Zh. Sultanov
Master's student, Faculty of Information Technologies
janibeksultanov97@gmail.com, orcid.org/0000-0003-1674-3362
L.N. Gumilyov Eurasian National University, Kazakhstan

ANALYSIS OF METHODS FOR DETECTING FACES IN AN IMAGE

Abstract: In this article, computer vision is considered as modern technology of automatic processing of graphic images, and the relationship between the terms "computer vision" and "machine vision" is investigated. A diagram of a typical computer vision system is given and the possibility of using a system based on an artificial neural network for image analysis is considered.

The article analyses the current situation with the use of computer vision systems and the possibility of its application. This article presents face recognition algorithms for existing categories, including: empirical method; feature method – invariant feature; use the template specified by the developer for identification; study the method of detecting the system by external signs.

The empirical method of "top-down knowledge-based methods" involves creating an algorithm that implements a set of rules that image segments must satisfy in order to be recognized as faces. Feature-invariant approaches (Feature-invariant approaches) based on bottom-up knowledge constitute the second group of face detection methods. The methods of this group have the ability to recognize faces in different places as an advantage. Use the template set by the developer for identification (template matching method). Templates define specific standard images of face images, for example, describing the attributes of different areas of the face and their possible mutual positions. A method for detecting faces by external signs (a method for performing the training stage of the system by processing test images). The image (or its fragments) is somehow assigned a calculated feature vector, which is used to classify the image into two categories – human face/non-human face.

Keywords: computer vision, face recognition, Kotropoulos & Pitas method.

The task of detecting faces in images is typically the initial stage in the process of tackling more difficult problems such as identifying features, facial characteristics, or facial emotions. Furthermore, information regarding the presence and quantity of individuals in the image can be used in automated visitor accounting systems, access control systems at institutions, airports, and metro stations, automated accident prevention systems, intelligent human systems, and computer systems. It is used in photography to automatically focus on people's faces and stabilize facial photos to aid with emotion identification; it is used to expand the field of stereo vision when creating a 3D display system.

The existing facial recognition algorithms can be divided into four categories:
- Empirical method;
- Method of feature-invariant feature;
- Use the template specified by the developer for identification;
- The method of training the system to detect through external signs.

The empirical technique of "top-down knowledge-based methods" entails developing an algorithm that implements a set of rules that picture segments must obey in order to be recognized as faces. This collection of rules attempts to codify empirical information about a certain element of a person in a picture, as well as suggestions that a person receives while making a decision: whether or not he sees a face. The simplest rule:
- The central part of the face has brightness and uniform color;
- There is a significant difference in brightness between the central part of the face and the upper part;
- The face contains two symmetrical eyes, nose and mouth, and the brightness is very variable compared to the rest of the face.

A powerful image reduction approach [1] for reducing interference and computing processes, resulting in a drastically reduced image in advance (Figure 1). It is easy to identify sections with a uniform distribution of brightness (the predicted area in which the face is placed) in this drawing, and then ensure that there is an area with dramatically variable brightness in the room: these areas can be employed to varying degrees.

Histogram construction [2] the method used to determine the area of the image with a "face" for constructing vertical and horizontal histograms (Fig. 2). The candidate's area was requested for facial characteristics (1).

$$HI(x) = \sum_{y=1}^{n} I(x, y) \, ; VI(x) = \sum_{x=1}^{m} I(x, y) \qquad (1)$$

This approach was employed in the early phases of computer vision development because to the minimal computational power needs for image processing for processors. The approaches presented above have strong indications for recognising faces in photographs with a uniform backdrop and are simple to implement using machine code. Several comparable algorithms were later developed. However, these approaches are unsuitable for processing photos with a high number of complicated faces or backgrounds. They are also extremely sensitive to head tilt and rotation.

