www.cadanda.com
Computer-Aided Design AND APPLICATIONS

# Design and Application Analysis of Virtual Reality-based Visual Biostatistics System: A Multi-Perspective Approach

Han Zeng[1*]

[1]Simon Fraser University, Vancouver, V5A 1S6, Canada

Corresponding Author: Han Zeng, E-mail: zenghanz@sfu.ca

**Abstract.** In order to improve the effect of visual biostatistics, this paper combines the visual recognition algorithm to construct a multi-perspective visual biostatistics system, and deduces the upper limit formula of the codeword capacity of optical orthogonal codes. Moreover, this paper analyzes several common methods of constructing optical orthogonal codes in combination with the actual needs of biostatistics. In addition, this paper improves the greedy algorithm and accelerated greedy algorithm, compares the two algorithms through Matlab simulation, and simulates and analyzes the correlation characteristics of optical orthogonal codes. Finally, this paper constructs the system architecture of the visual biostatistics system. The simulation study shows that the multi-perspective visualization biostatistics system has a good bioinformatics statistical effect.

**Keywords:** multiple perspectives; visualization; biostatistics; intelligent system; Virtual Reality
**DOI:** https://doi.org/10.14733/cadaps.2023.S14.33-49

## 1 INTRODUCTION

With the rapid growth of the data of biologically relevant systems, researchers have paid more attention to the rapid and effective simulation and prediction of these complex data through statistical methods. Moreover, the intrinsic relationship between the structure of biomolecules and their properties/activities is analyzed with the help of predictive models [6].

Virtual Reality (VR) has emerged as a valuable tool in various fields, including bioinformatics, for addressing the challenges associated with sampling surveys.The most scientific method to solve the problem of bioinformatics is sampling survey. Sampling survey is widely used because of its low cost, high speed and high accuracy. It is a set of behaviors for the purpose of obtaining a sample and relies on a series of programmed algorithms to accurately describe the population. Moreover, sampling survey is a basic theory that is difficult to understand, but is widely used in real society [15]. It involves almost all walks of life, and especially plays a pivotal role in the investigation of biological resources on which we depend. It involves not only the application of general sampling survey methods, but also the application of sequential sampling and marking, and the application of recapture methods is more common. In recent years, with the development of computers, the

complicated and primitive means have been replaced [16], and many departments have also developed special statistical software for sampling surveys. However, most of the sampling survey statistics of biological resources are operated by hand on the computer, which is time-consuming and labor-intensive, and is prone to errors, and the sampling method is not easy to master.

There are two types of errors in sampling surveys: sampling error and non-sampling error [19]. Sampling error is a random error without systematic deviation. The size of the error can be calculated by formula. It decreases with the increase of the sample size, which is inevitable in sampling survey. It is mainly determined by the degree of difference in the marker values of each unit of the population, the choice of sampling method, the number of sample units, and the estimation method. The non-sampling error lacks a satisfactory estimation method, so it is difficult to calculate. The main reasons for non-sampling errors are [7]: 1) Errors in the questionnaire design process in the design stage; the design of questionnaires and sampling survey methods is unreasonable or unscientific, such as the questionnaire is too complex, contains many professional terms or has too many questions , causing the respondents to be bored; the sampling frame is incomplete, such as missing the target overall unit or including non-target overall units, resulting in overestimation or underestimation; 2) The results of the survey are biased due to non-response to the questionnaire during the survey stage, and the survey process Medium: The staff's negligence or low professional quality, such as the staff's inability to communicate well with the respondents, resulting in the respondents' refusal to answer, the investigators subjectively check the sampling units to save time and effort, or make fraudulent: 3) Data In the aggregation stage, non-sampling errors may occur due to possible errors in filling in the adjustment data, inputting the data into the computer, or backward data processing technology. In the process of sampling survey, we should try to reduce non-sampling error as much as possible, and the control of non-sampling error can be divided into three stages [18]: 1) Design stage, the questionnaire design is scientific and rational, the purpose is clear, and the design questions should be To make it easy for the respondents to answer the question, a roundabout survey method should be used for privacy issues. The questionnaire questions should be mostly multiple-choice questions, and subjective questions should be used as little as possible. The choice of the survey method should also be timely and convenient. Staff with high personal and professional qualities should handle the case of refusal to police properly; 3) In the data aggregation stage, the survey data should be strictly reviewed, appropriate data processing technology should be selected, and data entry should be careful and not careless main idea.

