DOI: 10.17587/prin.13.331-343

Overview of Face Recognition Algorithms for Person Identification[1]

D. V. Alexandrov, dmalex-m2@yandex.ru, School of Software Engineering,
National Research University Higher School of Economics, Moscow, 109028, Russian Federation,
Department of Corporate Information Systems "MIREA — Russian Technological University", Moscow, 119454, Russian Federation

Corresponding author:

Dmitriy V. Alexandrov, Professor, National Research University Higher School of Economics, Moscow, 109028, Russian Federation, MIREA — Russian Technological University, Moscow, 119454, Russian Federation
E-mail: dmalex-m2@yandex.ru

Received on July 13, 2021
Accepted on March 28, 2022

Trends in computer vision and pattern recognition and capabilities of modern computers contributed to a considerable amount of research of these areas application in facial recognition systems. The purpose of this paper is to investigate the most significant methods of face recognition. In the first two sections of current paper, the methods of face recognition and identification are presented. The analysis of these methods covers the most important features of the pattern recognition area. An application of groups of methods is considered for different purposes. This paper contains comments for capabilities of algorithms under observation. The third section reveals result of the algorithms testing using real-world datasets and examples.

Keywords: *person identification, face recognition, neural networks, machine learning*

For citation:

Alexandrov D. V. Overview of Face Recognition Algorithms for Person Identification, *Programmnaya Ingeneria,* 2022, vol. 13, no. 7, pp. 331—243.

УДК 004.93'1

Д. В. Александров, д-р техн. наук, проф., dmalex-m2@yandex.ru,
НИУ ВШЭ, РТУ МИРЭА

Обзор алгоритмов распознавания лиц для идентификации личности

Тенденции в областях компьютерного зрения и распознавания образов, а также возможности современных компьютеров способствовали значительному объему прикладных исследований в сфере создания систем распознавания лиц. В статье представлено исследование наиболее значимых методов распознавания лиц. В первых двух разделах статьи представлены методы распознавания и идентификации лиц. Анализ этих методов охватывает наиболее важные особенности сферы распознавания образов. Рассмотрено также применение групп методов для различных целей. Даны комментарии по применимости исследуемых алгоритмов. В разд. 3 представлены результаты тестирования алгоритмов с использованием реальных наборов данных и примеров.

Ключевые слова: *идентификация личности, распознавание лиц, нейронные сети, машинное обучение*

[1] The article is based on the materials of the report at the Seventh International Conference "Actual problems of Systems and Software Engineering" APSSE 2021.

Introduction

Since the beginning of time people were aware of keeping their goods safe, and that is the main reason for the invention of locks. Through thousands of years people were changing the mechanism of door locks and standalone locks in order to create a better construction that will not let people without a valid token — usually, a key, — get any further. For some time, people were only saving money under lock and key. Treasures or other worthy goods could be traded or exchanged for making a living. Shortly afterwards people decided to keep their passport, personal mail, under lock and key, bank clerks started locking their records — since the moment information found a value.

Previously it was possible and convenient to present information only on paper, or a close substitute. The situation remained unchanged until the end of 19th century, when technical revolutions led to various important inventions, like the phone and the telegraph, which consequently gave birth to the first electronic computer. Since the creation of the first portable personal computer the market has grown and changes at a staggering speed — the quantity of information, that could be stored on a hard drive, increased from 240 bytes to 4—5 terabytes, which is more than around 10^{11} times. However, the quality of the information changed as well — the first computers stored only numbers to make calculations, nowadays it stores photos, videos, documents, programs and much more. The first electronic password was implemented by MIT (Massachusetts Institute of Technology) researchers in the 1960's, when there was a need to distribute files between different users, so that each file had a 'lock', that could be opened with a 'key' — password. Soon it evolved in a huge variety of alternatives — 2-factor password verification, physical USB key, graphical representations instead of numbers etc. These alternatives have no leader as all of them have vulnerabilities — if a thief accesses the phone of his victim, two factor authorization process does not work, password or finger drawings are easy to steal. The first idea of an application of human identity verification in electronic systems came from police methods — fingerprints. Soon scientists started researching methods of implementation of various biometric characteristics for identification.

A reasonable impetus to development of biometric identification systems gave the distribution of modern smartphones with its capabilities. Fingerprint sensors are now built-in and being used for identification, purchases and many other purposes in most of middle-class or higher smartphones, as well as a frontal camera, which in most cases permits the production of a high-resolution photo of the person who is holding the phone.

However, facial recognition systems that are in the market today are often a subject to hacking as there are various methods of corrupting the systems. For example, modern Facial Recognition System (FRS) should distinguish the person who is holding the phone and the picture that is being held below the camera. In order to replace current authorization methods with FRS it is crucial to obtain the same or better error rate, which is very low in the case of standard verification systems, having, despite that fact different problems, it is easy to forget the password or, in case of low password complexity — to crack it.

Modern FRS use different types of neural networks to accomplish their aim. It is interesting that most of theoretical basics, needed to build such systems, were presented in a previous century but they could not find their application in wide areas until recently due to the lack of technology and computational power. Since that moment some new areas were defined such as computer vision, machine vision, pattern recognition or graphics visualization, all of them have different aims and problem solving methods, which are, nevertheless, mostly based on an image process and some other data, that is being gathered by different sensors.

One of the most recent and promising methods of identifying a person is called a 3D recognition. In existing solutions an identification system analyses position of eyes, lips, ears, jaw and nose shape and distance ratio between these objects; some of them also count complex eye characteristics — eyes and eyebrow shape and color. In case of 3D recognition the contour of eye sockets, nose and chin is also being analyzed. An important benefit of the 3D method is that it can recognize the face from different angles. Another advantage is stabilization of light unlike other methods. In the case of a badly lit environment, attempts to recognize a face, are very likely to have a lower precision rate due to possible shadow presence, which always spoils calculations.

It is vital to mention, that the above-described advantages of a 3D recognition method can be achieved only in cases of using either three different cameras in an array, or using specific sensors for obtaining a 3D model of the head. Even though in terms of this work it will not include any of these hardware methods of facial recognition, multi-dimensional models are still going to be judged in order to verify if they still can be useful for facial recognition.

Based on an experience of existing solutions development it is fair to expect, that in most cases it is going to be used in combination with different methods in order to achieve a considerable result. This approach is now being used in a huge number of distributed systems to increase an accuracy of problem solving and to prevent system break-in. As it was mentioned before, it is crucial to realize, that the system should not only identify a person by his face, but also should be aware of a possible corruption with different existing techniques.

