Face detection accuracy study based on race and gender factor using Haar cascades

Elizaveta Rudinskaya
Department of Technical Cybernetics
Samara National Research University
Samara, Russia
ea.rud@yandex.ru

Rustam Paringer
Department of Technical Cybernetics
Samara National Research University
Samara, Russia
RusParinger@ssau.ru

***Abstract*—Object recognition on images is very often used in modern life. Many factors greatly affect the accuracy of detection and recognition. Belonging to a particular gender or race is one of them. This article develops a methodology that allows evaluating algorithm for this identification, as well as evaluating the impact of the training sample on the detection result. The results stability when detecting Haar cascades to a data set of human biological characteristics was considered. Special attention was paid to finding the necessary stages of the methodology, searching for the values of the algorithm parameters, selecting a training sample, and obtaining results with an optimal detection accuracy. The most detected class by Haar cascades was European Man with a score of 11/12 The best detection result was shown by the cascades "frontalface_alt", "frontalface_alt2" and "eye_tree_eyeglasses" with an accuracy of 5/6.**

Keywords—cascade Haar, image detection, cascade classifiers, computer vision (key words)

I. INTRODUCTION.

At present, due to the rapid development of various surveillance systems the task of detecting objects, faces, in particular, in images or videos (for example, in security tracking systems) is of great current interest [1, 2, 3]. A huge contribution in this area was made by the Viola-Jones method – a face detector that can find faces in real-time with high accuracy [4, 5].

The cascades are a family of universal algorithms that can be trained to detect any object in an image if the necessary data set is available. Taking into account modern capabilities and new methods, the task of detecting faces using machine learning methods (algorithms) becomes even more attractive and feasible. This is why the field of face detection is promising and in demand. In particular, it is used to find a person and track them [6]. Facebook also uses a detection algorithm to detect faces in photos and recognize them.

Information about the number of people is required in many access control systems of various types of institutions, such as airports, metro stations, in institutions that automatically record the number of visitors and many other observation systems [7, 8]. It is a growing number of particular methods for solving different problems [9, 10]. Detection is also used as the first step in solving the problem of facial recognition in images [11] because it facilitates the method. Many studies are aimed at improving the algorithm, for example, to increase accuracy or performance but do not always pay enough attention to the cause of any imperfections.

You can often hear about new achievements of neural networks, new methods, and approaches for solving a particular problem. After reading the articles you can conclude that very often there is a situation of false recognition caused by incorrect detection which can potentially lead to critical consequences.

One of the first stages of developing a machine vision algorithm is the learning stage when the algorithm is trained to work and perform a task on a specific training data set. However, few people take into account that the final result will depend on the correctness of this data set. For example, a correctly or incorrectly selected data set naturally affects the behavior of the algorithm in the future. False recognition occurs due to factors that were not taken into account when preparing the training sample. These factors include different lighting, a different angle of rotation of the photo, background, fundamentally different quality of the photo, and other factors. Therefore, it is very important to provide a careful and deliberate approach to the preparation of the training data set.

The problem of the influence of training samples on algorithms and their positioning is not fully understood, is poorly disclosed and is not taken into account by anyone. So, if the algorithm for detecting faces was trained on a single data set which consisted mainly of representatives of, for example, the Caucasian race, then it would be more correct to call it the algorithm for detecting Caucasian faces. In other words, since insufficient attention is paid to the training sample, the algorithm can not be positioned as a universal one. It can only be positioned as an algorithm that identifies the main features of the training sample. This means that we can assume that a properly selected training sample is required to prevent false recognition and improve the accuracy of the selected method.