Another type of face recognition method is thematic invariant approaches (invariant approaches to functions) based on bottom-up knowledge. There is no attempt to properly codify the processes taking place in the human brain, which is a solution to the problem from the other side. By default, proponents of this strategy attempt to detect patterns and properties of face photos in order to locate unchanging facial traits independent of angle or location.
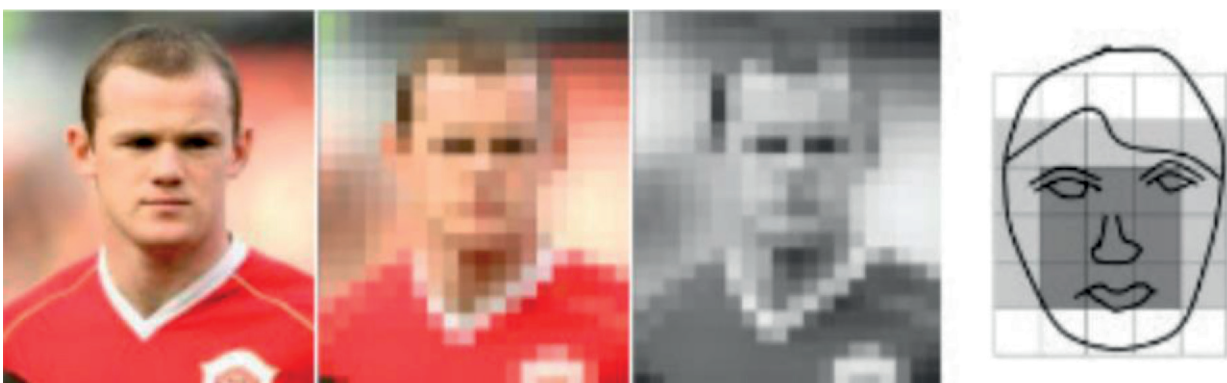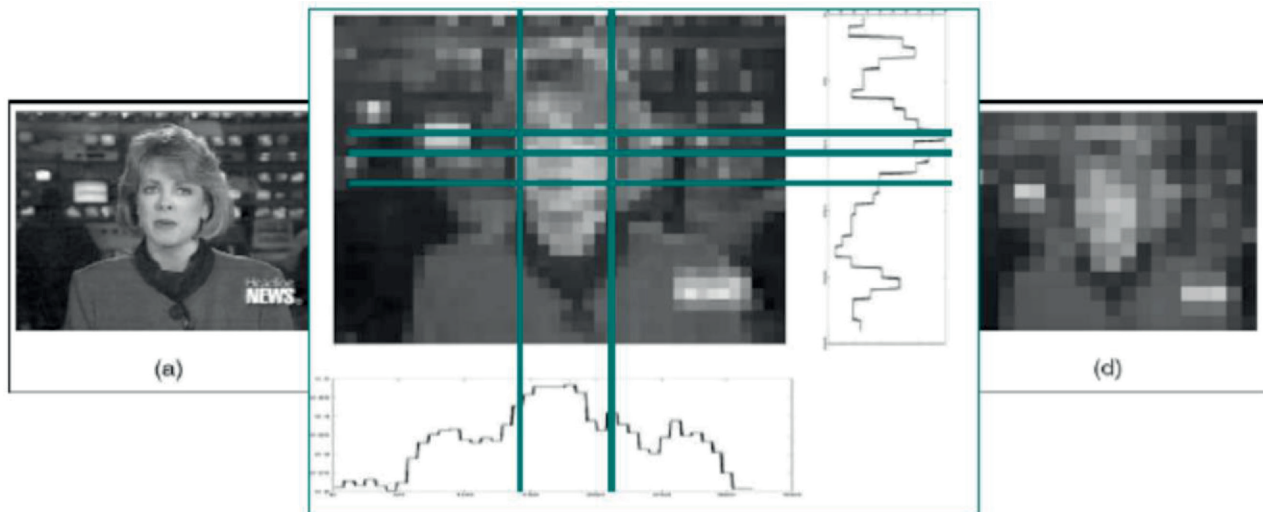

Fig. 1. The Yang & Huang method

Fig. 2. The Kotropoulos & Pitas method

The main stages of the algorithm of this set of methods:
- Detection of obvious facial features in the image: eyes, nose, mouth;
- Detection: face borders, shape, brightness, texture, color;
- Combine all the detected immutable functions and test them.

Face identification in complicated settings [3] entails determining the optimal geometric arrangement of facial characteristics. For this, Gaussian filters with a variety of ladders and orientations are utilised. Following that, a random search is utilised to locate recognised items and match them with the locations of others (Fig. 3a). The feature clustering approach [4] searches for the region of interest in a picture by using the second derivative of the Gaussian filter (Fig. 4a). The boundaries surrounding each such region are then grouped using a threshold filter. The identified data are then combined using a Bayesian network to pick face traits.