In recent years, a large number of research results have been achieved in the analysis of network motifs. The research in the literature [17] shows that some network motifs are basic information processing modules in gene regulation networks, and they cooperate with each other to complete the information in gene regulation networks. Process work. Literature [5] found the same three network motifs with important functions in the gene regulatory networks of bacteria and yeast, which may imply that motifs with the same structure in the same type of network have the same network function. Literature [3] proposed that these three important network motifs also appeared in the signal transduction network and the developing transcription network. In recent years, in the continuous research on biological network motifs, the concept of motif cluster, superfamily, can help people to further understand the function of biological network. The discovery of these network motifs, which are the basic functional modules of the network, is an important way for humans to understand the biological network and even the function of the genome. Especially now, the research on the function of the genome is still in the exploratory stage, and the biological network motif is more reflected. Its important value [12]. Although network motifs have been proven to be of great significance for future biological research, the discovery of network motifs is a difficult and algorithmic task, especially for those with a large number of network nodes [13]. Therefore, it is urgent to improve and research the discovery algorithm of network motifs. The network motif discovery algorithm can generally be divided into three main steps: (1) generate a set of suitable random networks according to the input network [2]; (2) perform subgraph mining on the input network and

the generated random network, ( 3) Compare the frequency of occurrence of subgraphs to identify network motifs. Most of the work focuses on the improvement of the subgraph mining in the second step. Literature [8] proposed a statistical-based biological network motif discovery algorithm, and made certain revisions to the concept of motifs, and achieved very good results. The good effect is not only improved in time complexity, but also more able to reflect the evolution process of the motif. However, the author only carried out a mathematical theoretical analysis of the algorithm, no simulation experiments, and no software published. Reference [11] proposed a biological network motif discovery algorithm based on expectation maximization algorithm. Because the network motif discovery algorithm is a new research direction and involves many fields with great research difficulty, it is a new research topic with great challenges at home and abroad.

Biological big data contains important information such as the origin of life, disease health, and crop cultivation. The efficient and accurate interpretation of biological big data is related to important issues such as human health and food security. Because biological big data has the characteristics of complex type, heterogeneous structure, high redundancy and huge volume, researchers need to use data visualization and other methods to understand its composition characteristics and internal connections, and then dig out more quickly and pertinently. Related knowledge information [1].

With the continuous development of biotechnology, massive data information is being accumulated rapidly, biology has entered the era of big data, and various new experimental technologies represented by second-generation high-throughput sequencing are rapidly generating and accumulating massive data sets. These data sets containing rich knowledge information are called biological big data [9]. In recent years, a number of large-scale biological big data output projects have been generated around the world, such as the Cancer Genome Atlas (TCGA), the 1000 Genomes Project, etc. These data are stored in the European Bioinformatics Institute (EBI), the US National Biological In public databases such as the Center for Technical Information (NCBI) [4].

Data visualization transforms the characteristic information of text or binary data in a database with a large amount of data, such as time or space-related information, into more intuitive and vivid diagrams or tables through computer graphics, statistics and other technologies. The abstract information is intuitively analyzed and represented, which helps to better understand the data, enhance the cognitive data, and mine the laws or intrinsic information between transactions contained in a large amount of data [14]. The role of visual analysis of biological big data mainly includes three aspects: helping researchers to quickly extract essential features from the original data set with large volume and lack of organizational context, providing theoretical guidance for research work; extracting a certain part of biological big data. The characteristics of dimensions can be visually displayed and emphasized in a graphical way; it can effectively deconstruct biological big data, remove redundant information and background noise, and obtain more scientifically meaningful data analysis results [10].