The goal of current research is to implement a facial recognition module for person identification by using the most appropriate stack of technologies and the most efficient and suitable algorithm of pattern recognition available.

The above aim will be accomplished by fulfilling the following research objectives:

- to observe existing methods of solving the face recognition problem and lying underneath theoretical basics;
- to implement the most relevant and accurate ones;
- to test it in a real-world circumstance;
- to select the most appropriate one or few of them and to build the FRS on top of it.

This paper consists of an introduction, three parts and a conclusion. The introduction reveals the relevance, an object, purpose and methods of research, reveals the theoretical and practical significance of the work. The first and the second parts deal with theoretical basics of facial recognition patterns, they describe an analysis of existing methods, and results of testing each method on real world examples. The third part is devoted to technology stack and program implementation itself, it mostly consists of practical results, obtained through the process of program development. In conclusion the results of study are summarized and final conclusions on the topic under consideration are formed.

1. Face recognition basics

Face recognition is a challenging problem in the field of computer vision and image processing. Through the history of technology development various algorithms were proposed and implemented as classifiers, that provide a sufficient precision rate and a high potential for a wide area of tasks.

The process of face recognition consists of multiple steps, that the system needs to go through to accomplish its aim:

- getting the image/video source;
- face detection;
- face normalization;
- feature extraction;
- descriptor feature matching;
- face identification.

For each step different methods of solving the problem may be used. Some older FRS did not follow exact steps, for example they could analyze the whole picture instead of an extracted frame with detected face. It could significantly increase execution time and accuracy.

Firstly, it needs to obtain a media file containing an object, that is going to be elaborated on. Media file may represent either a local video or image file stored in typical containers (png, jpg, mp4, flv) or a live stream from a camera, which is usually also represented in one of a wide known standards (RTSP, RTMP, HLS). Once the file or stream is acquired, the system tries to detect the face on each frame of presented media file. It may be mentioned that simple algorithms of face detection are not tied to specifications of an exact object "face", thus it can also recognize any kind of object — license plates, pedestrians, cars. More complicated groups of algorithms may pay attention to inner features of human behavioral — emotions, speaking qualities.

In order to detect a face, the system usually uses genetic algorithm and Eigenface technology [1—4]. The genetic algorithm is a search method based on the concept of natural evolution. In such an algorithm each chromosome (individual) is a possible solution to a given problem. Each individual chromosome may mate with others in case of a good fit during one generation. Thus, the least fitting individuals disappear from the population because no other chromosomes are mating with them. Chromosomes that survive through one generation then breed a new generation of individuals by crossing-over and mutating. The process is iterative, and the number of generations varies depending on the quality of the result needed. At the end of the process, it is expected to obtain an optimal solution to the problem. In case of detecting a face, genetic algorithm may be applied by the following steps:

- normalize an image (change the range of pixel intensity values);
- produce RGB extraction to obtain a gray-scale image;
- perform preprocessing and encode a sub-window of an initial image;
- using generated chromosomes, perform a search on an obtained sub-view using Eigenface (the template of human face) being rotated and moved around the given frame;
- applying a genetic algorithm to calculate parameters of template position on a given frame (center, scaling, rotation and match rate);
- in case of low fitness values — proceed to the next iteration of the loop and continue with a next sub-window.

Fitness function is a function of the difference between the intensity value of the input image and that of the template image measured for an expected chromosome value. Fitness function is defined as:

$$f(n) = 1 - \frac{\sum\limits_{(x,y)\in W} \left| f(x,y) - f_{n,t}(x,y) \right|}{B_{\max} \times xSize \times ySize}, \qquad (1)$$

where B_{max} is a maximal brightness of an image; *xSize* and *ySize* are width and height of the template image; f and $f_{n,t}$ are the intensity values justified for the n-th position of the chromosome for an initial image and template respectively.

On a good match an algorithm stops and fixes a possible face region. It may be represented as an oval but usually the smallest frame that contain the face is being used. Genetic algorithm flow includes the following steps.

- Getting the original color image.
- Preprocessing.
- Image enhancement.
- Filtering.
- Generic searching.
- Creating the face region.

Modern facial detection systems first detect possible human eye regions. Then using the genetic algorithm system, they generate all possible face regions which should include eyebrows, nose and mouth comers. Each sub-window of an initial image is normalized to avoid the shirring effect due to head movement or illumination effect. The fitness value of each frame is measured based on its projection on Eigenfaces.

To create a set of Eigenfaces (template of human face) it is necessary to:

- using a training set of face images where eyes and mouths should be aligned, lighting conditions should be similar and one resolution value is specified, create an image matrix where each column of the matrix is an image;
- subtract the mean;
- calculate eigenvalues and eigenvectors of the covariance matrix, which can also be called eigenfaces due to the possibilities presented as an image;
- choose principal components by sorting eigenvectors in descending order of its eigenvalues λ_i and setting a threshold ε on the total variance of $v = (\lambda_1 + \lambda_2 + ... + \lambda_n)$ and evaluate the number of principal components k which is the smallest number that satisfies to:

$$\frac{(\lambda_1 + \lambda_2 + ... + \lambda_k)}{v} > \varepsilon, \qquad (2)$$

where n is a number of components.

Eigenfaces may represent both existing and new faces. It allows to calculate the distance of a new face from the mean face. Eigenvalues represent the scale of difference between the mean image and the current eigenvector value. The information is lost when the image is projected on a subset of eigenvectors. However, selecting eigenvectors with the largest eigenvalues discards most of the available options and will grant a desirable output. Eigenfaces is based on the Principal Component Analysis (PCA) and remains one of the most robust and applicable methods of face detection [1, 4, 6].

Another possible option to identifying a face is the Fisherfaces algorithm, which is based on Linear Discriminant Analysis (LDA). Its usage may be efficient in cases of problems of recognition of faces under different illumination. Instead of presenting the whole variety of possible options with given weights (eigenfaces) Fisherfaces is trying to lower the distance between inner-class samples of data and maximize — between different classes, for example two images of one person under different lighting conditions or in a different pose may differ more than pictures of different people made under the same conditions. LDA may be used in order not only to detect a face but also to recognize it, meanwhile PCA has a poor performance in face recognition.

After feature extraction the system prepares the recognition stage and matches with descriptor data. From this point of flow, it depends on the aim of the system, in cases of trying to detect emotion on a human's face the system will analyze existing datasets of people with this emotion. In cases of face recognition, it will proceed to the whole trained data in order to find the closest match. There are various databases that can be used to train and test the facial recognition system with datasets containing various numbers of people, emotions and conditions affecting pictures.

