Solution of transport safety problems based on biometric systems with error-correcting coding procedures

A A Gladkikh¹, A K Volkov², A K Volkov², M A Y Damdam¹ and Ali S A Al-Mesri¹

¹Ulyanovsk State Technical University, 32, Severnyjvenec, Ulyanovsk, 432027, Russia
²Ulyanovsk Civil Aviation Institute, 8/8, Mozhaisky Street, Ulyanovsk, 432071, Russia

E-mail: volkovalex8@rambler.ru

Abstract. The paper considers models and methods for solving processing biometric data at the physical and channel levels. A classification of the parameters of such systems is given and those characteristics that play a decisive role in achieving the objective function in limited time intervals are distinguished from them. The work aims to analyze the possibilities of increasing the reliability of biometric devices at the channel level by using error correction codes. The practicability and necessity of error-correcting coding mean using the permutation data decoding algorithms capable of matching biometric identification adequately. Based on the results of statistical modelling of optical communication channels, the necessary and sufficient conditions for the application of permutation decoding tools for binary codes are determined. The problem of minimizing the memory size of the cognitive map of a permutation decoder is solved by allocating the permutation orbits and using the generating combinations of loops as indexes of the reference matrices. An efficient algorithm is proposed for finding a unique orbit number and the corresponding reference matrix when the receiver generates an arbitrary permutation of characters from the set of permissible permutations.

1. Introduction

The solution to most of the promising transport safety tasks is directly related to the methods of biometric identification of the passengers passing through different control forms at security checkpoints in airports. Similar tasks, but only of a more complex nature, arise in the case of searching and identifying individuals by characteristic signs in the “smart” city system, which is provided with an integrated video stream analysis system and an appropriate data processing centre. Moreover, recent experience in localizing and eliminating the consequences of the COVID-19 pandemic shows that the methods of biometric identification of subjects are already a sufficient basis for identifying potential carriers of dangerous infections. In any case, the effectiveness of biometric means substantially depends on the time limits for making a decision. Therefore all components of the identification information process should contribute to the achievement of the objective function in a given time interval. The analysis shows that the most expensive in this respect is the process of comparing the obtained biometric template with the existing database, which in turn can be located at a considerable distance from the biometric system. Loss or distortion of information during its transmission over the communication channel is critical and not acceptable for such systems.
Thus, long-term storage and timely provision of information is a key function of biometric identification devices. The work aims to analyze the possibilities of increasing the reliability of such devices at the channel level by using error correction codes.

2. Time constraints and probabilistic characteristics of biometric systems
The classification of biometric systems is well known. Naturally, contact systems for realizing the identification of people in a stream are of little use. Therefore, it is most rational to use face recognition of such dimensions as 2D or 3D in such systems. By modern standards, a minimum time interval is required to fix only 12 to 40 parameters. Fixing one parameter requires calling the coordinates of two points on the face image and estimating the distance between them, which require up to 6 bytes of data. Thus, the minimum amount of data can be about 72 bytes, and the maximum amount—up to 240 bytes. The indicated volumes are not critical for modern data exchange systems, but the use of traditional means of error-correcting coding to protect data from interference can lead to time delays. In the articles [1, 2], to improve the reliability of functioning of biometric systems, it is proposed to use the method of permutation decoding (PD) with elements of cognitive data processing. Due to the use of the cognitive card of the decoder (CCD) the approach allows replacing the computational process with a search for a finished result in a card with a small number of additional actions.

The main probabilistic characteristics of biometric systems are false pass coefficient (second-order error) — the probability that the system incorrectly compares the input sample with an inappropriate template in the database; false rejection ratio (error of the first kind) — the probability that the system will be mistaken in determining matches between the input sample and the corresponding template from the database. It is noticeable that these characteristics also directly depend on the correcting properties of the used redundant code and the properties of the used communication channel. In the articles [3-5], some results of applying noise-resistant codes in biometric systems are presented.

3. Necessary and sufficient conditions for the implementation of PD
Virtually none of the known data transmission systems using redundant coding tools allows the decoder training procedure achieving the effect of accelerating the processing of received data. The receiver, with rare exceptions, is not able to predict the behavior of the information source, which can send any of a given set of combinations $I(t)$ through the interference channel. In addition, a priori, an unknown interference vector $Er(t)$ acts in the communication channel, which together with the value $I(t)$ enhances the total uncertainty $Rn(t)$ for the receiver. The only method for processing data at the reception, for which a certain deterministic component $Dn(t)$ can be indicated in the decoding process, is the PD method described in detail in [2].

