Research normal distribution, the trust interval and calculation of importance degree of geometrical characteristics a human face on the basis of photo portraits

Tofiq Kazimov and Shafagat Mahmudova

Institute of Information Technology of ANAS, Baku, Azerbaijan

Az1141 tofig@mail.ru, shafagat_57@mail.ru

Contact Author: Shafagat Mahmudova

BIOCOMP’13

Abstract— in article algorithm of automatic identification of persons on the basis of their photographs are considered. For identification of persons, the comparative analysis of control systems by bases of images created in the different periods is carried out and their applied possibilities are shown. The questions of definition of anthropometrical points of the person calculation of values of geometrical characteristics and their confidential intervals, algorithm of automatic addition of geometrical characteristics in a database and some other questions connected with them are described. To determine trust interval using Student method, first of all, subordination of the law of normal distribution of selection consisting of geometrical characteristics was investigated. The paper offers a new algorithm to find coefficients which determine importance degree of the values of geometrical characteristics used to identify a human face on the basis of photo portraits.

Keywords: recognition, anthropometrical, identification, geometrical characteristics, confidential an interval, importance degree, coefficient.

1 INTRODUCTION

Modern information and communication technologies (ICT) enable the development of various areas of great importance, as well as of biometric technology. The expansion of the fields of application of these technologies plays an important role in preventing a number of dangerous incidents. It is obvious that the prevention of dangerous manifestations, such as the prevention of international terrorism, transnationals organized crime, as well as illegal weapon and drug transportation is one of the main duties of each state. One of the methods in detecting and neutralizing hazardous manifestations is just the advantages of biometric identification technologies. Biometric technologies particularly strengthen reliable control passport-visa control and other identification documents.

Information on the dynamics of biometric technology market given by the world-famous International Biometric Group, gives a way to say: Taking into account the unique characteristics of a person chosen separately, biometric technology was organized on the basis of biometrics [1, 8, 9].

People differ significantly from each other for the sizes and the arrangement of such face elements as eyes, eyebrows, noses, ears, mouths, etc. Therefore, the first approach to the problem solution of automatic person face identification by photo portrait was based on the selection and comparison of some anthropometric face peculiarities. This method has been used in experimental criminalities for years. This technique was especially effective when a person did not have a photograph except the one in a passport [2].

Paper [3] is devoted to the recognition of a human face on the basis of a photo portrait. For face recognition based on a photo portrait, the authors developed 19 anthropometric face points. These points are chosen from the point of resistance to slight changes (caused by the angle, light, facial expression, cosmetics, age, and so on). An algorithm has been developed for calculating the values of the distances between these points and of geometrical characteristics of the human face. It is shown that the difference of the developed algorithm from the other existing ones is that compared with the other photos stored in the database it works even in the absence of any other information about the person except the image described in the photo [10, 11].

The development principles of “Recognition” biometric identification system (RBIS) are explained on the basis of algorithm given to identify a human on the basis of photo portraits in paper [4], and a database with a developed structure is organized for it. Various sized images of $n$ persons and individual data for each (first name, middle name, last name, date of birth, eye colour, height and etc.) were included in the database. The paper also describes an algorithm for default addition of the values of the geometrical characteristics, for search and identification of an image of a human face on the basis of photo portrait in the database [12, 13].

2 PROBLEM STATEMENT
Definition of accuracy of identification is a very important aspect of human face recognition based on photo-portrait. Definition of trust interval of geometrical characteristics is one of the main factors of recognition of a human face based on photo-portrait. In this article, we considered the process of detection of trust intervals of geometrical characteristics used in identification of human face based on photo-portraits. In works [4] [13], trust interval is a range containing the real value of parameter studying on existing reliability level during its main collection.

Advantage of experiments carried out with definition of trust interval to those performed without, are as following:

1. Insignificant effect of ultimate factors to main process during performance of experiments;
2. Faster and more accurate performance of experiments;
3. Performance of necessary number of observations during experiments.