Obviously, the facial recognition system may be constructed on a different flow or algorithms. However, while no new algorithm is proposed, the described scheme remains one of the most efficient, due to possibility to extract a module and replace it with another one. Some simple FRSs use only one method to detect and recognize a face, some complicated ones use a huge variety of different methods depending on a situation and aim of the system.

2. Face recognition methods

Keen interest to the technology of face recognition brought a huge variety of different algorithms of solving the problem. Some groups of methods, that are going to be observed below may appear as a key technique for the whole process of FRS flow, e. g., former systems used only PCA to detect and identify a face on a given picture or, inversely, modern systems may utilize mostly 3D based algorithms to construct an object, measure it and compare it with another one. Algorithms are going to be classified in groups by main feature of analysis core.

2.1. Classical methods of pattern recognition

This group of methods was named due to using mostly traditional approaches of image analysis and face detection and recognition. It may be divided into methods utilizing local features and methods utilizing holistic features, that can also be subdivided into linear and nonlinear methods.

Linear projection appearance-based methods have proved themselves as a good solution in cases of a "perfect" situation — all objects are accurately aligned; the illumination influence is minimal and facial emotions should not be presented. The main explanation for such a bad performance is that faces are projected on a complex nonlinear manifold in high dimensional space, thus the algorithms cannot efficiently extract desired features. Principal component analysis, linear discriminant analysis and linear regression classifier are the most well-known algorithms of classical group methods.

2.1.1. Linear holistic methods

2.1.1.1. PCA (Principal Component Analysis)

One of the eldest algorithms initially proposed by Karl Pearson in 1901, is being used in a wide variety of areas including data compression, econometrics and bioinformatics. The mean is being subtracted, covariance matrix on its base constructed and eigenvectors and eigenvalues are calculated. In general, main components are eigenvectors sorted by descending eigenvalues. Eigenvectors, that are above a specified threshold form a feature vector. This vector is multiplied by the initial dataset and transposition brings the result of PCA — new dataset that is constructed solely on vectors we have chosen. The main idea of PCA is to find the straight, that will ensure the lowest sum of squared distances to this straight from the corresponding points of an original dataset.

When applied to image processing, firstly, the image is being transformed from an N^2 dimensional matrix to a one-dimensional vector. This vector contains an intensity of grayscale image pixels row by row. To detect patterns in a set of images, a matrix of such vectors is constructed and PCA is applied. The data based on chosen eigenvectors allows to calculate the distance between a new image and the initial dataset in axes derived from PCA analysis [7].

2.1.1.2. LDA (Linear Discriminant Analysis)

The algorithm is based on Discriminant function analysis, which was originally developed by Sir Ronald Fisher in 1936 and may be named as a generalization of an original one. As well as PCA, it has a lot of applications — bankruptcy prediction, earth science, biomedical studies. The goal of an approach is to reduce the dimension of a dataset by projecting it on a less-dimensional subspace. First, d-dimensional mean vectors for different classes and scatter matrices are computed both for in-between class and within class. Then eigenvectors and eigenvalues of scatter matrices are evaluated and sorted just like in PCA. A $n \times d$-dimensional matrix of samples multiplied by an obtained matrix of k eigenvectors transform data onto the new subspace.

The main difference between PCA and LDA approaches, is that in the first one it is needed to build component axes that will maximize the variance whereas in the second one, component axes are being maximized to separate different classes and find similarity in single classes. In general, reducing the dimension of an initial dataset significantly reduces computational costs for a given classification task and meanwhile minimizes an error occurrence possibility in parameter estimation [8, 9].

2.1.1.3. LRC (Linear Regression Classifier)

Linear classifier is aiming to identify which class or group does the object belong to. Response vectors for each class are being constructed, and the vector y representing the image under observation is being evaluated via logistic regression as follows in (3),

$$y = \frac{\mathbf{e}^{X_i \beta_i}}{1 + \mathbf{e}^{X_i \beta_i}}, \tag{3}$$

where X_i represents a class-specific model — matrix of one-class dataset and β_i is a vector of regression parameters. The weight of each feature derives from the gradient descent method. The minimal distance between original and predicted response vector will point to the class which is the most likely the object belonging.

The main disadvantage of such an approach is that in typical FRS it is unpredictable, if an image dataset contains "perfect" samples or they may contain occluded data, e. g., in cases that 3 out of 4 parts of an image are shadowed, then by the majority vote, even if the last one contains perfectly distributed facial features, it is very likely to have face portions mixed with an occlusion. When applied in FRS, before using LRC image dataset is usually being cleaned for potentially contaminated pixels using other methods like PCA or other simple linear methods [10, 11].

2.1.1.4. ICA (Independent Component Analysis)

The method is analyzing the multi-dimensional data to find underlying factors — components, that are both statistically independent and non-Gaussian. It is mostly applied to solve the problem of "the cocktail party" to extract the noise from a signal and obtain a clean desired source. The covariance matrix for an image vector is calculated (4), where E is a prediction operator, factorization of analysis of X brings an equation on (5), where Δ is a diagonal real positive and F transforms the original data X into Z in such a way that components of Z are statistically independent. The whitening operation performs the transformation of vector X into U (6), that will have a unit covariance matrix, where Φ is an orthonormal eigenvector matrix and Λ is a diagonal eigenvalue matrix, that come from equation (7). The rotation operation then consequently extracts sources by minimizing the mutual information, normalizing the operation, then evaluates final independent components [12].

$$\Sigma_X = E\left\{[X - E(X)][X - E(X)]^t\right\}, \quad (4)$$

$$\Sigma_X = F\Delta F^t, \quad (5)$$

$$X = \Phi\Lambda^{1/2}U, \quad (6)$$

$$\Sigma_X = \Phi\Lambda\Phi^t. \quad (7)$$

ICA identifies the independent source components from their linear mixtures (the observables). ICA thus provides a more powerful data representation than PCA as its goal is that of providing an independent rather than uncorrelated image decomposition and representation.