In the previous article [12], Haar cascades were combined to increase the detection accuracy and to determine more successful combinations. It was revealed that when combining cascades you need to take into account the characteristics of each of them (for example, what features are used for searching, parameter values, etc.), so that detection is more qualitative. It was also necessary to take into account the possibility of intersecting the found areas, so as not to count the same person twice. However, all these optimization options failed to achieve the results that should normally correspond to cascades. An attempt to determine the reason for this behavior led to this study. After reviewing the article [11], which described an experiment with neural networks and people of different biological characteristics (the authors used Haar cascades for initial detection of the necessary areas), it was decided to conduct a similar study but with a more fundamental method using Haar cascades intended to detect faces rather than recognize them. Quite a lot of articles cover the problem of discrepancy between the results expected and obtained after the experiment. The authors of the methods call their solutions universal and guarantee a single result. Despite this, the new results of

experiments of other researchers who applied this method later do not correspond to the previous ones. [10, 13, 14] the Relevance of the research presented in the article is to find the probable problem of inaccuracies in the solution. The fact is that the first authors call their solution" universal", although it is not always so, as a result of which there are differences of opinion. The sample plays an important role and has a significant impact on the research result [15, 16, 17]. Because of the difference between the training and test samples, there may be inconsistencies in the research results. You can't call the method universal until you have performed an independent thorough check of the method's performance on various input data, or call it not universal in the General sense, but specifying which data the solution will give such a stable and high result on. The purpose of this study was to set up an experiment on one specific face detector which would allow us to formulate a methodology for verifying the adequacy of this detector's performance.

II. METHODOLOGY

To investigate the algorithm for correct operation with various input data, you must:

1. Formulate the research issues.
2. Select the algorithm that will be used for the research.
3. Prepare the correct sample that matches the problem and determine the expected results.
4. To conduct a study of the influence of the algorithm parameters on the accuracy of detection of the prepared sample.
5. Conduct an experiment.
6. Calculate the error value to determine whether the results fit into the confidence range of values which will determine the behavior of the algorithm in this sample.
7. Process the results of the experiment.

III. HAAR CASCADE.

Since the research problem has already been determined, by the second point of the methodology Haar cascades were used for research.

Haar Cascade is a method for detecting objects in an image based on machine learning, the idea of which was proposed in an article authored by Paul Viola and Michael Jones [4, 5]. Taking an image as input, the trained Haar cascade determines whether the desired object is in it, i.e. it performs the classification task by dividing the input data into two classes (there is a desired object, there is no desired object). A properly trained Haar cascade has a good classification execution speed as well as good resistance to deviations of various kinds. [18, 19]

Main features of cascades:

Scale factor (scalefactor) is a parameter that defines the size of the image each time it is displayed. It is used to create a scale pyramid (representing images so that we can detect both small and large faces using the same detection window). Its significance determines the thoroughness of research in each area.

Sliding window size (minsize) is a parameter that defines the minimum possible size of the object. In this way, smaller areas are ignored. This size is set independently.

The parameter that defines the number of possible neighbors (minNeighbors) is responsible for the number of neighbors that each rectangle can have. This parameter affects the quality of detected faces: the higher the value is, the fewer detections there are but the quality is higher. The optimal values are 3 – 6.

IV. EXPERIMENTAL RESULTS

By the third point of the methodology, a sample of 600 different photos was compiled for research, taken from the publicly available IMDb-WIKI set of images of people [20] and classified by gender and race into 6 components. Photographs of women and men of the Negroid, Mongoloid, and Caucasian races were used. All photos were divided into 6 classes of 100 photos each depending on their biological characteristics: Mongol Woman, Mongol Man, European Woman, European Man, Negroid Woman, Negroid Man. The experiment used independent Haar cascades with two constant parameters: a sliding square window with a side length of 30 pixels and a parameter that determines the number of possible neighboring Windows equal to 5.

The fourth step was to determine the correct parameters of the algorithm: when processing results only the value that corresponds to the number of required detections shown in table 1 was considered to be the correct detection. If the result differed even by one (for example, one eye was detected instead of two), the detection was considered incorrect.