In order to define the trust interval, values of geometrical characteristics calculated through “Recognition” identification system we used [4]. For this reason, initially, experiments were carried out on geometrical characteristics stored in the data base belonging to 102 people. Geometrical characteristics belonging to 102 people were distributed among 18 clusters based on identical characteristics.

2 situations can occur while working with RABIS:

1. Values of geometrical characteristics can be included in trust interval detected for them. In this case, system will continue its operation and pass on to the next level.
2. Values of some geometrical characteristics may not be included in trust interval included for them. In this case the system alerts the user and the distances suitable for them are re-calculated in order to determine those geometrical characteristics and process is continued.

Generally, values of 102*18 numbers of geometrical characteristics were used in order to calculate the trust interval [4]. Clusters by geometrical characteristics are indicated as Ns1, Ns2, and Ns18. Student method was used in order to define the trust interval [6,17].

To determine trust interval, used in human image identification on the basis of photo portrait, using Student method, first of all, subordination of the law of normal distribution of selection consisting of geometrical characteristics should be investigated.

Determination of normal distribution is of great importance for various reasons. In most cases, it is considered to be the best convergence to the function. The statistical distribution of numbers of natural phenomena is considered normal. For example, for determining the weight or volume of the goods, and measurement of height of the person having medical check-up, etc.

A new algorithm is proposed in this paper to find the coefficient which determines importance degree of the values of geometrical characteristics. Let us explain the essence of the algorithm. The values of geometrical characteristics of a quantity used for the identification are divided into the clusters of m quantity for the same sign. To determine the importance degree of the values of geometrical characteristics an identification process is carried out temporarily replacing each value of geometric characteristics of each person with the other values taken from the replacement interval, and the impact of the replacement in the recognition process is assessed.

3 Research Normal Distribution The Trust Interval

Normal distribution is defined by the main 2 parameters: average and standard error [16]. Finite population consisting of 17 random selection values within the geometrical characteristics values. The average value is calculated on the basis of finite population. As the result of calculations the average value \( \mu = 2.45787 \) obtained. Selection distribution was sorted out in finite selection on the based 2 values and per each selection average value calculated on the basis of them. The frequency of average values and their sum have been calculated and shown in Table 1.
TABLE 1. The frequency of average values and their sum

<table>
<thead>
<tr>
<th>Frequency, ( f )</th>
<th>( \bar{x} f )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sum f = 135 )</td>
<td>( \sum \bar{x} f = 331.662269 )</td>
</tr>
</tbody>
</table>

Mathematical expectancy on the basis of selection distribution has been calculated by the following formula:

\[
E(\bar{x}) = \frac{\sum \bar{x} f}{\sum f} = \frac{331,662}{135} = 2,46
\]

The average value of selection data

\[
\mu = 2,45787 = 2,46
\]

Thus,

\[
E(\bar{x}) = \mu \approx 2,46
\]

Standard error for the main normally distributed population has been determined by the following formula [16].

\[
SE(\bar{x}) = \sqrt{\frac{(N-n)\sigma^2}{(N-1)n}}, \quad (1)
\]

\( \sigma^2 \) indicates the main dispersion.

For the values of main geometric characteristics

\[
\sigma^2 = 0,3551, \quad N = 102, \quad n = 2
\]

If the measure of the main population more than the selection population \( (n/N = 0.0196 < 0.05) \) then

\[
\sqrt{\frac{(N-n)}{(N-1)}} = 0,995 \approx 1
\]

And standard error equals to:

\[
SE(\bar{x}) = \sqrt{\frac{\sigma^2}{n}} = 0,2511
\]

As it is obvious from the calculations, normal distribution of values of geometrical characteristics has been proved (Figure 1).
\( t_\beta \) - Student factor selected from the schedule.

\[ x_1^*(n=1,n) \] - indicates the elements of main collection.

While using formula (2) for cluster \( N_{s1} \),

Confidence interval of value of geometric characteristics used for human face recognition based on photo portrait was determined by the Student method.