In this paper, an intelligent visualization algorithm is used to construct a multi-perspective visualization biostatistics system from multiple perspectives, and an intelligent method is used to conduct biostatistics to improve the effect of ecological management.

## 2 VISUAL RECOGNITION ALGORITHM

### 2.1 Generation Of Optical Orthogonal Codes

Optical orthogonal codes is a commonly used method to solve the multiple access method in the visual biometric system. It is a set of {0, 1} binary sequences with good autocorrelation and cross-correlation. Since most of the current visual biometric identification systems use light intensity modulation, and the light intensity has no negative value, unlike the traditional code division multiple access (CDMA), there is no negative value in the binary sequence of the optical orthogonal code. Optical orthogonal codes are first used in optical fiber communication code division multiple access,

which enables many asynchronous communicators to complete communications efficiently and stably. Since the code division multiple access system in visual biometric identification does not require synchronization between the two communication parties, the flexibility of the system is greatly increased.

Optical orthogonal codes are closely related to constant weight error correction codes and difference sets. Furthermore, many off-the-shelf techniques have been applied to the analysis of the construction of optical orthogonal codes. However, due to differences in disciplines, we only borrow the theoretical results we care about.

For a set of optical orthogonal codes with code length n and code weight w, we use $(n,\omega,\lambda_a,\lambda_c)$ to represent them. It must satisfy the following two properties:

1) Autocorrelation properties:

$$\sum_{t=0}^{n-1} x_t x_{t+\tau} \le \lambda_a \tag{1}$$

Among them, $x \in C,\tau$ is an integer, $0 < \tau < n$.

2) Cross-correlation properties:

$$\sum_{t=0}^{n-1} x_t y_{t+\tau} \le \lambda_c \tag{2}$$

Among them, $x \ne y \in C,\tau$ is any integer.

The optical orthogonal code here is defined in terms of periodic correlation, that is, the '+' in the formula is modulo n addition. $\lambda_a$ and $\lambda_c$ are the autocorrelation and cross-correlation limits, respectively.

The cyclic shift of the optical orthogonal code does not affect the autocorrelation and cross-correlation characteristics of the optical orthogonal code. We assume that $C'$ is a set of codewords obtained by cyclic shift of a set of optical orthogonal codes C. If C' is still a set of optical orthogonality, then we treat C' and C as the same codeword.

We look at optical orthogonal codes from the perspective of set theory. A set of $(n,\omega,\lambda_a,\lambda_c)$ optical orthogonal codes can be regarded as a set of integers of w tuples modulo n, where a set of w tuples represents a codeword, and the number in each set represents the position of the codeword '1'. Therefore, the auto-correlation and cross-correlation characteristics of optical orthogonal codes can be expressed as:

1) Autocorrelation properties:

$$|(a+X)\cap(b+X)| \le \lambda_a \tag{3}$$

Among them, $x \in C,a,b$ is any integer modulo n and $a \ne b$.

2) Cross-correlation properties:

$$|(a+X)\cap(b+Y)| \le \lambda_c \tag{4}$$

Among them, $x \neq y \in C$ , and $a,b$ is arbitrary integers. We need to pay attention to $a+X=\{a+x:x\in X\}$, and the integer $a,b$ must be the result of the operation modulo n. It can be seen that when the code length n is much larger than the code weight w, it is very simple and clear to express the optical orthogonal code from the perspective of set theory.

When the autocorrelation limit and cross-correlation limit of a set of optical orthogonal codes are equal, we can express $(n,\omega,\lambda)$ in short form, where $\lambda$ represents the auto-correlation limit and the cross-correlation limit. For example, $C=\{1101000\}$ is a codeword of a set of optical orthogonal codes (7, 3, 1). In the notation of sets, it is C={0, 1, 3}(mod7). For example, c={1100100000000, 1010000100000} are two code words in a set of optical orthogonal codes (13, 3, 1). It is expressed in the notation of a set as C={{0, 1, 4}, {0, 2, 7}} (modl3).

