Journal of Multimedia Information System VOL. 5, NO. 1, March 2018 (pp. 15-20): ISSN 2383-7632(Online)
http://dx.doi.org/10.9717/JMIS.2018.5.1.15

Triangle Method for Fast Face Detection on the Wild

Karimov Madjit Malikovich[1], Tashev Komil Akhmatovich[2], Islomov Shahboz Zokir ugli[3*], Mavlonov Obid Nizomovich[4]

Abstract

There are a lot of problems in the face detection area. One of them is detecting faces by facial features and reducing number of the false negatives and positions. This paper is directed to solve this problem by the proposed triangle method. Also, this paper explans cascades, Haar-like features, AdaBoost, HOG. We propose a scheme using 12-net, 24-net, 48-net to scan images and improve efficiency. Using triangle method for frontal pose, B and B1 methods for other poses in neural networks are proposed.

Key Words: FPS, Facial, Haar-Like, HOG, AdaBoost, Cascade, Convolutional, Landmark, Capsular, Triangle Method, Pose, Feature, Neural Network.

I. INTRODUCTION

Face detection is one of main field of computer vision such as object detection. Face detection is finding of known or unknown faces from video frame, online tracking video or image. The solution to the problem involves segmentation, extraction, and verification of faces and possibly facial features from an uncontrolled background. Also there are a lot of image preprocessing application areas, such as content-based image retrieval, video coding, video conferencing, crowd surveillance, object detection, and intelligent human–computer interfaces. Face is a dynamic object in video like other objects and it can change from current pixel to other. There for face detection from video or online camera is difficult, also it takes a lot of times and processing capabilities. There are more than 24 frames (images) per second (FPS) and researchers don't take all frames for training. For example a video file consists of 24 frames per second and we can choose three frame and train.

The main functions of face detection is to determine:

- whether human faces appear in a given image;
- where these faces are located at.

The expected outputs of this step are patches containing each face in the input image.

Sufficient face detection procedure helps to take good results in face detection. How long of face detection error rates are caused to reduce face recognition rate.

II. RELATED WORK

In the survey written by Yang et al. [1], face detection algorithms are classified into four categories: know-ledge-based, feature invariant, template matching, and the appearance-based method. Every method has their advantages and disadvantages, and we use know-ledge-based method for our researches. Because, by using this method we don't need saving original image and more memory. We analyze only face features.

These knowledge-based methods use and encode human knowledge such as typical face has what kind of features (eyes, nose, mouth, eyebrow, chick and others). Usually, the rules capture the relationships between facial features. These methods are designed mainly for face localization, which aims to determine the image position of a single face [2].

In knowledge-based methods, detection procedure classifies images based on the value of simple features.

Manuscript received: February 22, 2018; Revised March 07, 2018; Accepted March 18, 2018. (ID No. JMIS-2018-0008)

Corresponding Author (*): Islomov Shahboz Zokir ugli, Phd student at Tashkent university of information technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan, Tel: +998977049555, shaxboz4044@gmail.com

[1] Professor, Director of Government testing center of the Republic of Uzbekistan, Tashkent, Uzbekistan, info@dtm.uz

[2] Professor, Dean of the Computer engineering faculty of Tashkent university of information technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan, k.tashev@tuit.uz

[3] Ph. D. student Tashkent university of information technologies named after Muhammad al-Khwarizmi, Amir Temur 108, Tashkent, Uzbekistan, shaxboz4044@gmail.com

[4] Teacher, Samarkand branch of Tashkent university of information technologies named after Muhammad al-Khwarizmi, Samarkand, Uzbekistan, obit.mavlonov@mail.ru

There are many approaches for using features rather than the pixels directly. The most common reason is that features can act to encode ad-hoc domain knowledge that is difficult to learn using a finite quantity of training image-window. For this system, there is also a second critical motivation for features: the feature-based system operates much faster than a pixel-based system. Therefor researchers in this area advise using any frames per second for improving efficiency and faster detection. Face detection systems use cascade, Haar-Like features, Histogram Oriented Gradient (HOG) and AdaBoost.

The cascade face detector proposed by Viola and Jones [3] utilizes Haar-Like features and AdaBoost to train cascaded classifiers, which achieves good performance with real-time efficiency.