This group's approaches offer the benefit of being able to detect faces in a variety of settings. Even if there is a minor disturbance of the face with other objects, noise, or brightness, the proportion of trustworthy recognition is greatly diminished. The image's intricate backdrop also has a significant influence. The methodologies under examination are empirical in nature, which is both their strength and weakness. Because of the high variety of the identification object and the sensitivity of the kind of face to shooting and lighting circumstances, we may easily assign face detection in images to highly complicated tasks. When conducting a series of relatively easy checks, using the thumb rule allows you to generate a certain face pattern and lessen the burden. Nevertheless, despite the absolutely reasonable premise to try to use and repeat the successful thing of the human vision recognition tool, the first type of method is still far from its prototype in terms of effectiveness, since it was decided to choose this path, first of all in the process of solving the problem of pattern recognition, the processes occurring in the brain are far from being fully studied, and the empirical knowledge of people mastered by researchers at the conscious level, secondly, it is difficult to effectively turn informal, because excessively strict rules in some cases will lead to the fact that the staff will not be found. On the contrary, overly general rules will lead to a large number of false detections.
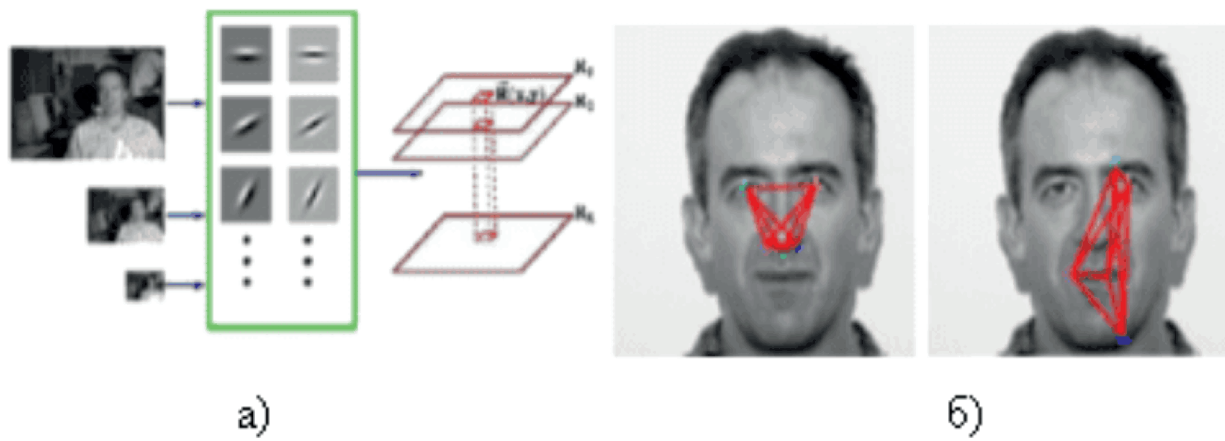
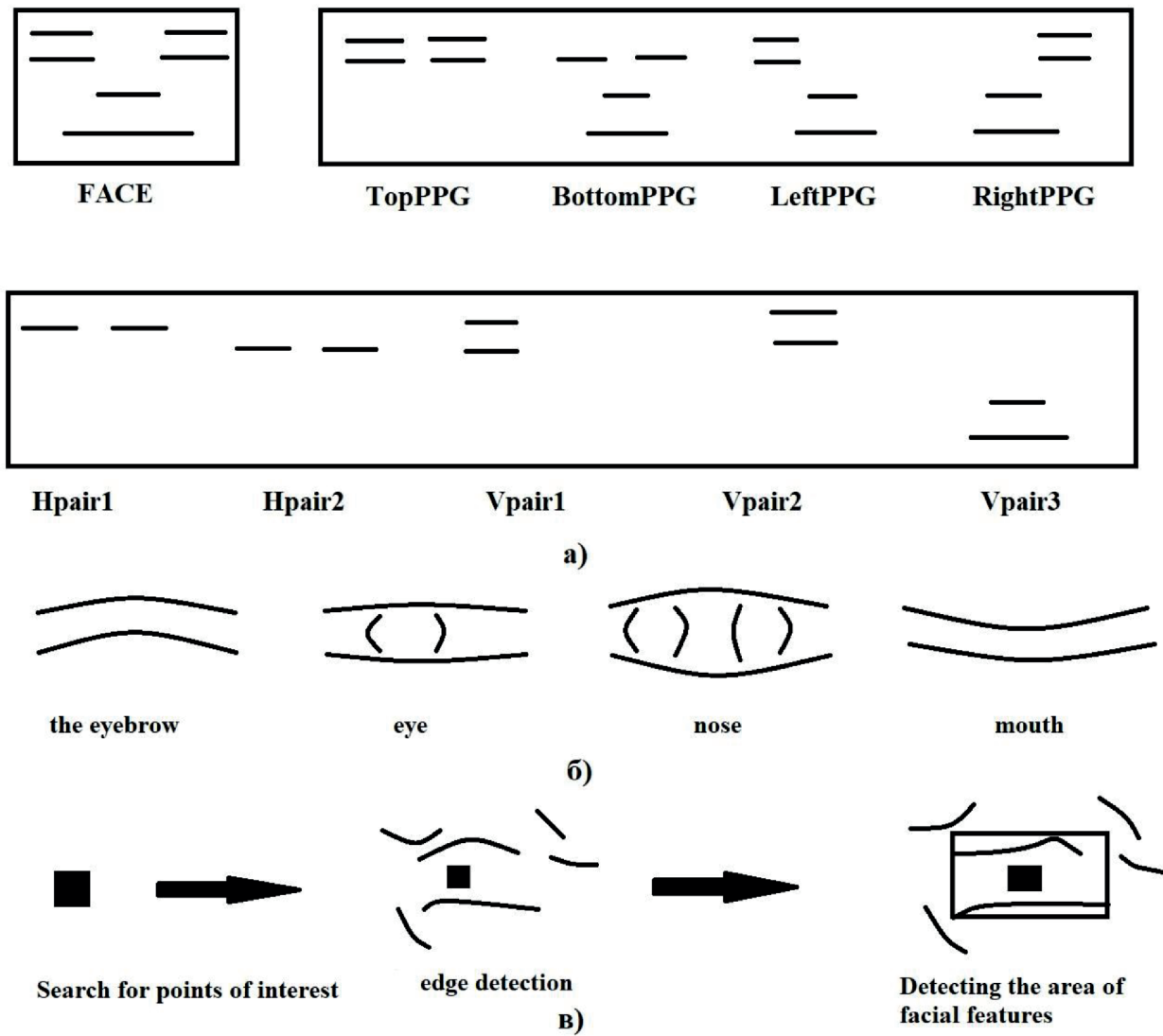Fig. 3. a) Multiscale multidirectional filtering, b) True and random triggers