TABLE I. EXPECTED VALUES OF DETECTION RESULTS

Cascade	eye	eye_tree_eyeglasses	frontalcatface_extended	Frontalface	frontalface_alt	frontalface_alt2	frontalface_alt_tree	frontalface_default	lefteye_2splits	lowerbody	profileface	smile	righteye_2splits
Expected results recognition	2	2	1	1	1	1	1	1	1	1	1	1	1

Next – the fifth point of the methodology.

The first stage consisted of estimating the error of detecting objects in images using each of the ready-made Haar cascades for each classification group at a certain value of the scale factor, which varied between 1.01 and 1.5 with a step of 0.01.

The results obtained are shown in illustration 1with the vertical axis being the ratio of the number of photos where there was no detection to the number of all images, and the horizontal axis is the value of the scale factor.

Analyzing the results of the first experiment we can conclude that the smaller the scale factor value, the more correct detection is possible. Accordingly, the maximum number of correct detections was achieved at a scale factor value of 1.01.

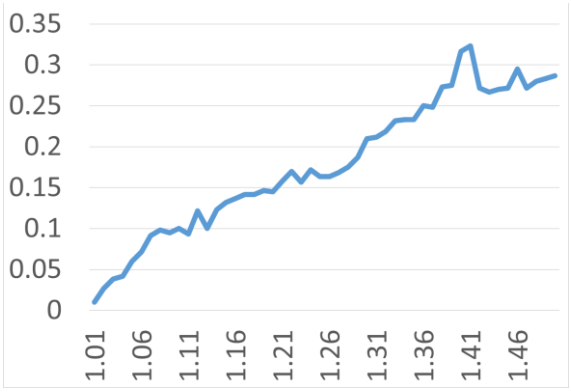

Fig. 1. Percentage of false detections depending on the scale factor value.

The second stage of the experiment was to study the accuracy of face detection for each image separately. The same values of the scale factor and the necessary values of the obtained results were used as were presented above. It was assumed that an image could be detected in cascades if at least one of the independent cascades gave the correct image detection for at least one parameter value. Table 2 below shows the results for the number of detections. The first line shows all the detection values that were obtained during the experiment, and the second line shows the number of photos that had the corresponding number of detections

Based on this table we can conclude:

- The maximum number of images was achieved when detecting faces was realized in seven different cascades, with eight cascades detection being in the second place.
- Out of 600 images only one image was not detected by any cascade. It is shown in illustration 2 below
- The maximum number of detections per photo achieved is 12 (out of a possible 13)

Fig. 2. The image was not detected by any cascade.

Then using one large table with the results, we made tables in which we counted the number of correct detections by each cascade not by individual images but by each of the six classes at a certain value of the scale factor.

A stable cascade to biological characteristics can be called one in which error value will differ no more than the standard deviation from the average value (the sixth point of the methodology).

After receiving the results, the average value of the number of correct detections for all classes was calculated using the formula

$$\bar{X} = \frac{1}{n}\sum_{j=1}^{n} x_j,$$

where $\bar{X}$ is the average, n is total number of values, x_j – each element.

Next, we determine Sigma

$$\sigma(X) = (D(X))^{1/2}, \text{where } D(X) = M(X^2) - \left(M(X)\right)^2$$

Here $D(X)$ is the variance, and $M(X)$ is the expectation.

The confidence range of values was calculated using the formula:

$$\bar{X} - \sigma(X) < \mu < \bar{X} + \sigma(X)$$

After finding it, an analysis was performed: which cascades gave a more stable result.

Table 3 shows a table with a scale factor value of – 1.01 where the number of correct detections is indicated.

- The results that correspond to the confidence range are highlighted in green
- The table illustrates that the "smile" cascade detected Negroid Man better than the other ones but its value is outside the confidence range
- The maximum values were obtained using the "frontalface_alt" and "frontalface_alt2" cascades
- Mongol Woman and Mongol Man have approximately the same detection rate.
- European Woman is in first place in terms of the average number of correct detections, followed by Mongol Woman and Mongol Man and Negroid Man is the least detectable.

The results were evaluated based on the average value of a particular class.

