

Face Detection Methods: A Literature Review

Chaoran Zhu

***Abstract*— Face recognition and verification is currently an active research area. Face detection, as one of the first and most important steps of face recognition and other countless applications, has gained a significant position in today’s research area spanning several disciplines such as image processing, pattern recognition, computer vision, neural networks. In this paper I try to present an overview of face detection methods, including traditional face detector and deep learning-based face detector. I also tried to analyze the advantages and disadvantages of these various methods and their application.**

II. RELATED WORK

The process of face detection is shown in fig.1 below [2].

***Keywords***— **Face recognition, Face detection, Computer vision.**

I. INTRODUCTION

A process of detecting and locating faces from a single or series of images is called face detection [15]. Face detection is the cornerstone of all applications revolving around automatic facial image analysis including face recognition and verification, face tracking for surveillance, facial behavior analysis, facial attribute recognition, gender/age recognition [1] etc.

The goal of face detection is to determine whether or not there are any faces in the image and if the image is present then it return the image location and extent of each face [2]. While this appears as a trivial task for human beings, it is an extremely tough task for computers, and has been one of the top studied research topics in the past few decades.

In this paper, we mainly divide into two categories: traditional methods and modern deep learning methods.

The first face detection method was developed in the early seventies [3]. Since then, more and more new methods have been invented, and the accuracy and speed have also been greatly improved. Traditional methods based on so- called hand-crafted features and combined with some machine learning techniques, such as principal component analysis, linear discriminant analysis or support vector machines. However, the difficulty of engineering features that were robust to the different variations encountered in unconstrained environments made researchers focus on specialized methods for each type of variation, e.g. age- invariant methods, pose invariant methods, illumination- invariant methods etc. [4].

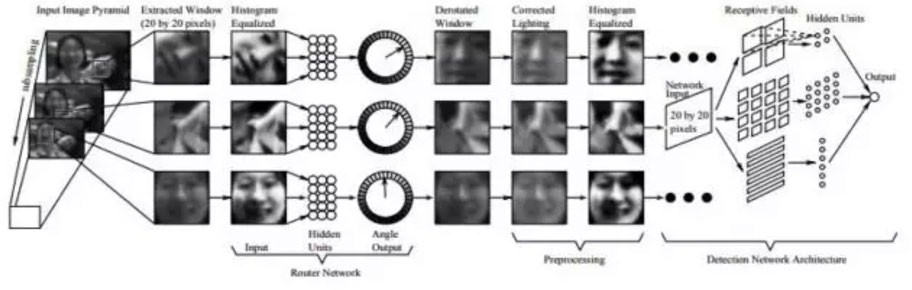
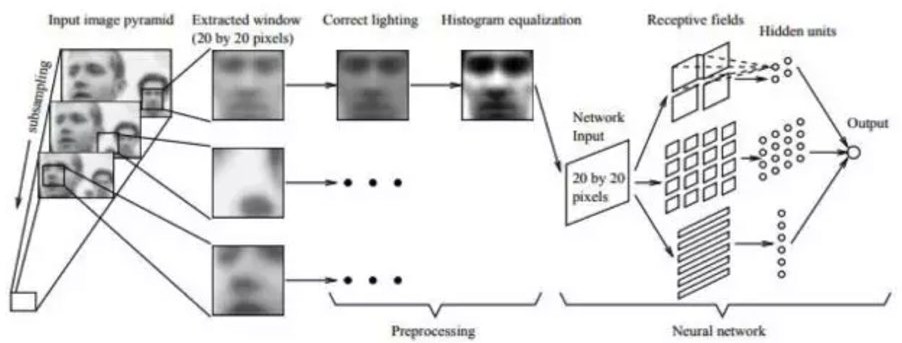
Nowadays, with the development of Internet, the accessibility of large-scale datasets and the growing speed of computer processors and graphic processors, the traditional face detectors have been superseded by deep learning methods based on convolutional neural networks (CNNs). The main advantage of deep learning methods is that they can be trained with very large datasets to learn the best features to represent the data and thus achieve very high accuracy. [4]

Face detection is basically an image segmentation problem as the image is to be segmented into two parts: one containing faces and the other representing non-face regions in images or videos [5]. First there are human faces in real world. We capture them as inputs of the system. The inputs are images or frames of a video. Then we do the pre- treatment or the pre-processing, including normalization and remove the noise. In the next step we extract the characteristics from the images or frames. The system are able to decide face and non-face area based on all the information learned from learning or so-called training process and finally output the locations. The last 3 step can be also simply named classifier and output.