2.1.1.5. 2DPCA (2D Principal Components Analysis)

2DPCA differs from its ancestor PCA with an approach of treating images like 2D matrices instead of creating vectors of the image. An image covariance matrix is being constructed using an original matrix and its eigenvectors are evaluated for further feature definition. It allows an extracted feature from an object to be classified. Firstly, it obtains a mean matrix between all objects presented in class and evaluates a scatter matrix. Next, it computes a projection subspace by constructing orthonormal vectors to corresponding largest eigenvalues of the scatter matrix. Finally, projecting images on each of the obtained vectors derives the principal components vectors. Computing the sum of Euclidian distances between principal component vectors and tested ones and specifying a threshold value allows us to get an object prediction. This method is computationally lower than PCA and brings the same accuracy rate of identifying objects [13].

Reconstructive approaches (such as PCA and ICA) are reported to be robust for the problem of contaminated pixels, whereas discriminative approaches (such as LDA) are known to yield better results in clean conditions. In order to deal with disadvantages of linear methods some nonlinear extensions were proposed — KPCA and KLDA [14].

2.1.2. Nonlinear holistic methods

2.1.2.1. KPCA (Kernel Principal Component Analysis)

The method preserves simplicity of PCA and uses kernel methods carrying standard linear operations of PCA to reproduce kernel Hilbert space. The underlying idea is that it is almost always possible to linearly separate N points in $(d - 1) \geqslant N$ dimensions. This method grants a high level of accuracy, however, remains time-consuming to be implemented in real-world applications. There exist some methods that are involved in the process feature extraction in order to lower execution time costs.

2.1.2.2. KLDA (Kernel Linear Discriminant Analysis)

An approach also known as generalized discriminant analysis is also an extension of linear method LDA, which due to performing in a new feature space allows nonlinear mapping to be examined. Despite a strong theoretical validity KLDA, KPCA and other extensions over linear methods do not produce significant improvements when applied in the real-world environment. There are various methods based on GDA (General Discriminant Analysis) and being used for face recognition purposes, KLDA may be classified as one of them [14, 15].

2.1.2.3. LLE (Locally Linear Embedding)

LLE is a nonlinear dimension-reductive technique that maps data onto a single global coordinate system of lower dimension. Projected data is being analyzed to identify closest neighbors by any method (Euclidean distance, Pearson coefficients etc.). This approach shows its usefulness in case of applying it to data where objects are presented in a not-proper way, e. g. a half-turned face may be reconstructed as a fully presented one and successfully compared with a proper facial object.

Another approach called Locality Preserving Projection is behaving in the same way as LLE using slightly different methods, both are considered as possible solutions for training purposes, but questionable in case of projecting new data items [16].

Observed approaches above, both linear and nonlinear, may be attributed as holistic based. Below, local features-based groups of methods are going to be observed. The main difference is that features in the holistic-based groups of methods represent the optimal variances of pixel data, this data is used to identify humans, in case of local features-based methods — data, which is being analyzed, already contains facial feature content which has characteristics which allow us to identify one individual from another [17].

2.1.3. Local feature-based methods

2.1.3.1. LBP (Local Binary Patterns)

Unlike in previously observed groups of methods once the face on the image is detected, for each pixel an LBP is constructed and then the image is being divided into small regions from which LBP histograms are extracted and formed into a single feature vector. This approach allows to identify people and to deal efficiently with facial emotional states. In further research, after initial proposition of an algorithm it has been determined that LBP combined with the HOG descriptor (Histogram of oriented gradients) in terms of facial recognition may significantly improve the computing time and precision rate [18, 19, 29].

2.1.3.2. iSVM (improved Support Vector Machine)

The performance of an ancestor — SVM, based on the kernel function is lacking in efficiency, it was proposed to enlarge a spatial resolution around the margin by a conformal mapping, and so the distance between classes is growing, increasing the identification precision rate. The algorithm preserves a good computational speed with a high precision rate while working with datasets containing occluded images [19, 20].

2.1.3.3. LDP (Local Directional Pattern)

LDP method is a texture pattern that differs from its close opponent LBP with a detailed analysis of features; by applying LBP it is only possible to find typical one-level objects like curves and edges, in case of LDP it analyzes different types of edges and curves therefore going deeper. It is possible due to usage of Kirsch masks and evaluating eight directions on each pixel. Hence, the analysis of data goes deeper, and can find some features of objects of one class. Each pixel goes through the procedure of applying a certain threshold on obtained direction numbers, transforms above values to 1, below — to 0 and evaluates the number that identifies the information about one pixel. After the LDP application the image is represented by an LDP descriptor (histogram).

The approach uses nearly the same algorithm for computing histograms that will represent sample data. However, the possibilities of LDP are not limited with detection of low-level only objects, the method is able to work with contrast and those features observed mostly, that the human eye would have caught the first time [21].

2.1.3.4. SIFT (Scale Invariant Feature Transform)

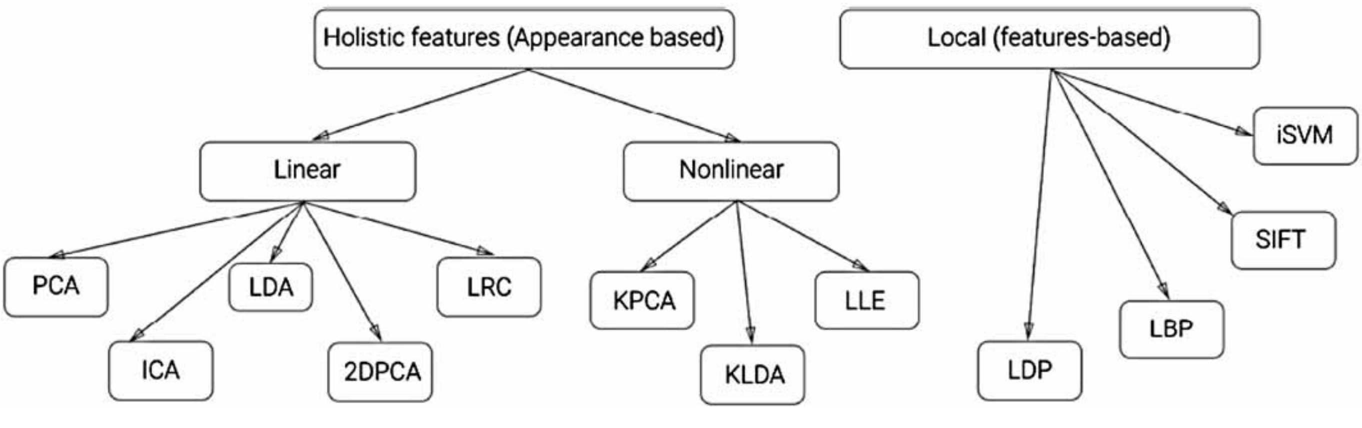
The method was initially proposed to extract some features from image objects, but soon was applied for object identification. SIFT does not pay attention to rotation and scaling and it also works well for pose and emotion invariants. This method is not computationally intensive and proves itself to be a good competitor to linear methods of face recognition [22, 23].