Fig. 4 a) The model of the face and its features; b) The paired boundaries of the facial features;
c) The selection of the facial feature area

Recognition using templates set by the developer (Template Matching Methods)

Templates define a specific standard image of a face, for example, describing the properties of individual parts of the face and their possible mutual location.

Face recognition using a template consists of checking each area of the image according to a given template.

Features of the approach:
- two types of templates:
  1) non-deformable
  2) deformable
- templates are pre-programmed, untrained
- correlation is used to find the face in the image

The usage of templates as the brightness coefficients of a pair in two locations is included in the approach of employing three-dimensional numbers for face identification [5]. The full image must be supplied for comparison with the defined form in order to establish the position of the face. Furthermore, this should be done on several stairwells (Fig. 5).

The control point distribution model [6] is a statistical model that represents objects with variable shapes. This method's ability to differentiate the forms of varied objects in a training set with a limited number of shape parameters is a beneficial characteristic (Fig. 2a). 6). This exact and concise parametrization may be utilised to create efficient classification systems.

Pattern recognition has the benefit of being relatively easy to deploy and producing decent results on photographs with less complicated backgrounds. The key drawback is that the sample must be calibrated in addition to the facial image. People questioned the practicality of employing it due to the enormous difficulty of calculating models from various angles and the modification of the face.
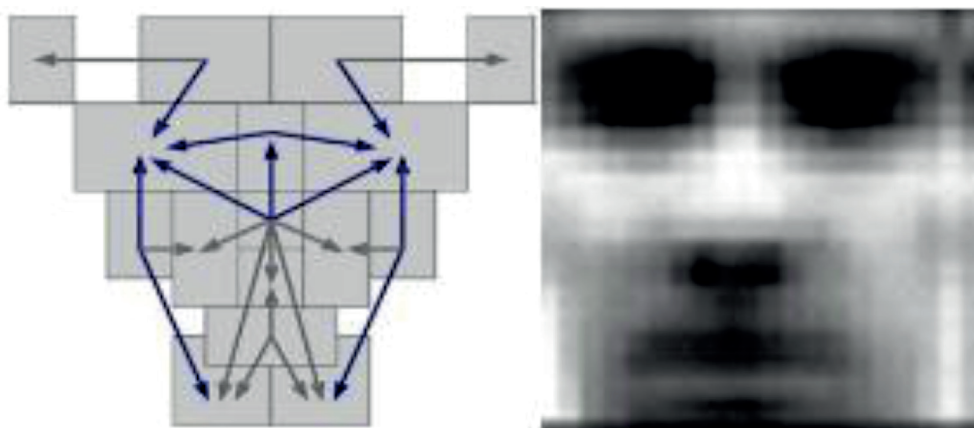