2.1. Cascades

For detection of face, facial features are used three kinds of features. The value of a two-rectangle feature is the difference between the sum of the pixels within two rectangular regions. The regions have the same size and shape and are horizontally or vertically adjacent (see Figure. 1). A three-rectangle feature computes the sum within two outside rectangles subtracted from the sum in a center rectangle. Finally a four-rectangle feature computes the difference between diagonal pairs of rectangles. All features must be gray color and rectangle features can be changed.

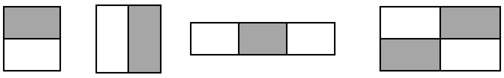


Fig. 1. Rectangle features (two, three and four rectangle features).

Image is divided into 12x12, 24x24, 48x48 windows by 12-net, 24-net, 48-net and compare based on features. Rectangle features can be computed very rapidly using an intermediate representation for the image which we call the integral image. The integral image at location x, y contains the sum [4] of the pixels above and to the left of x, y, inclusive:

$$ii(x,y) = \sum_{x' \leq x, y' \leq y} i(x',y') \ (1)$$

where $ii(x,y)$ is the integral image and $i(x,y)$ is the original image. Using the following pair of recurrences:

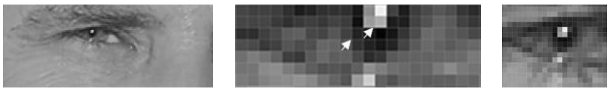
$$s(x,y) = s\big(x, y-1+i(x,y)\big) \ (2)$$

$$ii(x,y) = ii(x-1,y) + s(x,y) \ \ (3)$$

where $s(x,y)$ is the cumulative row sum.
Once computed, any one of these Haar-like features can be computed at any scale or location in constant time. In the domain of face detection it is possible to achieve fewer than 1% false negatives and 40% false positives [5] using a classifier constructed from two Haar-like features. As a result each stage of the boosting process, which selects a new weak classifier, can be viewed as a feature selection process. Our first task is minimizing false negative rates and we triangle method using capsule and convolutional neuron networks (section proposed method).

2.2. Histogram oriented gradient (HOG)

By using HOG [6], we can find face features easily. To find faces in an image, we'll start by making our image gray (value of gray color 0-255) because we don't need color data to find faces. Then we'll look at every single pixel in given image one at a time. For every single pixel, we want to look at the pixels that directly surrounding it. Our goal is to figure out how dark the current pixel is compared to the pixels directly surrounding it. Then we want to draw an arrow showing in which direction the image is getting darker. If we repeat that process for every single pixel in the image, we end up with every pixel being replaced by an arrow. These arrows are called gradients and they show the flow from light to dark across the entire image (see Figure. 2).



Eye in gray

Detection darker direction

Eye in HOG

Fig. 2. Detection eye by HOG.

This might seem like a random thing to do, but there's a really good reason for replacing the pixels with gradients. If we analyze pixels directly, current dark images and current light images of the same person will have totally different pixel values. But by only considering the direction that brightness changes, both really dark images and really bright images will end up with the same exact representation.

2.3. AdaBoost

For faster and efficiency, detection and boosting the classification performance of a simple learning algorithm is used AdaBoost learning algorithm [3]. AdaBoost takes weak classification functions from stronger, because strong features are used for detection easily, but we need to work with weak features. For learning weak functions we use capsule neuron networks in our triangle method.

The conventional AdaBoost procedure can be easily interpreted as a greedy feature selection process. Input part has several values and it helps to separate weak and strong features.

2.4. Convolutional Neuron Network (CNN)

For construction strongly and efficiency face detection system, Li and others [7] proposed using cascaded convolutional neuron network (CNN). Face detection by cascaded CNNs requires bounding box calibration from face detection with extra computational expense and ignores the inherent correlation between facial landmarks localization and bounding box regression. We know that in one frame will be one or a lot of faces. Also, by scanning a window we can see that there are strong and weak face windows. Zhang and others [8] used multi-task CNN to improve the accuracy of multi-view face detection, but the detection recall is limited by the initial detection window produced by a weak face detector. By multi-task cascaded face detection uses fourteen different face features such as a Figure 3. Also, Jeffry Hinton has introduced a newer, capsular neuron network [9] identification system, which has a more detailed face identification approach. By Hinton's capsules we can detect and find facial features from low resolution images.