The m optical orthogonal codes with parameter $(n,\omega,\lambda_a,\lambda_c)$ can be expressed in the form of codeword blocks:

$$C_1=\{a_{i1},a_{t2},\ldots,a_{i\omega}\} \tag{5}$$

Among them, i represents different codewords, and $a_m$ represents the position where the nth '1' appears in the i-th codeword. In the i-th codeword of the optical orthogonal code, the sum of the difference sets of all '1' codes can be expressed as

$$D_i=\{a_{ij}-a_{ik},j,k=1,\ldots,\omega;j\neq k\}(mod\ L) \tag{6}$$

Then, the set of all differences of the orthogonal code can be expressed as:

$$D=\{D_i,i=1,\ldots,m\} \tag{7}$$

According to the correlation properties of the optical orthogonal codes, it is not difficult to obtain the following properties of the difference set sum representation of the optical orthogonal codes:

Property 1: The correlation characteristics of optical orthogonal codes can be explained as follows: 1) The autocorrelation characteristics can be understood as the repetition of elements in any difference set and $D_i$ does not exceed $\lambda_a$. 2) The number of elements in the intersection $D_i\cap D_j$ of any two difference sets cannot exceed $\lambda_c$.

Property 2: The upper limit of the number of codewords of the optical orthogonal code $(n,\omega,\lambda_a,\lambda_c)$ can be expressed as $\Phi(n,\omega,\lambda_a,\lambda_c)$. Next, we discuss the upper limit of the codeword of the optical orthogonal code $(n,\omega,1)$. Since the autocorrelation limit of the optical orthogonal code is 1, there is no repeated element in any difference set $D_i$ of the optical orthogonal code codeword. Therefore, the number of elements of the set is $|D_i|=\omega(\omega-1)$. Since the cross-correlation limit of the optical orthogonal code is 1, the intersection of any two difference sets of the orthogonal code word is $D_i\cap D_j=\varnothing$. Therefore, it is not difficult to obtain the upper limit of the capacity of the orthogonal code as:

$$\Phi(n,\omega,1)\le\frac{n-1}{\omega(\omega-1)} \tag{8}$$

For example, $x=\{0,1,3,7\}$ and $y=\{0,5,13,22\}$ are two codewords in the codeword (49, 4, 1), and their difference sets can be calculated as $D_x=\{1,2,3,4,6,7,42,43,45,46,47,48\}$ and $D_y=\{5,8,9,13,17,22,27,32,36,40,41,44\}$, respectively. It can be seen that there are no repeated elements in $D_x$ and $D_y$, so the symbolic property is 1, that is, it conforms to the autocorrelation property. $D_x\cap D_y=\varnothing$ conforms to the cross-correlation property.

Optical orthogonal codes have other applications. In spread spectrum communication, the lower the correlation between frequency hopping patterns, the better, so optical orthogonal codes can be used to generate good frequency hopping patterns. However, there is an important factor to consider, the frequency hopping pattern can usually be represented as dots on a rectangular checkerboard, indicating that only one frequency can occur per time slot. To get the frequency hopping pattern from the optical orthogonal code, we can put an optical orthogonal code into a matrix. The '1' in each column indicates that the time slot has the corresponding frequency for communication.

Optical orthogonal codes can also be applied to many other situations, as long as they are related to autocorrelation and cross-correlation, such as signal design for radar and sonar.