Thus, a number of classical face recognition algorithms were described (figure). The holistic (appearance-based) group include linear and nonlinear methods using slightly different steps of solving the problem, but still quite similar in terms of working with data. Local feature-based group differs in extracting feature descriptors and classifying them with multiple possible characteristics.

Classical methods are used in any kind of FRS in our days, but not only because of robustness of this group of methods but due to their stableness and ability to obtain desired data while applying other algorithms just as one step of the process.

2.2. Applied methods

In the previous section algorithms and their execution steps were observed, in order to improve efficiency of different phases of face recognition — face detection, feature detection and extraction, analysis etc. — some additional methods may be used. In this section its underlying mechanics are described and their application in the real-world environment is considered.



Classical face recognition methods

2.2.1. Artificial Neural Networks (ANN)

Neural networks are being used widely in completely different problem-solving processes — its relevance may be explained by robustness and efficiency of neural network application. Particularly, in face recognition, neural networks are usually used to compare a recognized face from an image with the data stored on a database — the data which was used to train neural network. ANN is a type of neural network that uses "nodes" of three types — input, hidden and output nodes. Gathering initial data from input neurons and using the transfer function to apply NN's knowledge and passing values through neurons multiple times, in the end it reflects the meaningful in output nodes.

In order to significantly improve a precision rate of a neural network it is better to train it firstly on a 'clear' data — images that were taken in perfect conditions and are not spoiled with any kind of occlusion. Thus, obtaining a low sum-squared error neural network may further consume "noisy" images and still make high-rated predictions [6, 24].

One of ANN's applications is Self-Organizing Map (SOM), which the application usually considers for dimensionality reduction and feature extraction. It projects data to one or two-dimensional spaces placing the nearest points close to each other on a new space, therefore, providing partial invariance to rotation and image deformation. After the SOM application a soft nearest neighbor decision method may be applied in order to identify unlabeled subjects [26].

A multilayer perceptron (MLP) is a type of artificial neural network that maps an input data on a set of appropriate outputs, it consists of multiple layers, obligatorily connected to each other, and uses the supervised learning techniques, e. g. Back Propagation Algorithm (BPA). BPA, which is basically a method for training the weights in a multilayer neural network. Its principle is to model a given transfer function in the way that we expect the output to be presented. It is applied in the middle of the training process in order to correct obtained weights and improve the output of neural network nodes. Recently it was nearly replaced with more efficient Support Vector Machines, but now it is being under observation for further research due to successes in deep learning. The significant disadvantage of ANN application in facial recognition is a necessity of a solid amount of data to produce a satisfactory prediction [25].

2.2.2. Gabor wavelets

Gabor filters is one of the available methods to extract local features from an image. Mostly it can be applied on the step of image processing — to detect and represent features on an image by applying a filter — a function, which convolves an image to a Gabor representation of an image, it captures the properties of orientation selectivity and spatial localization, optimizing space localization and frequency domain. The drawback of this method is in high dimensionality of feature vectors that are extracted by Gabor filters. In order to solve this problem usually some typical methods of dimensionality reduction are applied — PCA, LDA or their descendants.

Gabor wavelets' characteristics are similar to human visual system characteristics; they are able to work with samples of one person presented in different pose or illumination. When dealing with face recognition in Gabor transformation needs to be used over 40 filters — typically 5 scales and 8 orientations.

This approach proves to be robust even when dealing with one or a small amount of sample data. However, selecting the most important list of features that were extracted by Gabor wavelets is a time-consuming and computationally intensive method, so it is not useful when creating a real-time system of face recognition. The Simplified Gabor Wavelets (SGW) proved their unstableness to pose and facial expressions changes, thus losing all meaning of applying such approach [27].

Elastic Bunch Graph Matching (EBGM) is one of feature-based algorithms that uses Gabor Wavelets transformation of each feature point obtained with standard methods. The method is based on dynamic link structures. A set of fiducial points on an image represents a fully connected graph and is labeled with the Gabor filters applied to a window around initial points. A set of graphs of fiducial points is combined in a stack structure, which is called a Face Bunch Graph. In order to recognize a new face image, its image graph is compared with sample data and the similarity is evaluated by the lowest distance between obtained values [5].

2.2.3. Face descriptor-based methods

Descriptor-based methods were already observed earlier, this approach is used in Local Binary Patterns, Scale Invariant Feature Transform and other species of image description methods. Unlike in other global-based methods where the entire image is observed for features detection, in FDB methods, local features of the image are evaluated in neighbor pixels and then form a global description of the whole image. The image is being divided onto sub-images, each of them present a local feature histogram, which afterwards is being concatenated into a global descriptor which can identify an image and its local features.

Methods that use such an approach are in general, robust and competitive even to completely new techniques of face recognition. It may be used as well for facial expression changes or to identify some inner characteristic of an object under observation — e. g., skin color, nationality and psychological predisposition [28].

2.2.4. 3D methods

Three-dimensional facial recognition is obviously a more accurate and robust approach to solving the problem of person identification. It does not depend on the illumination changes, facial expressions or poses. Thus, the model of a person's face gives the rather appropriate estimation of a person's identity. This group of methods is more likely to develop further in the close future due to a necessity of human recognition in an uncomfortable context with a high precision rate. Modern algorithms cannot provide an acceptable efficiency yet.

There are several drawbacks to this approach. First, in order to obtain a desired quality of model construction and respective human identification, it is necessary to use specific hardware — sensors, specific cameras or just several usual cameras connected in an array. Second, in most FRS it is not possible to obtain an object's 3D model because an aim of the system is to identify a possible intruder for example. However, for specific needs it is an acceptable approach — e. g., for foreign passport issue, hardware is used to capture both 2D and 3D images of an object for different needs. A 2D image is for person identification by an employee and 3D is for FRS verification. Also, it may be used in the case when the person significantly changes his outlook — e. g., cuts his beard or hair.

Methods exist for obtaining a 3D model of a person without using specific hardware — by taking several pictures of an object from different angles and then through object reconstruction, obtaining the model, which can then be used for identification. In the fields of 3D model construction there are also 3D Generic Elastic Models (3D GEM) and 3D Morphable Model (3DMM) which can construct a model just from one single frontal image. Its algorithm consists of the approaches mentioned before but aimed specifically for 3D reconstruction. Hereby, PCA is being used for minimizing the difference between principal components of 3D model's 2D acquired images of an object at different angles [30, 33].