There are also some evaluation methods to assess the system. First is the accuracy rate, that is, the proportion of the detected faces to the total faces (true positive). The second one is the false detection rate (also known as false alarms), that is, the number of errors (actually not faces) in the face detected by the detector (false positive). We hope that the detection rate can be as high as possible, and the false detection rate be as low as possible.

If we make false detections represented by x coordinate, and the accuracy rate represented by y coordinate. Then we

1



can draw the ROC curve. The ROC curve provides a intuitive way to compare different face detectors and is widely used.

and the model. Usually it needs a threshold to determine whether the detection area contains a human face.

Template matching has a variety of forms, the most traditional is to use a fix template. In addition, Yuille et al. [16]. proposed the deformable templates that using the parameter adjustable function. Cootes et al [17]. proposed the point distribution model (PDM) and a great idea of active shape model (ASM), which can flexibly change the shape of the template to suit the uncertainty of the target shape. Kass et al. [18]. proposed the active contour models called Snakes. This is the most common method used for boundary detection and image segmentation and can also be used for face boundary positioning.

III. TRADITONAL METHODS

Traditional method can be divided into 3 categories. They are feature-based methods and image-based methods (or called holistic methods). For feature-based methods, the most popular are skin color-based method, grayscale-based method and template-based method. They are fast and need less memory. Although their accuracy is not as high as other methods, they can be used in combination with other methods. For holistic methods, there are artificial neural net- work (ANN) method, the great V-J system and others.

A. *Skin color-based face detection Method*

Skin color-based face detection is a very popular method in face detection. Different color space has different ranges of pixels which represents skin region and non-skin region [6][8]. A color in a computer is identified by three base colors. The way computers store the color photos is using three planar matrices. The value in each of the planar matrices represents a base color value. Human skin has a characteristic color that can be distinguished from the outside world. Therefore, the colors of the skin are only represented by a certain range of the base color value. By finding the range value, the skin can be filtered out from the background. In situations where color description plays an integral role, the HSV color model [7] is often preferred over the RGB model [6].

The advantages of this method are that it is able to correctly locate all faces in the images with almost at right scale. And it’s more robust to noise and shape variations. Its processing is also much faster than processing other facial features. However, sometimes non-face-skin color region is also detected because many objects in the real world have skin-tone colors, such as some kinds of leather, sand, wood, fur, etc. [2].

*D. Artificial neural network.*

An early representative result is the method proposed by Rowley et al. [11] [12]. They used neural networks for face detection and trained a multi-layer perceptron model with 20x20 faces and non-face images. The methods are shown in Fig.2 and Fig.3 below. The method of [11] is used for front face detection.

**Fig. 2**

The method of [12] solves the multi-angle face detection problem. The whole system consists of two neural networks. The first network is used to estimate the angle of the face, and the second is used to determine whether it is a human face.

B. *Grayscale Method*

Grayscale is the brightness information of a face image. The grayscale features include organ characteristics (profile of the face, symmetry, etc.), grayscale distribution features (histograms, etc.). Although the face varies greatly in appearance, it follows certain rules, such as symmetry and the position distribution of the five senses. Grayscale method determines faces mainly by finding an image block that satisfies these rules or by finding an outline of the face by edge detection. In fact, grayscale method is a quite old method that it is hardly used alone. But the ideas have been widely used in face detection methods.

**Fig. 3**

E. *Viola Jones Face Detection system*

In 2001 Viola and Jones proposed a face detection algorithm [9]. It uses a simple Haar-like feature and a cascaded AdaBoost classifier to construct the detector. This face detection framework is capable of processing images extremely rapidly and achieving high detection rate. [2].

There are 4 very important parts in the system.