### 2.2 Design of Optical Orthogonal Codes

The codeword capacity $\Phi(n,\omega,\lambda)$ of optical orthogonal codes can be drawn from the relevant conclusions of algebraic coding theory. A set of error correction codes is a set of binary n-tuples, and each set of n-tuples of binary numbers is a codeword, and the number of '1' is the code weight. Hamming distance is the number of '1' in different positions between two codewords. The most basic problem in algebraic coding theory is to find the maximum number of codewords for a given code length of n and a code distance of at least d. $A(n,d,\omega)$ represents the maximum capacity of an error correction code word with a code length of n, a code weight of w, and a code distance of at least d. It is difficult to know the exact value of the codeword capacity $A(n,d,\omega)$ of the error correction code. After many people's research, it is concluded that the Johnson circle is as follows:

$$A(n,2\delta,\omega)\le\frac{n(n-1)\cdots(n-w+\delta)}{w(w-1)\cdots\delta} \tag{9}$$

We can derive the upper bound $\Phi(n,\omega,\lambda)$ for optical orthogonal codes according to the Johnson bound formula for error correction codes.

Theorem 1:

$$\begin{aligned}\Phi(n,w,\lambda)&\le(1/n)A(n,2w-2\lambda,w)\\&\le((n-1)\cdots(n-\lambda)/w(w-1)\cdots(w-\lambda))\end{aligned} \tag{10}$$

Proof: For any set of optical orthogonal codes $C(n,w,\lambda)$, all cyclically shifted codewords of the codeword are denoted by $C'$. It is not difficult to obtain the newly obtained codeword capacity $|C'|=n|C|$, and the code weight of the newly obtained codeword is also $w$. Since the autocorrelation limit and the cross-correlation limit of the optical orthogonal code C are both $\lambda$, that is, in the codeword $C'$, the number of any two codewords'1' in the same position is at most $\lambda$. Therefore, the code distance of the codeword $C'$ is at least $2w-2\lambda$. Therefore, we can get:

$$|C'|=A(n,2w-2\lambda,w) \tag{11}$$

Because of $|C\eta=n|C|$, theorem 1 is proved.

When $\lambda_a \neq \lambda_c$, we set $\lambda=max(\lambda_a,\lambda_c)$, and the above theorem still holds. When $\lambda$ is smaller, the result obtained by Theorem 1 is closer to the exact value of the codeword capacity.

We can construct a new set of optical orthogonal codes from a set of known optical orthogonal codes $C(n,w,\lambda_a,\lambda_c)$.

Method 1: For a set of known optical orthogonal codes $C(n,w,\lambda_a,\lambda_c)$, we can regard it as another set of optical orthogonal codes $C(n,w,\lambda_a',\lambda_c')$, where $\lambda_a' \geq \lambda_a, \lambda_c' \geq \lambda_c$.

Method 2: For m known optical orthogonal codes $C(n,w,\lambda_a,\lambda_c)$, we can construct $\binom{m}{2}$ parameters as $(n,2w-2\lambda_c,2\lambda_a+2\lambda_c,w+3\lambda_c)$ codeword $C'$. For any two codewords x and y in codeword C, we can construct a codeword z in codeword $C'$. First, we set $z'=x\vee y$, where $\vee$ represents a bit operation. Since the code weights of x and y are both w, and the overlap of '1' does not exceed $\lambda_c$, the code weight $wt(z')$ of the codeword z' is at least $2w-2\lambda_c$. Then, a 1 of $wt(z')-(2w-2\lambda_c)$ at any position in the codeword $z'$ becomes a 0 to obtain the codeword z. In this way, the code weight of each codeword z in the codeword set $C'$ is equal to $2w-2\lambda_c$. The autocorrelation value of codeword z does not exceed the autocorrelation value of codeword $z'$

$$\begin{aligned}
&|(x\vee y)\cap(x\vee y \text{ Cyclic displacement })| \\
&\leq |x\cap(x \text{ Cyclic displacement })| \\
&\quad +|x\cap(y \text{ Cyclic displacement })| \\
&\quad +|y\cap(x \text{ Cyclic displacement })| \\
&\quad +|y\cap(y \text{ Cyclic displacement })| \\
&\leq \lambda_a+\lambda_c+\lambda_c+\lambda_a=2\lambda_a+2\lambda_c
\end{aligned} \tag{12}$$