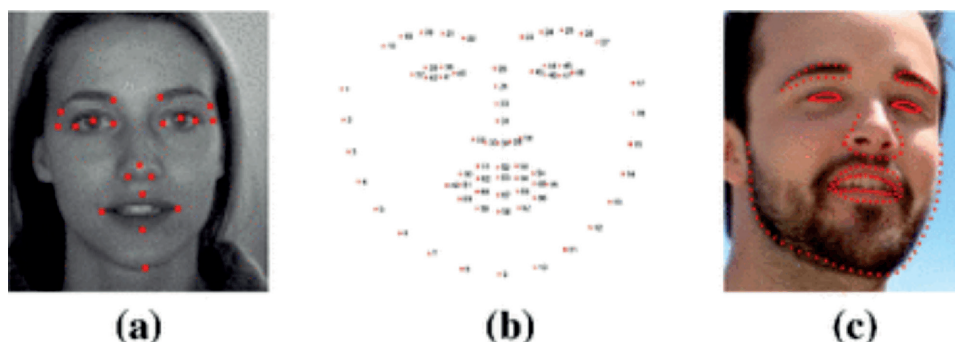Fig. 5. The method of face detection using three-dimensional shapes

(a) (b) (c)

Fig. 6. Models of the reference points distribution

Methods for detecting a face by external signs (methods for which it is necessary to conduct a training stage of the system by processing test images)

The image (or a portion of it) coincides with the generated vector of features used to categorise images into two classes: face and non-facial. Typically, approaches based on the building of a mathematical model of the face picture are used to discover all rectangular image fragments of all conceivable sizes and verify each of the pieces for the existence of a face.

Since the complete search scheme has obvious disadvantages, such as redundancy and high computational complexity, the authors use various methods to reduce the number of fragments under consideration. Basic principles of methods:

- Scholastics: each of them is scanned by a window and represented by vectors of values
- Block structure: the image is divided into cross sections or non-overlapping parts (Fig. 7) of various scales and is evaluated using vector weight estimation algorithms.

Fig. 7. Examples of splitting an image into sections

To teach the algorithm, a library of human and "non-human face" images and other images was manually prepared (Fig. 8).

It is critical to emphasize that the most crucial goal is to identify a powerful classifier. They will be given the greatest priority for inspecting the items discovered in the photograph. Because of their similarities, the number of weaker classifiers should be decreased, as should the deletion of classifiers originating from noise emissions. Let's go over the major methods for doing these tasks:

- Artificial neural network (neural network: multilayer perceptron);
- Analysis of the main components (PCA)
- Factor analysis (Factor analysis);
- Linear discrimination analysis;
- Vector Carrier Method (SVM);
- Naive Bayesian Classifier (naive Bayesian classifier);
- Hidden Markov model (hidden Markov model);
- Distribution – based method;
- Combination of FA and the method of main components (Mixture of PCA, Mixture of factor analyzers);
- Sparse network of windows (Sparse network of winnows (SNoW));
- Active Models (Active Appearance Models);
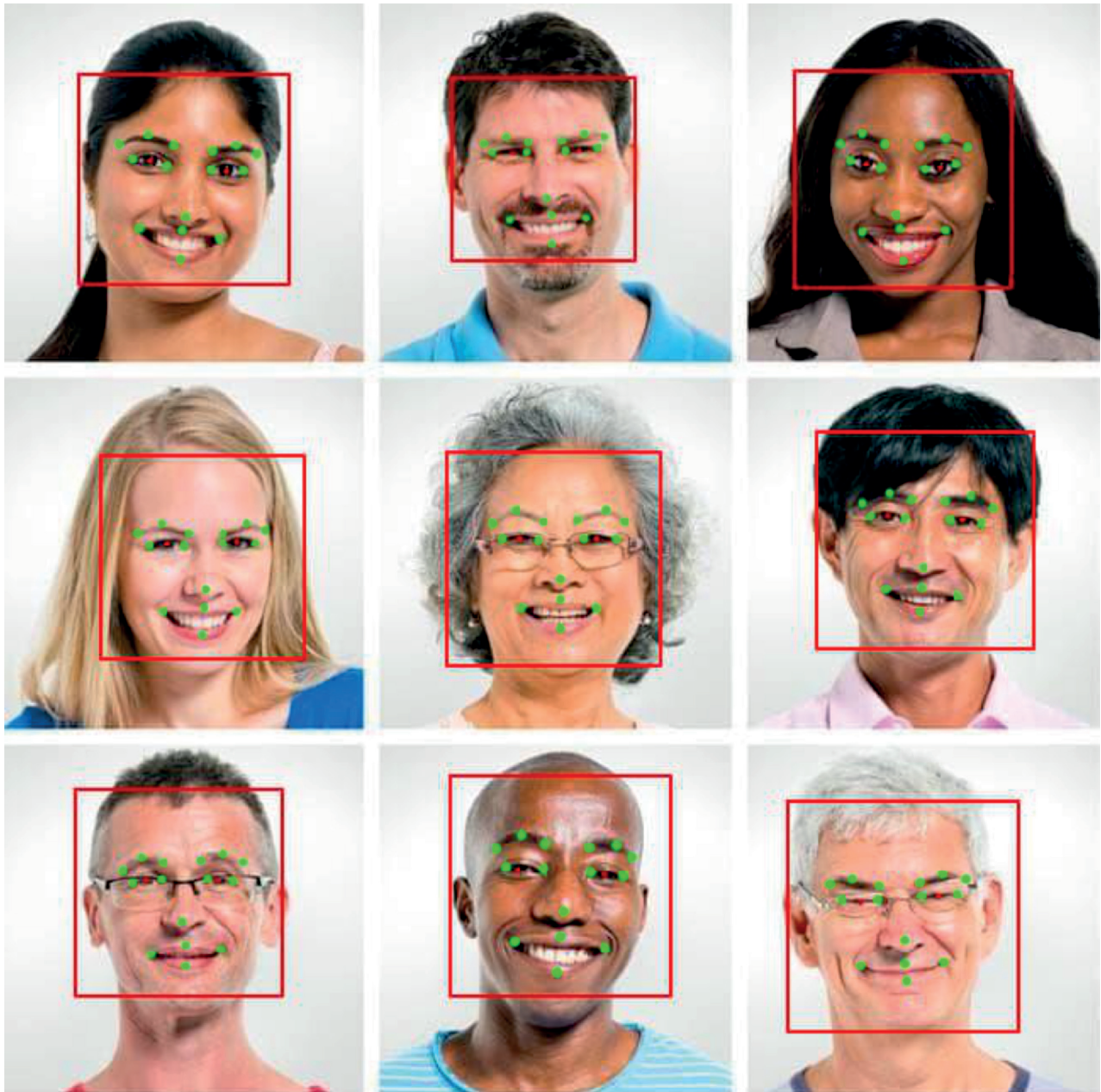- Adapted improvement and the Viola-Jones Method based on it [7, 8], etc.

