nose, cheeks, mouth and chin.

(a) (b)

Figure 1.1 Face regions automatically obtained in an aligned face image.

Propose to subdivide a face image into eight regions of interest that were defined using 68 landmark points detected in the face alignment step. The complete set of regions is shown in figure 1.1, where each region corresponds to a functional part of the face (i.e. forehead, eyebrows, eyes, nose, mouth, cheeks, chin and the complete face region). This subdivision allows us to take advantage of the common geometry shared by faces: the regions contain the face parts despite changes in pose, small alignment errors and morphological differences between individuals.

Our regions are somehow similar to those defined. In addition to those shown in figure 1.1, they consider a region for the hair zone and also a region for the area between mouth and nose. Since project is only interested on age and gender we propose to simplify this subdivision. Men and women from almost all ages can both have long hair and similar haircuts or hairstyles, so we consider that a hair region is not only poorly discriminative for both gender and age, but could also provide misleading information. Also, we discard the use of an additional region for the area be- tween mouth and nose, instead we enlarge our mouth region enough to cover this face part, in a more compact region. For the cheekbones, unlike the ellipses used, we employ triangular shapes that we believe cover the area better, especially when there are changes in pose and diverse facial expressions. Our regions are automatically obtained based on landmarks estimation, and this increase the robustness against variations in pose and expressions. It should be noticed that in the case of eyes, eyebrows and cheeks, each region is composed by two parts corresponding to both sides of the face image.

Wide Resnet(WRNs):

The depth of a neural network is the number of layers, but width usually refers to the number of neurons per layer, or for convolutional layers, the number of feature maps per layer. Residual block with identity mapping can be represented by the following formula:

$$x_{l+1} = x_l + F(x_l, W_l)$$

where xl+1 and xl are input and output of the l-th unit in the network, F is a residual function and Wl are parameters of the block.

In residual networks consisted of two type of blocks:

- Basic - with two consecutive 3×3 convolutions with batch normalization and ReLU preceding convolution: conv3×3-conv3×3 as in figure 1.2
- Bottleneck - with one 3×3 convolution surrounded by dimensionality reducing and expanding 1×1 convolution layers: conv1×1-conv3×3-conv1×1 as in figure 1.2(b).
- Wide-dropout - Residual networks already have batch normalization that provides a regularization effect, however it requires heavy data augmentation, which we would like to avoid, and it’s not always possible. Add a dropout layer into each residual block between convolutions as shown in figure 1.2(d) and after ReLU to perturb batch normalization in the next residual block and prevent it from overfitting.

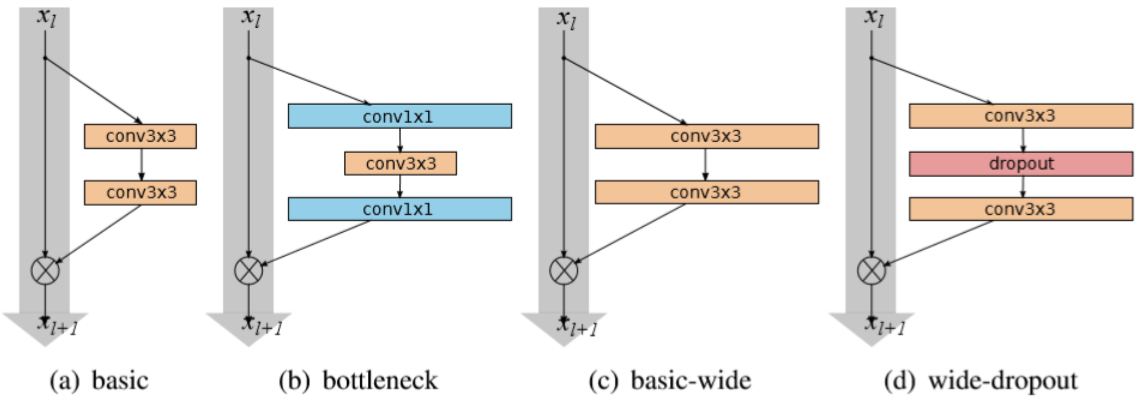

Figure 1.2 Architecture of WRN

Original architecture is equivalent to k=1. Groups of convolutions are shown in brackets where N is a number of blocks in group, down sampling performed by the first layers in groups conv3 and conv4. Final classification layer is omitted for clearance. In the particular example shown, the network uses a ResNet block of type B(3,3).