Using fuzzy calculation, confidence interval of the values of the geometrical characteristics can be determined. It should be noted that on the basis of the research carried out by the authors, real interval values of the distances between anthropometric points of a human face have been established. With the help of fuzzy calculation, to find interval values of the geometrical characteristic in accordance with the same distances

\[
S_i^*/S_{i,1} \leq P_i^* \leq S_i/S_{i,1} \quad i = \overline{1,n}
\] (5)

formula was used.

Here the real maximum value of (i) anthropometric distance of a human face was indicated as - \( S_i^* (i = \overline{1,n}) \), and the real minimum value as - \( S_i (i = \overline{1,n}) \).

The value of geometric characteristics found with the help of calculation was indicated as - \( P_i^* (i = \overline{1,n}) \).

The real interval values of geometric characteristics found through (5) have been shown in table 3.

When compared \( P_i^* (i = \overline{1,n}) \) values in table 3 with \( P_i (i = \overline{1,n}) \) values in table 2 the following terms will be ensured.

\[
\text{Max} (P_i) \leq \text{Max} (P_i^*) \quad i = \overline{1,n}
\]

\[
\text{Min} (P_i) \geq \text{Min} (P_i^*) \quad i = \overline{1,n}
\]

It has been cleared out that, the values of \( P_i \) don’t exceed beyond the real value interval.

The paper [6] provides information about the algorithms developed for normal distribution of the values of geometrical characteristics used in the recognition of a human face on the basis of photo portrait and to define trust interval of the geometrical characteristics. It is shown that the determination of normal distribution of geometrical characteristics is of great importance for various reasons. \( m \) selection value is randomly taken from the values of geometrical characteristics and its normal distribution is investigated [6, 14, 13, 15].

The paper [5] provides information about the algorithms developed to define trust interval of the values of geometrical characteristics in the recognition of a human face on the basis of photo portrait. On the basis of the conducted researches, the real interval values of distances between the anthropometric points of a human face are established. With the help of fuzzy calculation, interval values of geometric characteristics values, proper to the same points, are found [6, 7, 15].

Finding coefficients which determine importance degree of the values of geometrical characteristics used to identify a human face on the basis of photo portraits is of great importance for the recognition process from the various views. Determination of coefficients determining importance degree of the values of geometrical characteristics for identification leads to the reduction of the number of values of insignificant geometrical characteristics, as well as to the improvement of identification quality and to the decrease of time spent for the identification.

4 IMPORTANCE DEGREE

ALGORITHMS

Let us mark the wanted photo portrait possessing geometric characteristics of \( m \) quantity, i.e., \( m \) sized point with \( F^* (p_1^*, p_2^*, \ldots, p_m^*) \), photo portraits in the database with \( F_i (p_{i1}, p_{i2}, \ldots, p_{in}) \), \( i = \overline{1,n} \).

If we mark \( F_i \) points and distances of \( F^* \) point with \( S_i \), then

\[
S_i(F^*,F_i) = \sum_{k=1}^{m} (p_k^* - p_{ik})^2 , \quad i = \overline{1,n}
\] (6)

Where \( n \) indicates the number of photo portraits in the database. Let us divide
geometrical characteristics of the photo portraits in the database into $K_j \ (p_y, i = 1, n), \ (j = 1, m)$ clusters of $m$ quantity. If we mark replacement intervals of the parameters of each $K_j \ (j = 1, m)$ cluster with $[\alpha_j, \beta_j], \ (j = 1, m)$, then inequality
\[ \alpha_j \leq p_y \leq \beta_j, \quad i = 1, n, \ j = 1, m, \quad (7) \]
can be right for any $p_y$.