Figure 8. Example of a library of faces

Let's take a look at a couple of these traits. The methodology of artificial neural networks is now the most widely utilised way for dealing with facial recognition challenges in photos. An artificial neural network (Ann) is a mathematical model that replicates a network of connected neurons that interact with one another. In contrast to traditional programming, neural networks are taught. Technically, training comprises the search for communication coefficients (synapses) between neurons (Fig. 9).

The vector machine support approach is used to lower the size of the object's space without considerably reducing the object's training set's information value. Using the principal component approach on a set of vectors in linear space allows us to go to a spatial base, where the set's primary variance is oriented along the first few axes of the base (called the main axis). The extension of the subspace obtained on the main axis is the best of all spaces in terms of best describing your training set. This is a series of algorithms used for classification

and regression analysis tasks that are comparable to the "learning with a teacher" approach. This method belongs to the family of linear classifiers. The vector machine support method is based on the need for linear class separation.

The goal of most classifiers is to reduce classification mistakes (also known as empirical hazards) in the training set. On the contrary, using the vector machine support technique, a classifier may be built that minimises the estimation of the upper limit of the predicted classification error (including for unknown objects not included in the instruction set). The search for a hyperplane in object space that divides the categories of face photos from "no face" images is the usage of vector support machines to tackle the challenge of face recognition. It is doubtful that complicated classifications, such as human faces and "human" pictures, can be separated linearly. However, classification with support vectors enables devices to create default feature vectors in potentially multidimensional spaces (much bigger than the picture space) where classes may be separated linearly. Using the default projection of kernel functions does not result in complicated calculations, allowing linear classifiers to be used successfully for continuous linear classes.

The Viola-Jones method seems to be the most promising today in terms of high performance, low false positives and most correct facial recognition. the basic principles on which this method is based are as follows:

- The image can be used as a complete representation, which allows you to quickly calculate the necessary items;
- Use the Haar logo to find the desired object using it (in this case its face and characteristics);
- Boosting is used (with English boost-improvement, improvement) to select the most appropriate function for the desired object in this part of the image;
- All flags are accepted when entering the classifier, and the result is "True" or "False";
- Cascade Flag is used to quickly remove windows where the face is not found.