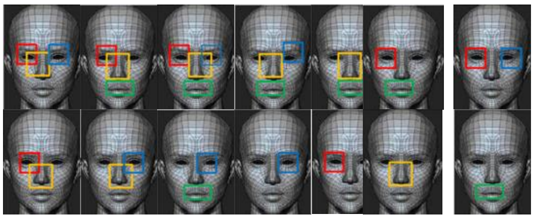


Fig. 3. Different face characters.

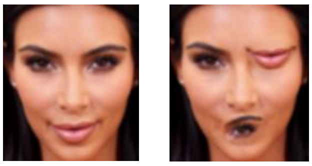
Zhang's method proposed a new framework to integrate these tasks using unified cascaded CNNs by multi-task learning. This method consists of three stages. In the first stage, it produces candidate windows quickly through a shallow CNN. Then, it refines the windows by rejecting a large number of non-faces windows through a more complex CNN. Finally, it uses a more powerful CNN to refine the result again and output five facial landmarks positions.

If we return to the neuron networks it studies every point of research area weak or strong results of function. Using neuron networks for face detection is realized problems with pose, expression, and lighting. For protection from these problems are used CNNs and it gives high quality performance. This CNN cascade operates at multiple resolutions, quickly rejects the background regions in the fast low resolution stages, and carefully evaluates a small number of challenging candidates in the last high resolution stage. Also, CNN is used to improve localization effectiveness, and reduce the number of candidates at later stages. Kaipeng Zhang and others said that deep CNNs [10] achieve substantial improvements and increase of detection rate in face identification in the wild. Classical CNN-based face detection methods simply stack successive layers of filters where an input sample should pass through all layers before reaching a face/non-face decision. Inspired by the fact that for face detection, filters in deeper layers can discriminate between difficult face/nonface samples contextual CNN (see Figure.3). In deep learning CNN differ with traditional CNN which enables different layers to be trained by different types of samples, and can focus on handling more difficult samples. Also, Kaipeng Zhang and his commands used Body Part Sensitive Learning.

But, Jeffry Hinton developer of "Deep learning" said that there is a problem with CNN which it works by accumulating a variety of features at each level. It begins with finding edges, then shapes, and then actual objects. However, information about the spatial relationships of all functions is lost. We can explain the principle of the CNN in the following way:

```
if (2 eyes && 1 nose && 1 mouth) {
It's a face!
}
```

It is not true, because in this code doesn't attend to the real face rules (see Figure. 4). In Figure 4.a is shown real face, but in Figure 4.b not face. Our function is solving this problem. For solving this problem, we propose using triangle method for detection faces by facial features easily, faster and without negatives.



a. Face b. Not face

Fig. 4. Face or not face pose.

2.5. Proposed Method

Ref. [11] is given recommendations and optimizing ways of collecting face databases for face detection and recognition. By these databases, we are trained faces and face features. For searching facial features, we scan the image or video frame by 12, 24, 48-net (see Figure. 5) [12] and send to the neuron network (see Figure. 7).

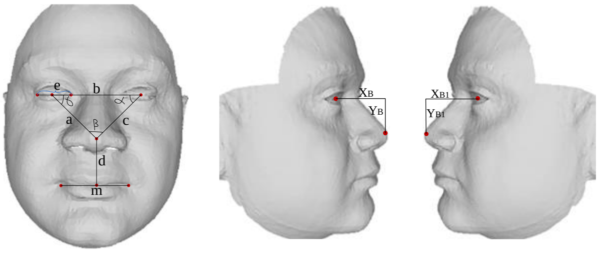
12-net. 12-net refers to the first CNN in the test pipeline. 12-net is a very shallow binary classification NN to quickly scan the testing image. Densely scanning an image of size WxH with 4-pixel spacing for 12x12 detection windows is equivalent to apply the 12-net to the whole image to obtain a $\left(\left[\frac{W-12}{4}\right]+1\right)*\left(\left[\frac{H-12}{4}\right]+1\right)$ map of confidence scores. Each point on the confidence map refers to a 12x12 detection window on the testing image.