To determine importance degree of geometrical characteristics of each photo portraits, i.e. of $p_y$ parameters for the identification, let us divide $[\alpha_j, \beta_j]$ cluster with $\jmath = 1, m$ interval of each $K_j \ (j = 1, m)$ cluster into the equal $10$ parts by $h_j$ step.
\[ h_j = (\beta_j - \alpha_j) / t, \ j = 1, m, \quad (8) \]
\[ x_{jk} = \alpha_j + kh_j, \quad t - \text{is an integer number}. \]

Replacing $x_{ij} \in [\alpha_j, \beta_j], \ (k = 0, t; \ j = 1, m)$ points consistently instead of the values of $l$ the $(l = 1, m)$ coordinates of $F_*$ points, we achieve the point of $(t + 1)m$ quantity. Let us mark them with $FT_{kl} \ (k = 0, t; \ l = 1, m)$.

Let us calculate $\omega_j$ coefficient indicating $ST_{kl} (F_i, FT_{kl})$ distances between these points and $F_i, \ (i = 1, n)$ points in the database and determining the efficiency degree of geometrical characteristics.
\[ \omega_j = \left( \frac{1}{m(t+1)} \sum_{i=1}^{m} \sum_{j=0}^{t} (ST_{kl} - S_{li})^2, \right)^{-1}, \ j = 1, m, \quad (9) \]

Note that, calculating the distance between the other photo portraits existing in the database and the two points in 16-dimensional space, the photo portrait of any person is compared with the following formula in the work [5].
\[ S_i(f^*, F_i) = \sqrt{\sum_{i=1}^{m} (p_i - p_k)^2, \ i = 1, n}, \quad (10) \]

In this paper, the formula (10) is replaced with the formula (6). The aim of the replacement is to accelerate the identification process and to reduce the time spent for the identification. Including the coefficient which determines importance degree of the values of geometrical characteristics, into the formula (6), we can increase the importance of the recognition and may not take into account insignificant geometrical characteristics. When in a database it is too many records, then this replacement very important.

Including the coefficient (9) into the formula (6), the following distance formula is achieved:
\[ S_i(f^*, F_i) = \sum_{i=1}^{m} \omega_j (p_i - p_k)^2, \ i = 1, n; \ j = 1, m. \quad (11) \]

5. EXPERIMENTAL TEST

\[ n = 102, \ \ k = n, \ \ \beta = 95\%, \quad \]
\[ t_{\beta} = 1.98, \quad \]
\[ m_{\beta} = \approx \pm 0.0353, \quad \]

We will get
\[ 2.21232796 \leq X \leq 2.35227203 \]

Here Student coefficients were taken based on the value of $t_{\beta}$ and values of $k$ and $\beta$.

In such manner, values of trust interval in accordance with other clusters are calculated and shown in table 2.
TABLE 2. Trust intervals of geometrical characteristics used in identification of human face based on photo portrait in accordance with other clusters

<table>
<thead>
<tr>
<th>$Ns_1$</th>
<th>$Ns_2$</th>
<th>$Ns_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2.21\leq 2.28\leq 2.35$</td>
<td>$0.05\leq 0.42\leq 0.79$</td>
<td>$5\leq 5.76\leq 6.51$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$Ns_4$</th>
<th>$Ns_5$</th>
<th>$Ns_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.18\leq 0.22\leq 0.62$</td>
<td>$0.9\leq 1.13\leq 1.35$</td>
<td>$0.4\leq 0.79\leq 1.08$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$Ns_7$</th>
<th>$Ns_8$</th>
<th>$Ns_9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1.94\leq 2.03\leq 2.11$</td>
<td>$0.25\leq 0.59\leq 0.92$</td>
<td>$1.21\leq 1.23\leq 1.25$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$Ns_{10}$</th>
<th>$Ns_{11}$</th>
<th>$Ns_{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.33\leq 0.665\leq 0.98$</td>
<td>$1.60\leq 1.61\leq 1.63$</td>
<td>$3.2\leq 3.243\leq 3.27$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$Ns_{13}$</th>
<th>$Ns_{14}$</th>
<th>$Ns_{15}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.25\leq 0.29\leq 0.33$</td>
<td>$0.01\leq 0.2935\leq 0.74$</td>
<td>$5\leq 5.76\leq 6.51$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$Ns_{16}$</th>
<th>$Ns_{17}$</th>
<th>$Ns_{18}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.18\leq 0.22\leq 0.62$</td>
<td>$0.33\leq 0.66\leq 0.98$</td>
<td>$1.60\leq 1.61\leq 1.63$</td>
</tr>
</tbody>
</table>