2.2.5. Video-based methods

These groups of methods are the least explored in the field of face recognition and, nevertheless, one of the most potentially efficient [31]. Instead of a single image capture and its further analysis, video may grant a significant size array of images with a variety of object representations from different angles with different facial expressions and under different conditions [32]. An approach was not relevant until recent times due to a small number of cameras that can capture needed data of a desired quality. In the last few years, the amount of video content in the global network increased dramatically — people are publishing video content from their smartphone cameras, the number of cameras in cities is also significantly raised due to the concept of Smart city — safety reasons, control reasons, e. g., track how the municipal services do their job etc.

The K-nearest neighbors or Gaussian mixture model may be used to identify and compare obtained image sets from several video sources. The drawback of a modern perception of the video-based approach lies primarily in an incompleteness of possible data extracted from video sources. Instead of analyzing people's behavior, manners and emotional states, modern systems just create several images and compare them utilizing existing methods of face recognition.

One of the most recent developments in these respective areas permitted capture of some facial expressions, e. g., an eye blink. Development of technologies will continue to improve and raise a functional part of such recognition, because most modern facial recognition systems may be corrupted just with a printed image of another person.

The main concept of FRS may be represented by the following. The system should act like a person who knows every other person that he may meet in his daily routine. Modern technologies and, particularly, the Neural networks concept are consistently developing to the perception of a human's natural systems features. This approach may lead to a more-than-human precision rate of person identification but containing significantly more information about people. Capabilities of modern computers allow to store an enormous amount of data, which can be processed and analyzed, modern trends in Big Data are also aimed to solve the problem of large data parts analysis. The basic overview of methods that were observed in this section is presented in the table 1.

Table 1

Summary overview

Method	Advantages	Disadvantages
Classical face recognition algorithms	These methods project a face onto a linear subspace spanned by the eigenface images. The distance from face space is orthogonal to the plane of the mean image, so it may be easily turned to Mahalanobis distances with probabilistic interpretation	In case of varying degrees of luminosity of images represented in dataset it may lead to lack of a precision rate. The problem is that it is still unclear how to select the neighborhood size or assign optimal values for them
Artificial neural networks	A radial basis function for an artificial neural network is naturally integrated with a non-negative matrix factorization. Also, other approaches for process simplification regarding ANNs native linearization feature and computation speed up. It may be a perfect solution for FRS which will work with partially distorted objects	This group of methods require a large set of images in order to produce an adequate estimation. It is inaccurate in the same way like other statistically based methods
Gabor wavelets	Gabor wavelets work very well with spatial orientation issues. Various biometric applications are based on this approach	This approach requires the construction of a relatively high-dimensional Gabor feature space. Due to the high computational power needed it is not the best solution for a real-time. FRS Also, it is sensitive to changes in illumination
Face descriptor-based methods	This involves extraction and comparison of the most discriminant local features in order to minimize the difference between the same object and maximize between different people. These methods are discriminative and robust to illumination and expression changes. They offer compact, easy to extract and highly discriminative descriptors	These group of methods require a huge computational power while extracting descriptors. However, it is simple once this is done
3D-based facial recognition	This is significantly more accurate compared to the traditional 2D capturing process. It is independent to pose and illumination of the object, which makes this approach more robust	At the moment, it is not applicable to real-world applications due to the high computational requirements. It also requires very strict calibration in order to get a viable output
Video-based recognition	The potential and main advantage of this approach is the possibility of capturing more information from a video frame, choosing a more appropriate image set, catching human actions — blinks etc.	Video-based recognition is not fully investigated, but it has a high potential for developing real-time applications. It needs to implement other groups of methods to obtain the distance between images and measure its similarity

3. Application of observed algorithms in practice

In the previous section, various methods of face recognition were examined. In this section, the development part of current research is going to be described.

3.1. Experimental results

The aim of this section is to describe how to build a face recognition system for person identification, which will possibly replace or substitute a typical password authorization, therefore, it is not needed to care about head position or occlusion in images — user of such a system would be interested in correct recognition and will prevent bad sample data perception. In order to estimate each algorithm's efficiency sample datasets are going to be used; there exist various large datasets for face recognition, each of them may be used for different purposes — estimating precision rate under different lighting conditions, with a huge number of people who look similar or containing a differently labeled pictures with facial expressions labeled.

In terms of this research, it is not needed to explore efficiency of facial expressions influence on face recognition precision rate, it is only needed to grant a sufficient precision rate for a person who can try to authorize in different lighting conditions. The FERET database containing around 2,400 images, the Yale database with around 150 images and LFW database with more than 13,000 images were considered to provide a sample data for further testing in terms of current research. All these datasets are distributed for free to improve quality of an overall face recognition technology and to obtain independent results while testing a FRS.

It is decided to conduct a testing part for the most stable appearance-based linear methods — Principal Component Analysis and Linear Discriminant Analysis — and for the most accurate local features-based methods — improved

Support Vector Machine, Local Binary Patterns and Scale Invariant Feature Transform methods. PCA and LDA proved themselves to be robust and efficient algorithms of face recognition, compared to modern iSVM and SIFT it will indicate the quality growth in facial recognition technologies, LBP represents a texture-based approach.

In order to conduct current experiments, the following scenarios were considered — 6 images available for training, 4 images to identify a person for each class and 8 images available for training, 2 images for person identification — test samples. Also, it was decided to create two subspaces — in open subspace training data consists of images under various illumination conditions and images to identify differed from training data; in closed subspace training data consists of images under different illumination conditions as well, but test samples were taken under the same circumstances. For testing purposes SciKit open-source Python library was used to implement a range of machine learning algorithms.

In all methods under observation an increase of training data led to an accuracy improvement, as well as a presence of similar images in training set and in sample image. Local feature-based methods proved themselves to be stricter than appearance-based methods. Improved Support Vector Machine proved itself as the most precise algorithm (table 2).

Table 2

Precision rates for different ratios of training/test datasets

Method	Open subspace		Close subspace	
	60/40	80/20	60/40	80/20
PCA	71.5	75	86.5	90
LDA	76.5	77	89.5	93
iSVM	79	84	91	95
LBP	78.6	81.3	90	94.5
SIFT	81	83	90	94

3.2. Implementation of facial recognition system

Current research involves the construction of the prototype of Face Recognition System. The system for security reason is divided into backend and frontend parts. The iOS application represents the client side of it, and Node.js + Python is responsible for server-side calculations. The prototype itself is implemented for authorization needs of any service that wants to provide facial verification. The system also provides an API that is used on client side.