Based on this assessment of the results of the experiment we can assume that the cascades were trained mainly on representatives of the female Caucasian race since they were more correctly detected than the other ones. Due to their greater similarity, men and women of the Mongoloid race came in second, followed by men of the Caucasian and women of the Negroid races, and the last place was given to men of the Negroid race.

After that, an evaluation criterion was introduced for ranking individual classes and cascades. The correct result is the one that entered the confidence interval. As a result, we can obtain: in the first place for correct detection is the class of European Man with a result of 11 out of 12 possible; then Mongol Woman, Mongol Man, Negroid Woman, and European Woman having 9 out of 12, but the class Negroid Man is only 3 out of 12. Among the cascades "eye_tree_eyeglasses", "frontalface_alt" and "frontalface_alt2" performed the best with the result of 5/6, the "frontalface_default" cascade performed the least detections – 1/2 and the remaining cascades performed the same with the result of 2/3.

V. CONCLUSION

A methodology was proposed and tested that allows us to find out the universality of the chosen method, in particular, to test cascades for the stability of the results of cascades for input data with images of people of different biological characteristics. Haar cascades cannot be positioned as a universal detector since they do not detect Negroid Man

well. We can assume that this is most likely caused by a problem with the training sample, for example, an insufficient number of photos of representatives of the negroid race were taken. Therefore, it is necessary to create an appropriate training sample that will match the task and position of the developed algorithm accordingly in the future. During the experiment, it was found that the lower the value of the scale factor, the more correct the detection is. It is better to start detection at scale factor values within the range of 1.01 – 1.1. When drawing up a training sample one should approach it with maximum responsibility and provide for various options for the outcome of events. When training the face detection algorithm it is necessary to take a fairly large number of images with people of different ages, genders, races, and other biological characteristics. It is also necessary to provide pre-processing of photos so that there is no blurring or inaccuracy in the input data. It was found that the most detectable class was European Man with a result of 11/12 and "frontalface_alt", "frontalface_alt2" and "eye_tree_eyeglasses" which have an accuracy result of 5/6 are best suited for initial detection. Based on this, we can conclude that a problem that was not previously addressed in scientific publications was demonstrated by the example of Haar cascades and their stability to various validation samples. The influence of the initial data on the research result is shown; more optimal values for accurate face detection using Haar cascades are found. It is concluded that the difference between the training sample and the initial set for further research does not allow the author to call his solution universal until it is thoroughly tested on different input data.

TABLE II. THE NUMBER OF DETECTIONS IN THE IMAGES

The number of detections	0	1	2	3	4	5	6	7	8	9	10	11	12
The amount of images	1	15	23	28	63	80	83	118	103	48	27	9	2

TABLE III. RESULTS OF CORRECT DETECTION OF EACH INDEPENDENT CASCADE BY SEPARATE RACIAL CHARACTERISTICS

Name class	eye	eye_tree_ eyeglasses	frontalcatface_ extended	frontalface	frontalface_alt	frontalface_alt2	frontalface_alt _tree	frontalface_ default	lefteye_2splits	profileface	righteye_2splits	smile
Mongol Woman	16	41	35	38	81	73	49	56	22	70	22	5
Mongol Man	24	30	27	47	80	77	51	69	22	61	27	9
Negroid Woman	29	42	23	29	81	78	43	68	19	55	26	10
Negroid Man	23	14	16	20	62	59	22	53	24	58	19	12
European Woman	18	40	38	41	87	79	63	58	18	63	25	7
European Man	24	32	32	38	81	77	43	60	22	59	20	2

ACKNOWLEDGMENT

The research was supported by the Ministry of Science and Higher Education of the Russian Federation (Grant # 0777-2020-0017) and partially funded by RFBR, project number # 19-29-01135.

REFERENCES

[1] L. Sirovich and M. Kirby, “Low dimensional procedure for characterization of human faces,” Journal of the Optical Society of America A, vol. 4, pp. 519,1987.