group name	output size	block type = $B(3,3)$
conv1	32×32	$[3\times3,\ 16]$
conv2	32×32	$\begin{bmatrix} 3\times3,\ 16\times k \\ 3\times3,\ 16\times k \end{bmatrix} \times N$
conv3	16×16	$\begin{bmatrix} 3\times3,\ 32\times k \\ 3\times3,\ 32\times k \end{bmatrix} \times N$
conv4	8×8	$\begin{bmatrix} 3\times3,\ 64\times k \\ 3\times3,\ 64\times k \end{bmatrix} \times N$
avg-pool	1×1	$[8 \times 8]$

Figure 1.3 Structure of residual network

The general structure of our residual networks is illustrated in figure 1.3: it consists of an initial convolutional layer conv1 that is followed by 3 groups (each of size N) of residual blocks conv2, conv3 and conv4, followed by average pooling and final classification layer. The size of conv1 is fixed in all of our experiments, while the introduced widening factor k scales the width of the residual blocks in the three groups.

Binary Classification:

The goal of a binary classification problem is to make a prediction that can be one of just two possible values.The gender prediction is a binary classification task. The model outputs value between 0~1, where the higher the value, the more confidence the model think the face is a male.The output of wide resnet is used and using binary qualification gender is recognized from the image captured. Binary Classification is implemented as shown in the diagram 1.4.

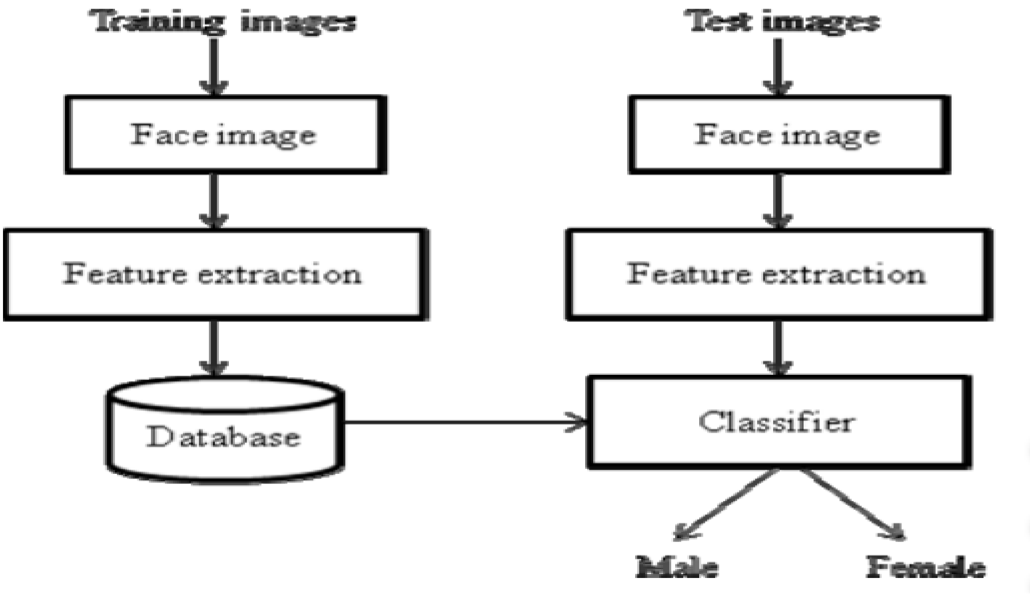

Figure 1.4 Implemtation of Binary Classification

1.4 Advantages of Proposed system

The face recognition is one of the biometric methods to identify individuals by features of the face. The biometric has a significant advantage over traditional authentication techniques as the biometric characteristics of the individual are unique for every person. A problem of personal verification and identification is an actively growing area of research. Face images are being increasingly used as additional means of authentication in applications of high security zone. Effective age group estimation using face features like texture and shape from human face image are proposed.

Deep residual networks were shown to be able to scale up to thousands of layers and still have improving performance. However, each fraction of a percent of improved accuracy costs nearly doubling the number of layers, and so training very deep residual networks has a problem of diminishing feature reuse, which makes these networks very slow to train. To tackle these problems, the architecture is used which decrease depth and increase width of residual networks. The resulting network structure is wide residual networks (WRNs).

For better performance, the geometric features of facial image like wrinkle geography, face angle, left to right eye distance, eye to nose distance, eye to chin distance and eye to lip distance are calculated. Project is implemented based on pretrained models.

Binary classifier & Wide Resnet is used for age & gender classification. In answer to this, we provide two contributions: a new and extensive data set and for the study of age and gender estimation, and a classification pipeline designed with an emphasis on making the most of what little data is available.