PROBLEMS OF BUILDING A DATABASE IN THE CONTEXT OF HUMAN IDENTIFICATION AND VALIDATION

Haykanush AYVAZYAN
Ph.D. in Economics
Robert HAKOBYAN
Master student at NPUA

Key words: identification, validation, database, features of the face

Introduction. Owing to the development of innovative technologies, miracles disappear from our lives one by one, becoming a daily routine, which in fact is not so bad, because in return we get the opportunity to benefit from these great achievements. And this also applies to face recognition programs.

Face recognition is still a new and little studied field, but already now face recognition programs are in great demand in almost all parts of everyday life. The aim of this work is to study the basic methods of holding and recognizing faces.

Economic importance. In modern organizations, it is very important to ensure the security of the organization's premises, which is why it is often necessary to hire a large number of security personnel. Replacing them with cameras ակսն identification system will avoid human error in the long run, making it possible to minimize security-related losses. The speed of the system is also important in identifying the unwanted person, because in case of slow identification the intruder will have time to harm the organization. This article addresses these issues by offering a database structure that can quickly compare an intruder 's face data with the facial features of authorized people stored in the database and determine if a person entering the site is eligible to be in the area.

Methodology. This is about an online program that can separate a person's face from the whole image or video, which allows you to recognize the person to whom the person belongs. Of course, the program can also recognize other objects in the image or video, but in this work we are talking about a specific human face. Modern facial recognition software programs are able to identify details such as:

- The sex of the person
- Approximate age
- Current state of mind

The face recognition process is done in these three main steps:

1. *The face detection process* is an important step as it detects and places a person's face in pictures and videos.

2. *The process of face fixation* transforms analog information (face image) into a series of digital information (data) based on a person's facial features.

3. *The face matching process* checks to see if two faces belong to the same person.

Today, this type of process is considered the most natural of all biometric measurements. And for good reason, we recognize ourselves not when we look at our fingerprints or irises, but when we see our face. [pandasecurity, 2021]. Before we go any further, let's quickly define two keywords: "identification" and "validation".

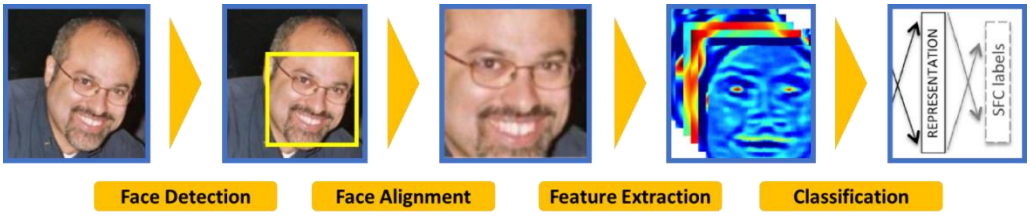


Image 1. The process of face recogniotion

Literature review. Biometrics is used to identify and validate a person, using a set of data specific to that person to be verified.

• Identification answers the question "Who are you?"

• Ratification answers the question "Are you really that person?"

Here are some examples: in case of face biometry, the 2D or 3D sensor "catches" the face. The face image is then converted to digital data using an algorithm before being compared to the images stored in the database.

Automated scanning systems can be used to identify or verify a person's identity in just a few seconds, based on their facial features, cheekbones, nose size, lips, ears, muzzle and other features.

Yes, the process can even take place in a crowd, in a dynamic, volatile environment. Many cell phone owners are already familiar with face recognition software, as it allows them to be identified and validated by the phone.

Of course, there are other means of recognition through the human body: fingerprints, iris examination, voice recognition, digitization of veins in the palm, etc. [thalesgroup, 2021].

What are the features of recognition? When it comes to pattern recognition, the face recognition system has to store hundreds of thousands of feature vectors. The number of vectors depends on the number of users. When identifying a person, the system not only recognizes the presence of a face in the circle, but also identifies the face of the circle and finds the user.