24-net. 24-net is an intermediate binary classification CNN to further reduce the number of detection windows. Remaining detection windows from the 12-net are cropped out and resized into 24x24 images and evaluated by the 24-net. A similar shallow structure is chosen for time efficiency. With this multi-resolution structure, the 24-net is supplemented by the information at 12x12 resolution which helps detect the small faces.

48-net. 48-net is the last binary classification CNN. At this stage of the cascade, it is feasible to apply a more powerful but slower CNN. Similar to the 24-net, we adopt the multi-resolution design in 48-net with additional input copy in 24x24 and a sub-structure the same as the 24-net. Or we can window size as a 48x48.

If we analyze all of the nets, 12-net gives best result for detect facial features but it detects slowly.

Fig. 5. Scan image by 12, 24, 48-net window.



A. Front B. Right side B1. Left side

Fig. 6. Face poses and features.

Triangle method. Finding faces by facial features is big problem. By our triangle method, we can solve this problem. Triangle method is used to sort facial and no facial images by calculating dependencies between key-points. This method consists of following steps (see Figure. 6.A):

1. Calculating distance between facial features: a, b, c and e. Here, a, b, c distance small or equal to the three times size of the eye (a, b, c <=3*e).
2. Calculating corners: α, β, γ. Here, these corners small than 90^0.

$$\alpha = \arccos(\frac{c^2+b^2-a^2}{2cb}) \quad (4)$$

$$\beta = \arccos(\frac{a^2+b^2-c^2}{2ab}) \quad (5)$$

$$\gamma = \arccos(\frac{a^2+b^2-c^2}{2ab}) \quad (6)$$

3. Y coordinate of eyes differ each other maximal size of eye (e).
4. Y coordinate of the nose is small than eyes.

If face pose is left side or right side (see Figure. 6.B and 6.B.1), than we can implement fallowing methods:

Here, m=size of the mouth,

e – size of the eye,
a – distance between left eye and nose,
b – distance between two eyes,
c – distance between right eye and nose,
d – distance between nose and mouth,
α, β, γ – corners (see Figure. 6.A).

2.5.1. B method

Coordinates of right eye and nose differ to the X_B, Y_B and y coordinate of the nose small than eye. Here $X_B, Y_B \leq 3 * e$. And finishing coordinates of the eye and nose must be opposite.

2.5.2. B1 method

Coordinates of left eye and nose differ to X_{B1}, Y_{B1} and y coordinate of the nose small than eye. Here, $X_{B1}, Y_{B1} \leq 3 * e$. And finishing coordinates of the eye and nose must be opposite.

Our neuron network consists of following spteps:

Step 1. We have detected Left eye, Right eye, Nose, Mouth and we can implement triangle method easily by left and right eyes, nose.

Step 2. We have detected Left eye, Right eye, Nose and we can implement triangle method easily by these features.

Step 3. We have detected Nose and Mouth. Distance between nose and mouth must be small than size of the mouth (d<m). Also, y coordinate of the mouth small then nose.

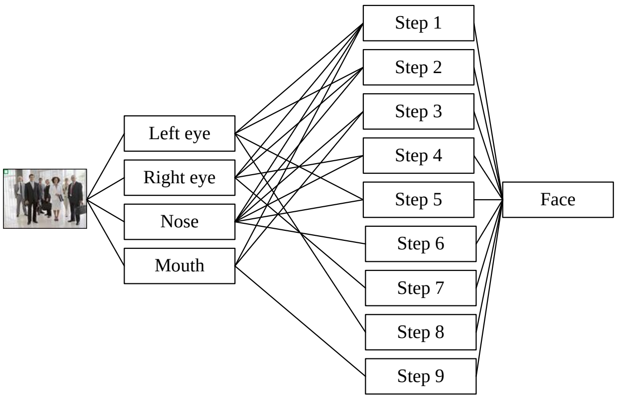


Fig. 7. Neuron network.

Step 4. We have detected Right eye and Nose. In this state, first we try to find left eye by triangle method. If result is not successful then we implement B method.

Step 5. We have detected Left eye and Nose. In this state, first we try to find right eye by triangle method. If result is not successful then we implement B1 method.

Step 6. We have detected Nose. We try to find left and right eyes and implement triangle method. If this method is not successful then we try to implement methods of Step 4 and Step 5.

Step 7. We have detected Right eye. In this state we try to find left eye and nose and implement triangle method. If this method is not successful then we try to implement B method.