The deterministic component $Dn(t)$ teaches the decoder to recognize specific permutations and issue ready-made data processing decisions depending on the situation in time for selecting reliable characters. Learning outcomes are usually recorded in the memory of the decoder, which is implemented in the format of the CCD. The memory capacity of the card depends on the dimension of the code. Since it is supposed to use non-binary code constructions, the question arises of the rational organization of such memory. The CCD takes into account possible permutations of symbols of code vectors, which directly depend on the distribution of reliability evaluation for symbols. A similar problem is solved due to the obligatory procedure for a soft processing of binary or non-binary characters (depending on the expansion of the Galois binary field, in which the redundant code vectors are processed) and selection of the most reliable characters from the tuple for the formation of the equivalent code (EC), which is actually a sufficient condition for the implementation of PD. A prerequisite for implementing the acceleration of the data decoding procedure is the presence of the CCD. Since the EC is formed on the receiving side, the influence of interfering factors is at the level.
of internal processor failures of the receiver. It is advisable to develop soft symbol values for dual types of modulation based on an erasing communication channel with a wide erasure interval under an analytical expression

\[ \lambda_i(z) = \left\lfloor \frac{\lambda_{\text{max}}}{\mu \sqrt{E_b}} \times z_i \right\rfloor, \quad npu \ 0 \leq z_i \leq \mu \sqrt{E_b}, \quad (1) \]

where \( \lambda_{\text{max}} \) – the maximum value of the soft solution (specified by the constructor); \( \mu \) – erasure interval value (usually \( 0 \leq \mu < 1 \)); \( E_b \) – signal energy per bit; \( z_i \) – the actually recorded signal value (as a rule, due to the influence of destructive factors it differs from the nominal one). It is important to note that in expression (1) there are no parameters characterizing the property of the communication channel, for example, in the format of noise dispersion. For non-binary codes, it is advisable to use the histogram comparison method.

4. The results of statistical modeling of the soft decision formation system

We evaluate the feasibility of PD for some codes. Let the code be the maximum decoded code for which \( d_{\text{min}} = n - k + 1 \). The asymptotic estimate of the energy gain of a code for similar codes is determined by the expression \( D_{\text{max}}(k) = 10 \lg(k - \frac{k^2}{n} + \frac{k}{n}) \). The estimate of the extremum of this function leads to an expression of the form \( 2k = n + 1 \). This means that the AP is effective when the length of the parameter \( k \) is half the code vector \( n \). The test results are presented in Figure 1.

Figure 1. Frequency of occurrence of soft decision indices in a code length vector \( n = 15 \), where: a) \( \rho = 0.6 \); b) \( \rho = 0.7 \); c) \( \rho = 0.8 \); d) \( \rho = 0.9 \).
Fundamentally the erasing communication channel can be replaced by a system with polar coding and a fuzzy set, but the evaluation of such a solution requires additional research.

5. The cyclic properties of permutation and their orbits

When organizing the PD procedure, the decoder, according to some criterion, must select $k$ reliable characters from $n$ accepted and in the general case $C^n_k$ such combinations can be created. It is important to note that in the set of combinations you can notice certain patterns that are cyclical in nature. In combinatorics, such structures are called orbits. For each orbit, you can specify the smallest number that underlies the cyclic shift of the combination of $k$ non-repeating elements (numbers). We call such a number the generating combination of the cycle (GCC). Adding a number to the GCC, you can specify the orbit number in the general system of permutations, which in the future it is advisable to associate with a specific generating matrix EC, which underlies the AP system. Since permutations represent a group, it seems convenient to associate this group with group codes that are formed over Galois binary fields or their extensions. Therefore, the developed material of the PD system is suitable for both binary codes and non-binary redundant codes.