For example, when it comes to face recognition, the system measures features such as eye distance, eye diameter, muzzle sharpness, and many other nuances. There are a total of 128 such features.

The reason for such sweaty measurements is that just face-to-face comparisons are not enough. Because depending on the light, haircut or level of shaving, the face can be identified by different shapes, and such details can affect accuracy.

The main purpose of this work is to identify the best solution for storing all these properties, because the more users there are in the system, the more properties vectors need to be stored.

In addition, it is desirable to note that these features can be applied not only to the face, but also to objects and even sounds. At the same time, the storage methods do not depend on the properties in question [Rattani, Roli, Granger, 2015, 1-8].

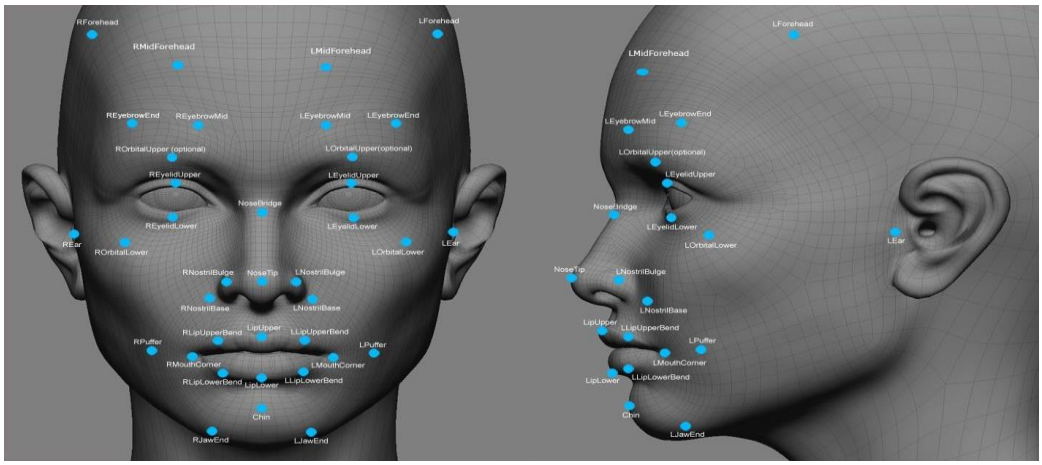
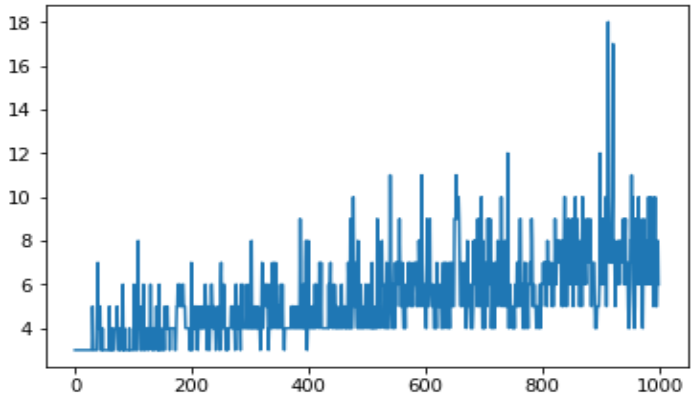


Image 2. Face features

Analysis: What is the complexity of storing feature vectors in recognition systems? The problem with this issue is that the system stores several vectors of each user's features for more accurate recognition. Therefore, our task is to design a database that can store millions of feature vectors. The problem is that no vector of properties will ever be equal to another, that is, complete validation is impossible. In this case, the principle of validation "as close as possible" is used. That is, the input vector has the least deviation from the base vector. This is calculated by the least squares method, ie the sum of the squares of the distances between the vector elements must be minimal.

Thus, there are two problems. The first is the large amount of data, which leads to lengthy information processing. The second problem is the lack of a search algorithm [Fernandez-Saavedra, Sánchez-Reillo, Liu-Jimenez, Miguel-Hurtado, 2015, 240-254].