Step 8. We have detected Left eye. In this state we try to find right eye and nose and implement triangle method. If this method is not successful then we try to implement method B.1.

Step 9. We have detected Mouth. In this state, we try to find noise and implement methods of Step 3.

III. DISCUSSION

We detected 640x480 size image using 12-net on the 4Gb RAM and simple CPU, it takes 36ms to detect one frame. If we use 2Gb GPU than detection time reduce until 10ms.

In real world, like Figure 7, detection system based on triangle method detects faces 93.8% accuracy and faster then Sparse coding and Adaboost LDA algorithms.

IV. CONCLUSION

Face detection is main part of the computer vision in terms of face recognition. There are several methods of the face detection methods, such as, know-ledge-based, feature invariant, template matching, and the appearance-based method. In this paper, we selected feature based face detection methods, because finding facial features is easy using 12, 24, 48-nets. In this case, main problem is to find faces and their localizing. We proposed triangle method to detect legal faces and reduce number of false negatives and false positives. Also, this method was implemented into CNN and studied each position of the faces by neurons (9 steps). Also, we recommend using 12-nets with high speed computers (supporting five or more frames per second).

We will be implement our triangle method to recognize humans and it will take sufficient result. Because, face detection is one of the main parts of face recognition and after that we can find all faces from image. Also, we can reduce number of false negatives.

References

[1] M. H. Yang, D. J. Kriegman, and N. Ahuja, “Detecting face in images: a survey,” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 24, pp. 34–58, 2002.

[2] Chao W. L, “Face Recognition,” GICE, National Taiwan University, 2007.

[3] P. Viola and M. J. Jones, “Robust real-time face detection,” International journal of computer vision, Vol. 57, No. 2, pp. 137-154, 2004.

[4] Crow F. C, “Summed-area tables for texture mapping,” ACM SIGGRAPH computer graphics, No.3 pp. 207-212, ACM, 1984.

[5] Malikovich, K. M., Axmatovich, T. K., Zokirugli, I. S., and Zarif, K. “Minimizing in Face Recognition Errors and Preprocessing Time,” In Proceedings of International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE), pp. 212, 2014.

[6] Dalal N., Triggs B, “Histograms of oriented gradients for human detection,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, pp.886-893, 2005.

[7] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, “A convolutional neural network cascade for face detection,” in IEEE Conference on Computer Vision and Pattern Recognition, pp. 5325-5334, 2015.

[8] C. Zhang, and Z. Zhang, “Improving multiview face detection with multi-task deep convolutional neural networks,” IEEE Winter Conference on Applications of Computer Vision, pp. 1036-1041, 2014.

[9] Sabour S., Frosst N., Hinton G. E. “Dynamic routing between capsules,” Advances in Neural Information Processing Systems, pp. 3859-3869, 2017.

[10] Zhang K et al. “Detecting Faces Using Inside Cascaded Contextual CNN,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3171-3179, 2017.

[11] Karimov M. M. Islomov Sh. Z. “Optimising And Recommendations For Collecting Face Databases,” International Journal of Research in Engineering and Science. pp. 2320-9364.

[12] H. Li, Z. Lin, X. Shen, J. Brandt, G. Hua, “A convolutional neural network cascade for face detection,” in IEEE Conference on Computer Vision and Pattern Recognition, pp. 5325-5334, 2015.

Authors

Karimov Madjid Malikovich received his MS degree in the Department of Computer systems from Tashkent state technical university, Uzbekistan, in 1979. He received PhD degree in 1995. After than he also received Doctor of Science in technical sphere, in 2005. His research interests include information security, protecting techniques, techniques of image recognition.

Tashev Komil Akhmatovich received his MS degree in the Department of Computer systems from Tashkent state technical university, Uzbekistan, in 2005. He received PhD degree in 2010. His research interests include data protecting techniques, monitoring, image recognition.

Islomov Shahboz Zokir ugli received his MS degree in the Department of Information security from Tashkent university of information technologies, Uzbekistan, in 2014. He is PhD student. His research interests include data protecting techniques image detection, face recognition, access control.

Mavlonov Obid Nizomovich received his MS degree in the Department of Information security from Tashkent university of information technologies, Uzbekistan, in 2014. He is PhD student. His research interests include data protecting techniques, VPN.