Because the development of the classifier is long, but the results of the face search are rapid, we concentrate on face recognition techniques in pictures. The algorithm performs effectively and detects face characteristics from a modest angle of up to 30 degrees. When the angle of inclination exceeds 30 degrees, the detection percentage plummets dramatically. This makes it hard to distinguish the face of a person returning from any angle in a normal application, making its usage in current production systems difficult or impossible owing to its expansion.
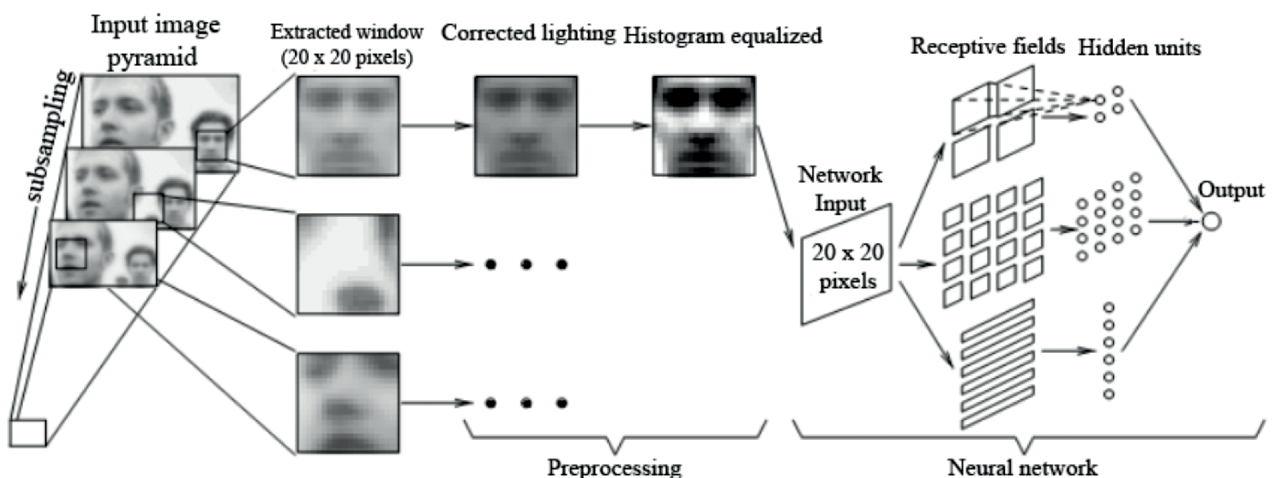

Fig. 9. Training of the system using a neural network

As a rule, the task of detecting a person's faces and facial features on a digital image is as follows: there is an image on which the desired objects are located. It is represented by a two-dimensional array of pixels of size $V * H$, in which each pixel is important:

- From 0 to 255 if the image is black and white;
- From 0 to 255 3 3, If the image is colored (components R, G, B). As a result of their work, the algorithm must determine their faces and characteristics and mark them-the search is performed in the active area of the image with rectangular signs that describe the found face and its characteristics:

$$Rect_i = \{x, y, w, h, a\}, \tag{2}$$

where x, y are the coordinates of the center of the i-th rectangle, w is the width, h is the height, and a is the angle of inclination of the rectangle to the vertical axis of the image. In other words, an approach based on a scanning window is used for drawings and photographs: the image is scanned by a search window (the so-called scanning window), and then a classifier is applied to each position. The system of training and selection of the most significant features is fully automated and does not require human intervention, so this approach works quickly.

The full representation of images in the Viola-Jones method is used to perform any actions with data. The full view allows you to quickly calculate the total brightness of an arbitrary rectangle in a given image, and regardless of the rectangle, the calculation time does not change.

The full view of the image is a string that corresponds to the size of the original image. Each of its elements stores the sum of the intensity of all pixels located to the left and above this element. The matrix elements are calculated using the following formula:

$$L(x, y) = \sum_{i=0, j=0}^{i \leq x \leq y} I(i, j), \tag{3}$$

where $I(i, j)$ is the pixel brightness of the source image. Each element of the matrix $L[x, y]$ is the sum of pixels in a rectangle from (0,0) to (x, y).

In the standard Viola-Jones method, rectangular features are used (Fig. 10) they are called haar-like wavelets:

$F = X-Y$, where X is the sum of the brightness values of the points covered by the bright side of the item and Y is the total of the brightness values of the points covered by the dark side of the object. The above-mentioned idea of integral pictures is utilised to compute them. The Haar notation expresses the brightness difference along the X and Y axes as a point value.

The algorithm for scanning a window with functions is as follows: the picture is inspected, the scanning window is chosen, and the functions to be utilised are chosen. The scan window then begins travelling across the picture in ascending order with 1 window cell (assuming the window's size is 24 X 24 cells). The scan is carried out in descending sequence of scale. Not only does the picture squeak, but so does the scan window (the disc size changes). All discovered characteristics are classified by the classifier, and the classifier "makes a choice."

It is just impracticable to calculate all of the characters on the desktop with little power usage throughout the search process. As a result, the classifier should only respond to the necessary subset of all functions. It is very natural that the classifier must be trained in order to locate persons for this subgroup. This can be accomplished through the use of computer-assisted training.