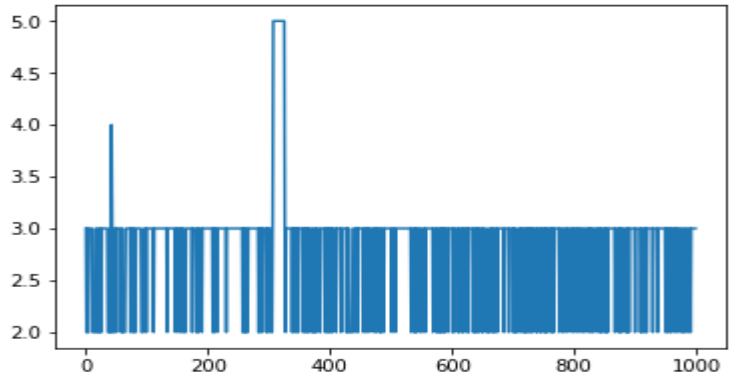
Method 1: In the first method, all the features are serialized in a single text column. Features are contained in one column, and there is a text field that contains an array of 128 features. In the first option, you need to select all the values in the properties column, which takes more time to complete the table.



Graph 1. Distribution of queries by response time (OX – number of requests, OУ - seconds):

As we can see in graph 1, as the number of recordings increases, so does the response time, and the average return time is about 5.3 seconds.

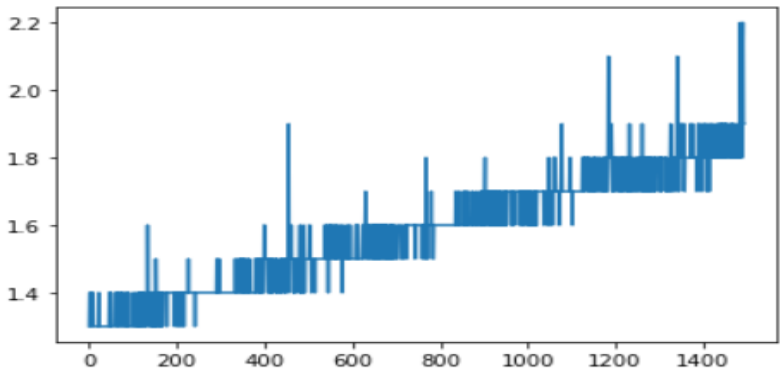
Method 2: In the second method, the properties are stored in 128 columns (f1, f2, f3,… f128)



Graph 2. Distribution of queries by response time (OX number of requests, OУ seconds)

As in column 2, regardless of the number of registrations, the waiting time remains almost unchanged and the response time averages 2.7 seconds.

Method 3: Vectors are stored in multiple tables in a one-to-many relationship. Only vectors (without attributes) are stored in the first table, and all properties of all vectors (one line for each property) are stored in the rest of the table. As you can see in the third graph, the third method is much faster than the first. However, the speed of query responses is increasing. Which, of course, is not a good sign.



Graph 3. Distribution of queries by response time (OX – number of requests, OУ - seconds):

The third method proved to be the most convenient. And although the second method will become faster and more optimized when processing large amounts of data over time, the simplicity and convenience of further use of the database can not be underestimated, which can only be achieved by using the third method.

The Construction of the Database: Thus, the database will be built according to the third method, which is the most convenient for the given work. Figure 3 shows a simplified scheme of the database, where 128 numeric values of facial features will be stored, such as eye size, color, nose size, ear size, lip curvature, muzzle sharpness, etc.