Figure 2. Trust intervals of geometrical characteristics used for identification of human face based on photo portrait

TABLE 3. The real minimum and maximum values of geometrical characteristics

<table>
<thead>
<tr>
<th>$P_1^*$</th>
<th>$P_2^*$</th>
<th>$P_3^*$</th>
<th>$P_4^*$</th>
<th>$P_5^*$</th>
<th>$P_6^*$</th>
<th>$P_7^*$</th>
<th>$P_8^*$</th>
<th>$P_9^*$</th>
<th>$P_{10}^*$</th>
<th>$P_{11}^*$</th>
<th>$P_{12}^*$</th>
<th>$P_{13}^*$</th>
<th>$P_{14}^*$</th>
<th>$P_{15}^*$</th>
<th>$P_{16}^*$</th>
<th>$P_{17}^*$</th>
<th>$P_{18}^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.42</td>
<td>0.04</td>
<td>2.8</td>
<td>0.1</td>
<td>0.352</td>
<td>0.4</td>
<td>0.68</td>
<td>0.1</td>
<td>1</td>
<td>0.2</td>
<td>0.21</td>
<td>0.33</td>
<td>0.54</td>
<td>0.01</td>
<td>2.8</td>
<td>0.1</td>
<td>0.23</td>
<td>0.21</td>
</tr>
<tr>
<td>2.8</td>
<td>0.8</td>
<td>6.7</td>
<td>1</td>
<td>1.79</td>
<td>1.29</td>
<td>2.22</td>
<td>1.6</td>
<td>2.031</td>
<td>3.33</td>
<td>1.73</td>
<td>3.56</td>
<td>2.4</td>
<td>1</td>
<td>6.8</td>
<td>1</td>
<td>1.33</td>
<td>1.8</td>
</tr>
</tbody>
</table>

As it is mentioned above, a large number of experiments have been carried out at TBIS on the basis of above mentioned algorithm in order
to calculate the coefficients which determine importance degree of the values of geometrical characteristics used to identify a human face on the basis of photo portraits.

**TABLE 4.** Values found on the basis of given algorithms

<table>
<thead>
<tr>
<th>P11</th>
<th>P12</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.21</td>
<td>0.09 0.09</td>
</tr>
<tr>
<td>2.22</td>
<td>0.08 0.08</td>
</tr>
<tr>
<td>2.24</td>
<td>0.06 0.06</td>
</tr>
<tr>
<td>2.25</td>
<td>0.05 0.05</td>
</tr>
<tr>
<td>2.27</td>
<td>0.03 0.03</td>
</tr>
<tr>
<td>2.28</td>
<td>0.02 0.02</td>
</tr>
<tr>
<td>2.29</td>
<td>0.01 0.01</td>
</tr>
<tr>
<td>2.31</td>
<td>0.01 0.01</td>
</tr>
<tr>
<td>2.32</td>
<td>0.02 0.02</td>
</tr>
<tr>
<td>2.34</td>
<td>0.04 0.04</td>
</tr>
<tr>
<td>2.35</td>
<td>0.05 0.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P13</th>
<th>P14</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.21</td>
<td>0.21</td>
</tr>
<tr>
<td>2.22</td>
<td>0.22</td>
</tr>
<tr>
<td>2.24</td>
<td>0.24</td>
</tr>
<tr>
<td>2.25</td>
<td>0.25</td>
</tr>
<tr>
<td>2.27</td>
<td>0.27</td>
</tr>
<tr>
<td>2.28</td>
<td>0.28</td>
</tr>
<tr>
<td>2.29</td>
<td>0.29</td>
</tr>
<tr>
<td>2.31</td>
<td>0.31</td>
</tr>
<tr>
<td>2.32</td>
<td>0.32</td>
</tr>
<tr>
<td>2.34</td>
<td>0.34</td>
</tr>
<tr>
<td>2.35</td>
<td>0.35</td>
</tr>
</tbody>
</table>