1. *Harr-like feature*

The Haar-like feature here is defined as the difference between the sum of the pixels of adjacent rectangular regions in the image. The following

C. *Template-based Method*

This method firstly establishes a standard target model and then calculate the similarity between the detection area

2

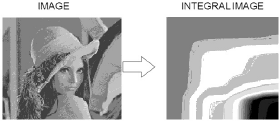
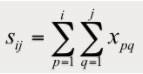
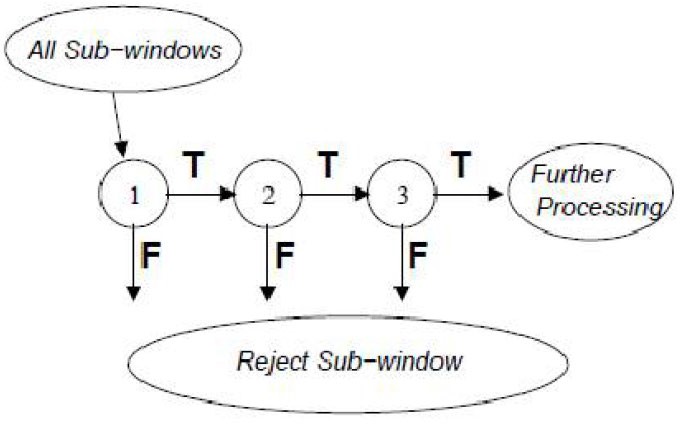


figure is a schematic diagram of the basic Haar-like features:

determine whether the sub-image in the window is a human face. This is called sliding window. The idea of cascading AdaBoost classifier is to use a plurality of AdaBoost classifiers to form a pipeline on this classification as shown in the Fig.5 below [2].

The Haar-like feature is the sum of the pixel values in the white rectangle, minus the sum of the pixel values in the black area. [10]. Such features capture information such as edges, changes and various features describe image change information in various directions. The facial features of the face have their own brightness information, which is consistent with the characteristics of Haar-like features.

*Integral image*

It is a new representation of an image allows the features used by face detector to be computed rapidly. Once integral image is computed, Harr-like features can be computed at any scale or location in constant [2]. Assume that there is an image whose pixel value at the j column and i row is xij, then the integral graph is defined as:

*2)*

**Fig. 5**

The essence of this idea is to quickly eliminate a large number of non-face windows with a simple classifier since that the face window only accounts for a very small part of all images. It helps increase the speed and achieve low false positive rates.

The advantages of Viola Jones Face Detection system are rapid image processing rate, high accuracy and very low false positive rate. However, it’s less effective on detecting non-frontal faces [2].

F. *Other methods*

One of the most popular approaches in this category is based on PCA. The idea, first proposed in[13],[14], is to apply PCA to a set of training face images to find the eigenvectors that account for the most variance in the data distribution. In this context, the eigenvectors are typically called eigenfaces due to their resemblance to real faces, as shown in Fig.6[4]. New faces can be projected onto the subspace spanned by the eigenfaces to obtain the weights of the linear combination of eigenfaces.

The Fig.5 below represented the conversion of image into an integral image [2].

**Fig.4.conversion of image into an integral image**

*Adaboost Algorithm*

The essence of the AdaBoost algorithm is a random forest consisting of multiple decision trees of depth 1 (also called decision tree stubs) using boosting method. Aim of this algorithm is to create strong classifier from linear combination of weak classifier For this, we use Adaboost Algorithm to select principal features and to train classifiers that would be using them [11].

*Cascading*

Since faces may appear anywhere in the image, the image is scanned from top to bottom and left to right with a fixed-size window during detection to

*3)*

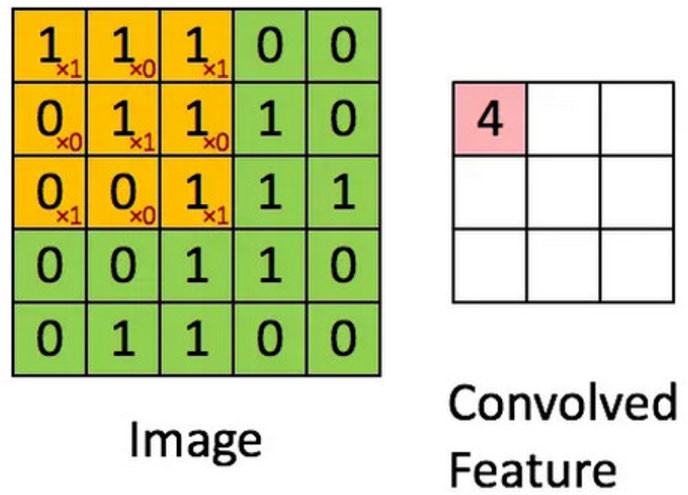
**Fig. 6: example of eigenfaces**

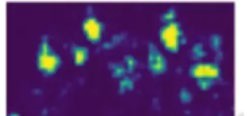
Support vector machines (SVMs) have also been used as holistic methods for face recognition. It projects the pattern into the higher-dimensional array space, and then form a decision surface between the face and the non-face. The segmentation boundary of this decision surface must be maximized.