The system is built on widely distributed products and frameworks. Nginx is one of the most efficient web servers. FFmpeg library deals with media files handling and converting. Python scripts are used to access OpenCV functionality that has proven itself as one of the best free solutions for commercial projects. Node.js logical center is a core of the system, that handles the data and grants user to access his account. All the data between client and server is transmitted via secure web-sockets — one of the most reliable communication channels.

After registration, the evaluated classifier for each person is stored in a database and then will be retrieved when the person will try to authenticate. In case of a small precision rate at authorization process user will be proposed to make an additional training session.

3.2.1. Useful features of different algorithms

Through analysis of the algorithms its interesting features were revealed. There exists a giant number of features that can be extracted from a person's face to analyze his behavior, his intentions, his emotions and many others. All the data may be used further with different purposes — from psychological analysis to capturing a suspicious person.

Some holistic linear and nonlinear methods may also be used for finding a person wearing glasses or emotion identification, e. g., ICA and KLDA. There exist various applications of these methods, they are mostly entertaining: "Guess your nationality" or "The most similar Hollywood superstar", that are distributed widely for fun. However, there also exist various important applications of the methods, such as crowd analysis, behavioral patterns and computer vision.

Local based algorithms may be good when analyzing shape and texture. It may reveal various details about the person — skin qualities, nationality belonging, what kind of accessories he is wearing. These methods work well with detecting shapes, lines, edges or angles; therefore, it can be applied in other various areas. In current project iSVM utilize a detection of lines that were presented in datasets, and by this, defend the system from breaking when a user places a photo instead of himself. In order to test the capabilities of the system a registered person wearing sunglasses tried to authenticate but did not hit the threshold. However, it is vital to say that the system gave a 0.67 precision rate to identify a right person.

The amounts of information that can be collected with existing methods consequently lead to the topic of Big Data as only computational expenses may stop the analysis of all people that can be caught by Closed Circuit Television (CCTV) systems all over the world. It is possible to catch gender, age, ethnicity, clothes, style, eye color — all of this is potentially priceless information that can be used for direct marketing, government or medical control including during the COVID-19 pandemic.

3.2.2. Prototype implementation

The implemented program allows registration and authorization using a created classifier for the user. It may be applied to any kind of existing applications — private photo storage, messenger, e-banking etc.

Implemented algorithm of suspicious movement detection catches strange device orientation changes and protects the initial authorization module from breaking it with an image of a person. The algorithm involves usage of a motion coprocessor built-in to iPhones.

Authentication process takes approximately 5—6 seconds to gather all needed data and requires the user just to look at the camera, it is permitted to tilt the head not more than 45 degrees.

At the end of authentication process the user will be notified about the result obtained by a classifier, in case of successful and correct process handling (shaking the phone is not permitted) and if the obtained precision rate is over the specified threshold the user is granted access to the account where photos are stored. For the process of registration nearly the same screens are being used with different content though, the user will spend around 10 seconds sending data, he will be asked to face the camera if he moves to much.

All the data is stored and processed on the server. Once the user completes his photo session, an array of photos is delivered to the server, each picture is analyzed, cropped, normalized and modified for better recognition purposes. Based on the linear Support Vector Machine approach it evaluates the classifier, using the benefits of this approach it constructs two classifiers — one for person recognition and one for person differentiation, so-called "within-class" and "in-class" classifiers. Then the training stage is over. However, it is possible to train classifiers multiple times, growing the size of the datasets and making the predictions better.

The application is able to analyze not only photos but videos, so each user's registration grants a 25 fps video of around 8 seconds, which contains around 40 suitable images that can be analyzed to compare emotional features of each person, features of the skin and the eyes under different circumstances, it is possible to require actions from user to approve his identity by shaking his head or blinking.

Conclusion

In terms of current research, the analysis of existing technologies was conducted, the most significant methods were tested and afterwards implemented in the FRS prototype. The analysis of technologies included the structure description and evaluation of existing facial recognition systems, groups of methods and different approaches to face detection, analyzing, processing and recognition were reviewed. The testing stage included a practical test of different algorithms applied on the same standard datasets in order to determine the most efficient algorithm. In the implementation the Support Vector Machine method, which proved itself the one of most precise, was built in the server-side logic of the system. A client-side iOS application allows simplified access to the classifier, it allows its modification — to increase the amount of new data — or its use to verify a person's identity.

A huge stack of technologies being used for the implementation of facial recognition systems. It is important to notice that with the modern growth rates of technologies the number of alternatives that might be used for direct or indirect application in FRS will grow dramatically.

This paper contains information collected, interpreted and structured for facial recognition purposes. It may help to choose an area of further investigation for either creating new approaches, or building a system based on existing methods. Video-based methods of analyzing data should be of interest as it is a wide area of possible investigation. It is also recommended to pay attention to 3D vision methods, technologies that allow a volumetric object to be built within the camera of a phone are already presented and being implemented and distributed in modern devices.

Nowadays, biometric methods of authentication are replacing standard methods of authorization and sooner will cover multiple different areas of human-machine interaction. The implementation and development of an algorithm for video analysis of human behavior patterns may be proposed as a possible continuation of this research.

The author expresses gratitude to Alexey Ryabov, Konstantin Kudelkin and Kirill Babich for fruitful cooperation in creating existing prototypes of the face recognition system, as well as in research on this topic.

References

1. **Singh A., Singh S. K., Tiwari S.** Comparison of face Recognition Algorithms on Dummy Faces, *The International Journal of Multimedia & Its Applications (IJMA),* 2012, vol. 4, no. 4, pp. 121—135.
2. **Sukhija P., Behai S., Singh P.** Face Recognition System Using Genetic Algorithm, *Procedia Computer Science*, 2016, vol. 85, pp. 410—417.
3. **Tareeq S. M., Parveen R., Rozario L. J., Bhuiyan A. A.** Robust Face Detection using Genetic Algorithm, *Information Technology Journal*, 2007, vol. 6, pp. 142—147.
4. **Panigrahy M. P., Kumar N.** Face Recognition using Genetic Algorithm and Neural Networks, *International Journal of Computer Applications*, 2012, vol. 55, no. 4, available at: http://research.ijca-online.org/volume55/number4/pxc3882613.pdf.
5. **Wiskott L., Fellous J. M., Kruger N., Malsburg C. V. D.** Face Recognition by Elastic Bunch Graph Matching, *Intelligent Biometric Techniques in Fingerprint and Face Recognition* / eds. L. C. Jain et al., CRC Press. 1999. Chapter 11, pp. 355—396.