[2] M. Turk and A. Pentland, “Face recognition using eigenfaces,” Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 586-591, 1991.

[3] W. Zhao, R. Chellappa, A. Rosenfeld and P.J. Phillips, “Face Recognition: A Literature Survey,” ACM Computing Surveys, vol. 35, no. 4, pp. 399-458, 2003. DOI: 10.1145/954339.954342.

[4] P. Viola and M.J. Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features,” Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2001.

[5] P. Viola and M.J. Jones, “Robust real-time face detection,” International Journal of Computer Vision, vol. 57, no. 2, pp. 137-154, 2004.

[6] V.V. Arlazarov, K. Bulatov, T. Chernov and V.L. Arlazarov, “MIDV-500: a dataset for identity document analysis and recognition on mobile devices in video stream,” Computer Optics, vol. 43, no. 5, pp. 818-824, 2019. DOI: 10.18287/2412-6179-2019-43-5-818-824.

[7] R.C. Gonzalez and R.E. Woods, “Digital image processing,” Technosphere, Moscow, 2005, 1072 p.

[8] D. Forsyth and Zh. Ponce, “Computer Vision: A Modern Approach,” Vil’yams, Moscow, 2004, 928 p.

[9] P.V. Bezmaternykh, D.A. Ilin and D.P. Nikolaev, “U-Net-bin: hacking the document image binarization contest,” Computer Optics, vol. 43, no. 5, pp. 825-832, 2019. DOI: 10.18287/2412-6179-2019-43-5-825-832.

[10] L. Stengcai, K. Jain, Z. Anil and L. Stan, “A Fast and Accurate Unconstrained Face Detector,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 2, pp. 211-223, 2016.

[11] M. Sarfaraz, Sh. Gupta, A. Wajid, S. Gupta and A. Musher, “Prediction of Human Ethnicity from Facial Images Using Neural Networks,” A Novel Cluster Algorithms of Analysis and Predict for Brain Derived Neurotrophic Factor (BDNF) Using Diabetes Patients, pp. 217-226, 2018.

[12] E.A Rudinskaya and R.A. Paringer, “Development of face detection algorithm using combinations of Haar cascades,” V International Conference on Information Technology and Nanotechnology (ITNT), pp. 6–12, 2019.

[13] V. Jain and E. Learned-Miller, ”FDDB: A Benchmark for face detection in unconstrained settings,” Technical Report UM-CS-2010-009, University of Massachusetts, 2010 [Online]. URL: https://people.cs.umass.edu/~elm/papers/fddb.pdf.

[14] P. Li, M.L. Prieto, P.J. Flynn and D. Mery, "Learning face similarity for re-identification from real surveillance video: A deep metric solution," Biometrics (IJCB) IEEE International Joint Conference, pp. 243-252, 2017.

[15] B.F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, M. Burge and A.K. Jain, "Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A," IEEE Conference on Computer Vision and Pattern Recognition, pp. 1931-1939, 2015.

[16] B. Yang, J. Yan, Z. Lei and S.Z. Li, "Fine-grained evaluation on face detection in the wild," IEEE International Conference on Automatic Face and Gesture Recognition, 2015. DOI: 10.1109/FG.2015.7163158.

[17] G. Guo and G. Mu, "A study of large-scale ethnicity estimation with gender and age variations," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 79-86, 2010.

[18] I.A. Kalinovskii and V.G. Spitsyn, "Review and testing testing of frontal face detectors," Computer Optics, vol. 40, no. 1, pp. 99-111, 2016. DOI: 10.18287/2412-6179-2016-40-1-99-111.

[19] A.V. Anastasov, "Using the Haar cascade to detect vehicle registration plates in an image," XII international scientific and practical conference of students, postgraduates and young scientists "Youth and modern information technologies", Tomsk, vol. 2, pp. 126-127, 2014.

[20] R. Rothe, "IMDB-WIKI – 500k+ face images with age and gender labels" [Online] URL: https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/.