There are also other import methods like LDA, FA, HMM and Bayes decision and so on. It’s impossible to discuss all the traditional methods here. Here I just discuss the methods that made a difference or is regraded as a milestone.

*4)*

3





kernels that occupy only tens or hundreds of pixels []. That help us reduce the memory requirements and improve the speed as well as efficiency. Other important ideas that help to improve a machine learning system are parameter sharing and equivariant representations. The parameter sharing used by the convolution operation means that rather than learning a separate set of parameters. Convolution is thus more eﬃcient in terms of the memory requirements and statistical eﬃciency.

V.DEEP LEARNING BASED METHODS

Most traditional methods aimed to address one aspect of unconstrained facial changes only, such as lighting, pose, expression, or disguise. There was no any integrated technique to address these unconstrained challenges integrally. By continuous efforts of more than a decade, “shallow” methods only improved the accuracy to about 95%, which indicates that “shallow” methods are insufficient to extract stable identity feature invariant to real-world changes. Due to this technical insufficiency, facial recognition systems were often reported with unstable performance or failures with countless false alarms in real- world applications .

But all that changed in 2012 when AlexNet won the ImageNet competition by a large margin using a technique called deep learning. Deep learning methods, such as convolutional neural networks, use a cascade of multiple layers of processing units for feature extraction and transformation. The main advantage of deep learning methods is that they can be trained with large amounts of data to learn a face representation that is robust to the variations present in the training data. Perhaps in the field of computer vision, one of the most successful deep - learning based neural networks is the Convolutional Neural Network (CNN), which use the convolution operations. And there are also many other intelligent methods derived from the CNN.

A. *CNN*

Convolutional networks, also known as convolutional neural networks or CNNs, are a specialized kind of neural network for processing data that has a known, grid-like topology. Convolutional networks are simply neural networks that use convolution, a specialized kind of mathematical linear operation, in at least one of their layers. It’s mentioned above that some traditional method uses the sliding window to scan the image. It usually costs much time. In CNN, we can use a special mathematic method of convolution, called the kernel function, to replace the sliding window and gain a great efficacy improvement. As

shown in Fig.7, it acts like a sliding window but can

A. *Other methods*

Many other deep learning-based methods are derived from CNN, and the famous ones are R-CNN and Cascade CNN.

R-CNN, where R corresponds to “Region”, meaning that CNN uses image area as input. It was first proposed in the end of 2013 and then inspired a lot of followers. R-CNN brings three major changes. Firstly, R-CNN “generates” windows to guess where there may be faces based on some characteristics of the image instead of “scan and detect”. Secondly, R-CNN can “learn” features automatically by itself. Thirdly, R-CNN use a new step in the detection process: bounding box regressions. The difference is that the regression model outputs not a discrete value, but a continuous value. It means predicting the position and size of the real suitable frame of faces. On the one hand, the bounding box regression provides a new perspective to define the detection task, and on the other hand, it has a significant effect on improving the accuracy of the detection results. Based on R-CNN, there are Fast R-CNN and Faster R-CNN and, as we can see from the name, they both improve the speed and the efficiency of the R-CNN.

Cascade CNN [19] is another kind of CNN-based method. It can be considered as a combination of traditional technology and deep learning. Like the VJ face detector, it contains multiple classifiers, which are organized in a cascade structure. Cascade CNN uses a convolutional network as a classifier for each level.

The convolutional layer is an essential feature of CNN. However, CNN usually includes not only the convolutional layer, but also the fully connected layer. A CNN that does not contain a fully connected layer is called Full Convolutional Network (FCN). Dense Box is a typical detector based on full convolutional network. We can consider the output of the FCN as a heat map, using heat to indicate the location of the faces area as shown in Fig.8 below. This can also be considered as categorizing every pixel on the image.