6. **Le T. H.** Applying Artificial Neural Networks for Face Recognition, *Advances in Artificial Neural Systems,* 2011, vol. 2011, pp. 1—16.
7. **Smith L. I.** A tutorial on Principal Components Analysis (Computer Science Technical Report No. OUCS-2002-12), available at: https://ourarchive.otago.ac.nz/handle/10523/7534
8. **Welling M.** Fisher Linear Discriminant Analysis, available at: https://www.cs.huji.ac.il/w~csip/Fisher-LDA.pdf
9. **Raschka S.** Linear Discriminant Analysis, *Bit by Bit 2014*, available at: https://www.researchgate.net/publication/264825166_Linear_Discriminant_Analysis_bit_by_bit
10. **Naseem I., Togneri R., Bennamoun M.** Linear Regression for Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, vol. 32, no. 11, pp. 2106—2112.
11. **Lai J., Jiang X.** Robust Face Recognition Using Trimmed Linear Regression, *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 2979—2983.
12. **Liu C., Wechsler H.** Comparative Assessment of Independent Component Analysis (ICA) for Face Recognition, available at: http://www.face-rec.org/algorithms/ica/liu99comparative.pdf
13. **Das D. K.** Comparative Analysis of PCA and 2DPCA in Face Recognition, *International Journal of Emerging Technology and Advanced Engineering*, 2012, vol. 2, issue 1, pp. 330—336.
14. **Yang M. H.** Face Recognition Using Kernel Methods, NIPS, 2001.
15. **Ji S., Jieping Ye. J.** Generalized Linear Discriminant Analysis: A Unified Framework and Efficient Model Selection, *IEEE Transactions on Neural Networks*, 2008, vol. 19, no. 10, pp. 1768—1782.
16. **Kadoury S.** Face Detection Using Locally Linear Embedding, 2005, available at: https://pdfs.semanticscholar.org/830d/35f4a9d8df629a35a05c514a6c45c64f4f62.pdf
17. **Tseng S.**Comparison of Holistic and Feature Based Approaches to Face Recognition. 2003, available at: https://www.semanticscholar.org/paper/COMPARISON-OF-HOLISTIC-AND-FEATURE-BASED-APPROACHES-Tseng/84bc48c85b2389f700779c1bda8d808dc4972451
18. **Rahim A., Hossain N., Wahid T., Azam S.** Face Recognition using Local Binary Patterns (LBP), *Global Journal of Computer Science and Technology Graphics & Vision*, 2013, vol. 13, issue 4, available at: https://globaljournals.org/GJCST_Volume13/1-Face-Recognition-using-Local.pdf
19. **Jia H. and Martinez A. M.** Support Vector Machines in Face Recognition with Occlusions, *CVPR,* 2009, pp. 136—141.
20. **Li Z., Tang S.** Face Recognition Using Improved Pairwise Coupling Support Vector Machines, available at: https://www.csie.ntu.edu.tw/~cjlin/courses/slt2002/cr1288.pdf
21. **Ishraque S. Z., Jabid T., Chae O.** Face Recognition Based on Local Directional Pattern Variance (LDPv), *Proc. International Conference on Computer Science and Technology,* Jeju, Korea, June 25—27, 2012, pp. 101—106.
22. **Gupta T., Garg L.** Face Recognition Using SIFT, *International Journal of Emerging Technology and Advanced Engineering.* 2014, vol. 4, issue 5, pp. 358—363.
23. **Sadeghipour E., Sahragard N.** Face Recognition Based on Improved SIFT Algorithm, *International Journal of Advanced Computer Science and Applications*, 2016, vol. 7, no. 1, pp. 548—551.
24. **Reda A., Aoued B.** Artificial Neural Network-Based Face Recognition, *First International Symposium on Control, Communications and Signal Processing,* 2004, available at: https://eurasip.org/Proceedings/Ext/ISCCSP2006/defevent/papers/cr1117.pdf
25. **Kasar M. M., Bhattacharyya D., Kim T.** Face Recognition Using Neural Network, *A Review — International Journal of Security and Its Application*s, 2016, vol. 10, no. 3, pp. 81—100.
26. **Chen Q., Kotani K., Lee F., Ohmi T.** Face Recognition Using Self-Organizing Maps, available at: http://cdn.intechopen.com/pdfs/10464.pdf
27. **Yi Jin, Qiu-Qi Ruan.** Face Recognition Using Gabor-Based Improved Supervised Locality Preserving Projections, *Computing and Informatics*, 2009, vol. 28, pp. 81—95.
28. **Wolf L., Hassner T., Taigman Y.** Descriptor Based Methods in the Wild, *Faces in Real-Life Images workshop at the European Conference on Computer Vision (ECCV),* Oct 2008, available at: https://talhassner.github.io/home/publication/2008_ECCV
29. **Ahonen T., Hadid A., Pietikainen M.** Face Description with Local Binary Patterns: Application to Face Recognition, *IEEE Trans Pattern Anal Mach Intell.,* 2006, vol. 28, no. 12, pp. 2037—2041.
30. **Tan J. S., Venkat I., Liao I. Y., Wilde P. D.** General Human Traits Oriented Generic Elastic Model for 3D Face Reconstruction, *BMVC 2016,* available at: https://www.researchgate.net/profile/Iman-Liao/publication/305768198_General_Human_Traits_Oriented_Generic_Elastic_Model_for_3D_Face_Reconstruction/links/57f6116108ae886b89816c4c/General-Human-Traits-Oriented-Generic-Elastic-Model-for-3D-Face-Reconstruction.pdf?origin=publication_detail
31. **Stallkamp J., Ekenel H. K., Stiefelhagen R.** Video-based Face Recognition on Real-World Data*, 2007 IEEE 11th International Conference on Computer Vision.* 14—21 Oct. 2007, pp. 1—8.
32. **Burgos-Artizzu X. P., Perona P., Dollar P.** Robust Face Landmark Estimation under Occlusion, *2013 IEEE International Conference on Computer Vision*, Sydney, Australia, 1—8 December 2013, pp. 1513—1520.
33. **Wojcik W., Gromaszek K., Junisbekov M.** Face Recognition: Issues, Methods and Alternative Applications, 2016, available at: https://www.intechopen.com/chapters/51031