In particular cases, using the values $n = 102, m = 18, t = 10$ the values proper to the values of the geometrical characteristics of 10 persons in accordance with the 1st value of the cluster have been calculated through the formulas (6), (8), (9) and (11). Some of the results of the conducted experiments are shown in table 4, figure3, figure.4, figure.5, figure.6 in the form of graphics.

**TABLE 5.** Coefficients determining importance degree of geometrical characteristics (for the values of the 1st, 2nd, 3rd geometrical characteristics)

| $\omega_1$ | 1.39 |
| $\omega_2$ | 0.70 |
| $\omega_3$ | 0.80 |

The values proper to the given formulas have been calculated for other persons in this way, as well.

**Figure 3.** Comparison of the distance values found in accordance with the values of trust intervals proper to the values of the 1st geometrical characteristics (10 persons)

**Figure 4.** Comparison of the distance values found in accordance with the values of trust intervals proper to the values of the 1st geometrical characteristics (females)
Figure 5. Comparison of the distance values found in accordance with the values of trust intervals proper to the values of the 1st geometrical characteristics (males)

Figure 6. Comparison of the distance values found in accordance with the values of trust intervals proper to the values of the 2nd geometrical characteristics (10 persons)

102 * 19 * 5 (9690) experiments have been carried out on the basis of the data of 102 persons through TBIS. The values \( \omega_j (j = 1, m) \) are shown in the table 5.

The authors have established biometric identification system in accordance with anthropometric points of a human face by the photo portrait on the basis of obtained scientific results.

The software system is capable to detect the most similar faces comparing any photo portrait of any person uploaded to the system with other existing ones in the base. Note that the rumours regard to the identity of the hero of the mysterious "Mona Liza" by the prominent Italian artist Leonardo da Vinci is still not calming down. The disputes in connection with who is described in the portrait have been going on over more than 500 years.

The portraits of Leonardo da Vinci (figure 7) and Mona Liza (figure 8) painted in different years were included in the system database by the authors as an experiment. Two versions of identification process were carried out through the system. In the 1st version the portrait of Mona Liza was included in the system base for identification and compared with the other ones existing in the database. Initially, the most similar portraits were Mona Liza (100%) and the portrait of Leonardo da Vinci (99.5%). In the 2nd version the portrait by artist was included in the system for identification. In this case, the most similar portraits were the portrait of Leonardo da Vinci himself (100%), and then the portrait of Mona Liza (99.5%).

5 Result

Student method was used for definition of trust intervals of geometrical characteristics used for identification of human face based on photo portrait. Definition and use of trust intervals results in fast and effective operation of human face identification program. This, results in preventing time loss during identification.

A new algorithm has been proposed to find coefficients which determine importance degree of the values of geometrical characteristics used to identify a human face on the basis of photo portraits:

1. A formula is given to calculate distances between the wanted photo portraits possessing \( m \) number of geometric characteristics with the points of photo portraits in the base;
2. A formula is given to calculate a step in appropriate intervals of each cluster in order to determine importance degree of the values of geometrical characteristics of each photo portrait for identification;
3. A formula is given to calculate \( \omega_j (j = 1, m) \) coefficients determining importance degree of the values of geometrical characteristics;
4. Including \( \omega_j (j = 1, m) \) coefficients a distance formula is given for the identification.

The given algorithm leads to the reduction in the number of values of geometrical characteristics used for identification, as well as to the improvement of identification quality and to the decrease in time spent for the identification.

Figure 7. Leonardo Da Vinci

Figure 8. Mona Liza
REFERENCES