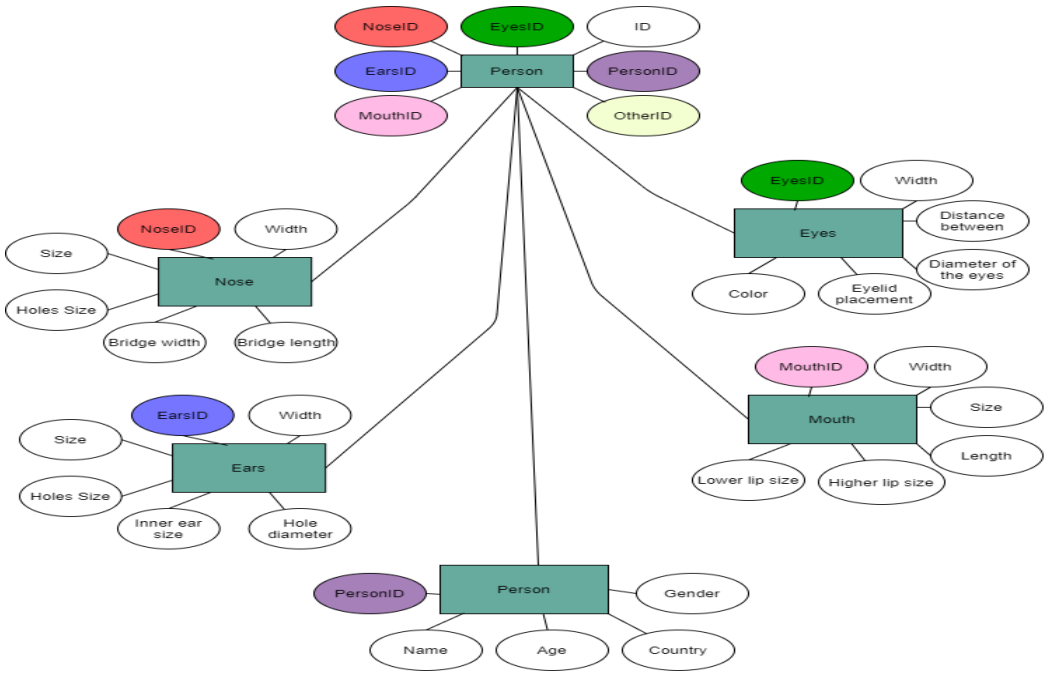


Image 3. Simplified database scheme

The database will also contain general information about the person: age, name, country of residence. Information can be increased or expanded as needed. It is not possible to get a completely accurate value in the face recognition process, so the face to be validated will always be obtained with some deviation from all the values in the database. That is why the principle of "as close as possible" is used in the validation process. That is, the input vector has the least deviation from the base vector. This is calculated by the least squares method, i.e. the sum of the squares of the distances between the vector elements must be minimal. Short information about the person to be returned after recognizing the face: name, gender, age, etc.

Conclusion: Thus, in the course of the work, 3 main approaches to storage in the face database were studied.

The third method was chosen for this task, which requires more time-resources than working with a large amount of data, but instead makes it easier and simpler to work with the database.

References

1. https://www.pandasecurity.com/en/mediacenter/panda-security/facial-recognition-technology/, 2021
2. https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/biometrics/facial-recognition, 2021
3. A. Rattani, F. Roli, E. Granger, Introduction to Adaptive Biometric Systems, 2015, pp 1-8
4. B. Fernandez-Saavedra, R. Sánchez-Reillo, J. Liu-Jimenez, O. Miguel-Hurtado, Evaluation of biometric system performance in the context of Common Criteria, 2015, pp. 240-254

Haykanush AYVAZYAN, Robert HAKOBYAN

Problems of building a database in the context of human identification and validation

Key words: identification, validation, database, features of the face

In this paper, methods are studied for storing information in a database for the purpose of identification and validation, which, in the presence of a large amount of data, will not lose its speed, but at the same time will not suffer from a complex structure and not enough flexibility. It also describes the general appearance of the database tables, which will contain the necessary facial features. Automation of people's identification and validation systems in organizations can lead to long-term profits, as the need to have a large number of security personnel will be lost. The speed of the system is also important in identifying the unwanted person, because in case of slow identification the intruder will have time to harm the organization. This article addresses these issues by offering a database structure that can quickly compare the intruder's face data with the features of the authorized people stored in the database and determine whether the person entering the area has the right to be in the area or not.