There are various objects (pictures) in the context of algorithms that are classified in some way. A restricted number of photographs is provided in the category to which they belong

(for example, it may be the category "front position of the nose"). This collection is referred to as learning templates. It is uncertain if this item belongs to the class of other objects. It is important to develop an algorithm capable of classifying any object in the original collection. There is a stimulation method — a combination of procedures that help increase the accuracy of the analysis model — that can be used to accomplish such a complicated learning job.

An effective model that allows a small number of classification errors is called "powerful".
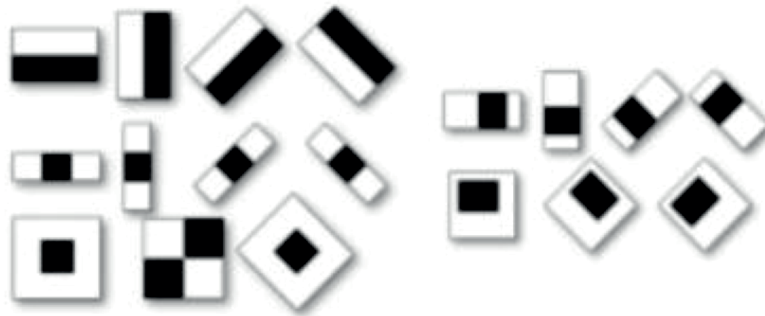
Fig. 10. Rectangular haar-like features

On the contrary, "weak" does not allow us to reliably separate the clock or make accurate predictions, allowing many mistakes in our work. The improved face search algorithm looks like this:

1. Definition of a weak classifier with rectangular characteristics;

2. For each movement of the Scan Window, calculate the rectangle function for each example;

3. Select the most appropriate threshold for each attribute;

4. Choose the best logo and the most suitable threshold;

5. The sample is overweight.

The cascade model of strict classifiers is actually the same decision tree in which each Rewe node is constructed by detecting almost all images of interest and areas of cast other than images. In addition, the nodes of the tree are arranged so that the closer the nodes are to the root of the tree, the less primitive it is, so it takes less time to make a decision. This type of cascade model is very suitable for image processing in which the total number of detected images is small. In this case, the method can quickly determine that this area does not contain image and proceeds to the next image.

It is worth noting that if a color image is introduced into the detection system, the speed of the algorithm can be significantly improved by preprocessing the image using color coding [9]. In addition, color coding helps to reduce the number of false positives.

To date, the Viola-Jones algorithm is the most popular because of its high speed and high accuracy of operation.

References

1. G. Yang and Thomas S. Huang. «Human face detection in a complex background. Pattern Recognition», 27 (1):53–63, 1994.
2. C. Kotropoulos, I. Pitas. «Acoustics, Speech, and Signal Processing», 1997. ICASSP-97, 1997 IEEE International Conference on p.2537–2540 v. 4.
3. T.K. Leung, M.C. Burl, P. Perona. «Finding Faces in Cluttered Scenes Using Random Labeled Graph Matching».

4. K.C. Yow, R Cipolla, «Feature-based human face detection», Image and vision computing 15 (9), p. 713-735, 1997.
5. Sinha, P. (1996). «Perceiving and Recognizing threedimensional forms» PhD thesis, Massachusetts Institue of Technology.
6. Lanitis, A.; Taylor, C.J.; Ahmed, T.; Cootes, T.F.; Wolfson «Image Anal. Classifying variable objects using a flexible shape model» Image Processing and its Applications, 1995., p.70-74.
7. 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), 2001, vol. 1, p-511 — p-518.
8. P. Viola and M. J. Jones, «Robust real-time face detection», International Journal of Computer Vision, vol. 57, no. 2, 2004., pp. 137-154.
9. Buchatskiy, A.N., Tatarenkov D.A., "Selection of the Optimal Color Space for Reducing False Positives Rate in the Viola-Jones Method", Actual problems of infotelecommunications in science and education, II International Scientific-technical and Scientific-methodological Conference. St. Petersburg, 2013.
10. L. Neumann and J. Matas. A method for text localization and recognition in real-world images. 2010
11. A.I. Dzhangarov, M.A. Suleymanova and A.L. Zolkin. Face recognition methods. IOP Conference Series: Materials Science and Engineering.
12. "Creating a face recognition model using deep learning in Python". https://sudonull.com/post/6434
13. Adil Sarsenov, Konstantin Latuta. "Face Recognition Based on Facial Landmarks", 2017 IEEE 11th International Conference on Application of Information and Communication Technologies (AICT), 2017.
14. "Intelligent Systems and Applications", Springer Science and Business Media LLC, 2019.
15. A.S. Miroshnikov, I.A. Berko, A.A. Berko. "Optimization Method for the Parallel Algorithm for Finding Faces in Graphic Images", 2021 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), 2021.