Let there be given a Hamming code (7, 4, 3) as a binary group. We number the elements of the code combination of the code from 1 to $n = 7$ and call such numbers the position numerators. In this case, for the code in question, it is permissible to form 35 permutations, which are conveniently divided into orbits from their GCC as shown in Table 1.

| Table 1. Permutation orbits for code (7, 4, 3) |
|---------------------------------------------|
| 1234 GCC$^1$ | 1236GCC$^2$ | 1245 GCC$^3$ | 1246 GCC$^4$ | 1235 GCC$^5$ |
| 2345 | 2347 | 2356 | 2357 | 2346 |
| 3456 | 1345 | 3467 | 1346 | 3457 |
| 4567 | 2456 | 1457 | 2457 | 1456 |
| 1567 | 3567 | 1256 | 1356 | 2567 |
| 1267 | 1467 | 2367 | 2467 | 1367 |
| 1237 | 1257 | 1347 | 1357 | 1247 |

It is noticeable that the table uses the right moves along the cycle, which we call direct. In the case of using the left motion along the cycle, we call it inverse. The weight of any permutation is the sum of the numbers in it. Then several properties of an arbitrary GCC, which play an important role in minimizing the amount of memory of CCD, become apparent.

**Property 1.** Any GCC has a minimum weight value (the sum of the permutation digits) relative to the weights of all other elements of the orbit that it forms. This property is explained by the fact that it is the GCC numbering boxes that occupy left positions in the series of cycle numbers from 1 to $k$, and the movement to at least one step to the value of $n$ only increases the permutation weight.

**Property 2.** In the case of a complete cycle, the total number of permutations of any orbit starting with the number 1 is always equal to the value of $k$. This is easily proved by the composition of the elements of the orbits in the full cycle.

**Property 3.** The rule of minimizing the weight of the permutation by attribute 1 in the left category allows realizing an accelerated search for the GCC and therefore precisely name the orbit number. Let the contents of only the top line be known from table 1 to the decoder in CCD, and let an arbitrary permutation be given in lexicographic format. By applying the inverse loop rule and minimizing the permutation, you can always find a permutation with minimal weight.

Example. Let for the code (7, 4, 3) a permutation of the form 7641. Its lexicographic form has a value of 1467, and since the combination is led by the number 1, due to the inverse cycle, we proceed to the permutation 4678 and minimize it, leading to the form 4678-3333=1345. The weight of the original permutation (the sum of all digits) is 18. The weight of the resulting new permutation is 13.
Remember: 18 > 13. The transformation cycle continues. The permutation 1345 due to the reverse cycle is converted to the form 3458. Next, 3458-2222 = 1236, the weight of the permutation 12. Remember: 13 > 12, the sign of inequality does not change. Next, we calculate: 2368-1111 = 1247. The permutation weight is 14. We remember: 12 < 14 and state that the sign of inequality has changed. Therefore, by property 1, the GCC for the obtained permutation is GCC\(^2\) = 1236 (see Table 1).

Using these properties allows reducing the volume of CCD \(n\) times and Table 1 takes on the form shown in Table 2.

| GCC\(^1\) | GCC\(^2\) | GCC\(^3\) | OKC\(^4\) | GCC\(^5\) |
|----------|----------|----------|---------|----------|
| 1234     | 1236     | 1245     | 1246    | 1235     |

By associating a specific GCC with the corresponding generating matrix and transforming it according to cyclic shifts, it is easy to obtain the generating matrix of the EC for the arbitrary permutation.

6. Conclusion
PD is a kind of soft decoding of block error-correcting codes and has doubtless advantages in terms of energy gain of the code regarding hard data decoding methods. PD is based on the calculation by the receiver for each received combination and the EC vector transmitted over the channel with interference. The complexity of the computational process for classical algorithms is unacceptably high. From a practical point of view, the situation changes dramatically for the better with the introduction of cognitive methods for data processing, when the memory volume of the CCD replaces a complex computational process.

Acknowledgments
The reported study was funded by RFBR and INSF, project number 19-47-730011.

References
[1] Gladkikh A A, Volkov An K, Volkov Al K and Ibragimov R Z 2019 Journal of Physics: Conference Series (ITBI 2019) \textbf{1333} 1-5
[2] Gladkikh A A, Volkov An K, Volkov Al K, Andriyanov N A and Shakhtanov S V 2019 Proceedings of the International Conference on Aviamechanical Engineering and Transport (AviaENT 2019) \textbf{188} 373-379
[3] Hao F, Anderson R and Daugman J 2006 IEEE Trans. Comput. \textbf{55}(9) 1081-88
[4] Alawi A Al-Saggaf 2018 J. UCS \textbf{24}(4) 341-356
[5] Peng L, Xin Y, Hua Q, Kai C, Eryun L and Jie T 2012 Expert Syst. Appl. \textbf{39} 6562-74