**Fig. 7**

achieve a very important feature, the sparse interactions. This is accomplished by making the kernel much fewer than the input. For example, when processing an image, the input image might have thousands or millions of pixels, but we can detect small, meaningful features such as edges with

**Fig. 8**

4

1. L. Sirovich and M. Kirby, “Low-dimensional procedure for the char-acterization of human faces,” Josa a, vol. 4, no. 3, pp. 519– 524, 1987.
2. M. Kirby and L. Sirovich, “Application of the karhunen-loeve proce-dure for the characterization of human faces,” IEEE Transactions on Pattern analysis and Machine intelligence, vol. 12, no. 1, pp. 103–108, 1990.
3. Muhammad Sharif, Farah Naz, Mussarat Yasmin, Muhammad Alyas Shahid and Amjad Rehman, “Face Recognition: A Survey”, Journal of Engineering Science and Technology Review 10 (2) (2017) 166- 177.
4. Yuille A L, Hallinan P W, Cohen D S. “Feature extraction from faces using deformable template” International Journal of Computer Vision, 1992, 8(2), 99-111.
5. Cootes T F, Edwards G J , Taylor C J. “Active appearance models”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(5), 681-785
6. Kass M, Witkin A, Terzopoules D. “Snakes: active contour models”, Proceedings of the IEEE Computer Society Conference on Computer Vision, 1987, 1, 239-268
7. Haoxiang Li, Zhe Lin, Xiaohui Shen, Jonathan Brandt, Gang

VI. CONCLUSION

On referring various methods, we come to understand that traditional methods are fast and require less memory and deep learning-based methods can achieve a high accuracy, a better robust and a larger range of applications. Traditional methods usually proposed early and the deep learning-based methods are more advanced. Different traditional methods are often used to solve different types of face detection problems while deep learning-based methods are more adaptable to changes in the real world. In the future, a popular research area may be the combination of traditional methods and deep learning.

REFERENCES

1. M.H. Yang, D. J. Kriegman, N. Ahuja, “Detecting faces in images: A survey”, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 24 (1), pp.34-58,2002.
2. Dr.V.K.Kadam, Deepali G.Ganakwar, “Face Detection: A Literature Review”, International Journal of Innovative Research in Science, Vol. 6, Issue 7, July 2017.
3. T.Kanade, “Picture processing by computer complex and recognition of human faces,” PhD Thesis, Kyoto University, 1973.
4. D.S.Trigueros, L.Meng, “Face Recognition: From Traditional to Deep Learning Methods” , arXiv.org, arXiv.1811.00116v1,31 Oct 2018.
5. Dr.V.K.Kadam, Deepali G.Ganakwar, “Face Detection: A Literature Review”, International Journal of Innovative Research in Science, Vol. 6, Issue 7, July 2017.

Hua. “A convolutional neural network cascade for detection.” 2015, computer vision and pattern recognition

face

[6]

Vandana S. BhatandJagadeesh D. Pujari, “Face detection system using HSV color model and morphing operations “,International Journal of Current Engineering and Technology,Proceedings of National Conference on ‘Women in Science & Engineering, Special issue.1, pp.200-204,2013.

S.Chitra, G.Balakrishnan, “Comparative Study for Two Color Spaces HSCbCr and YCbCr in Skin Color Detection”, Applied Mathematical Sciences, Vol.6, no.85, pp.4229 - 4238, 2012.

Devendra Singh Raghuvanshi,Dheeraj Agrawal, “Human Face Detection by using Skin Color Segmentation, Face Features and Regions Properties”, International Journal of Computer Applications, Vol. 38, No.9,pp.14-17,2012.

P.Viola and M.Jones. “Rapid object detection using a boosted cascade of simple features.” In Proceedings IEEE Conf. on Computer Vision and Pattern Recognition, 2001

[7]

[8]

[9]

1. Y. Amit, D. Geman, and K. Wilder,“Joint induction of shape features and tree classifiers”,IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19 ,Issue 11,pp.1300- 1305,1997.
2. Henry A Rowley, Shumeet Baluja, Takeo Kanade. “Neural network-based face detection.” 1998, IEEE Transactions on Pattern Analysis and Machine Intelligence.
3. Henry A Rowley, Shumeet Baluja, Takeo Kanade. “Rotation invariant neural network-based face detection.” 1998, computer vision and pattern recognition.

5