Therefore, the optical orthogonal code $C'$ satisfies the autocorrelation property. Among them, codewords $z_1$ and $z_2$ are represented by $z_1' = x_1 \vee y_1$ and $z_2' = x_2 \vee y_2$, respectively, where $x_1, x_2, y_1, y_2$ is the codeword in the optical orthogonal code C, and $\{x_1, y_1\} \neq \{x_2, y_2\}$. The cross-correlation value of codewords $z_1$ and $z_2$ does not exceed the cross-correlation value of $z_1''$ and $z_2'$:

$$\left|(x_1 \vee y_1) \cap (x_2 \vee y_2 \ \textit{Cyclic displacement}\ )\right| \leq \begin{cases} 4\lambda_c, \ \textit{If}\ \{x_1, y_1\} \cap \{x_2, y_2\} = \varnothing \\ w + 3\lambda_c, \ \textit{other} \end{cases} \tag{13}$$

Because of $\lambda_c \leq w$, the codeword $C'$ satisfies the cross-correlation property.

Method 3: For a given optical orthogonal code $C(n, w, \lambda_a, \lambda_c)$, we can construct a new set of optical orthogonal codes $C'(tn, tw, t\lambda_a, t\lambda_c)$. This method is relatively simple, and repeats the known optical orthogonal code t times to obtain a new codeword.

Greedy algorithms are used in many fields. Here, we use a greedy algorithm to construct optical orthogonal codes with regular parameters. We give lower bounds on optical orthogonal codes according to the greedy algorithm. In the case of different parameters, the accuracy of the two lower bounds is not the same. In practice, the codewords that may be constructed by the greedy algorithm are better than the lower bound budget. In addition, the greedy algorithm still has a lot of room for improvement.

The greedy algorithm constructs an optical orthogonal code with parameter $(n, w, \lambda_a, \lambda_c)$. First, the codeword set is an empty set. Then, we detect $\binom{n}{w}$ possible n-tuple codewords one by one. If the autocorrelation and cross-correlation characteristics are satisfied, the n-tuple is added to the optical orthogonal code codeword set.

The time complexity of the greedy algorithm is $\binom{n}{w}|C|w^2$ and the space complexity is $|C|n$. Next, we introduce the lower bound of the optical orthogonal code derived from the greedy algorithm.

Theorem 2:

$$\Phi(n, w, \lambda_a, \lambda_c) \geq \frac{\binom{n}{w} - \frac{n-1}{2}\binom{w}{\lambda_a + 1}\binom{n}{w - \lambda_a - 1}}{n \cdot \sum_{i=\lambda_c+1}^{\min\{n-w, w\}} \binom{n-w}{w-i}\binom{w}{i}} \tag{14}$$

Proof: For the molecular part, we only need to prove that there are at most $\frac{n-1}{2}\binom{w}{\lambda_a+1}\binom{n}{w-\lambda_a-1}$ autocorrelation properties that do not conform to optical orthogonal codes. For the denominator part, we only need to prove that for any n-tuple codeword, there are at most $n\cdot\sum_{i=\lambda_c+1}^{\min\{n-w,w\}}\binom{n-w}{w-i}\binom{w}{i}$ n-tuple codewords that do not conform to the cross-correlation properties of optical orthogonal codes.

First, we prove the molecular part: we set y to be an n-tuple codeword and $s_1,s_2,\cdots,s_w$ to denote its '1' position. If the n-tuple y does not conform to the autocorrelation property of the optical orthogonal code, then at least $\lambda_a+1$ elements in the difference set D of the n-tuple codeword are the same. The value of the same element in the difference set D is $\delta$, then there are at most $(n-1)/2$ choices for the value of $\delta$. Moreover, there are at most $\binom{w}{\lambda_a+1}$ ways to determine the position of the '1' in the n-tuple y in the difference set D, and at most $\binom{n}{w-\lambda_a-1}$ ways to select the positions of the remaining '1's. The proof of the molecular part is completed. The proof of the denominator part is as follows: For a given n-tuple codeword x, there are at most $\sum_{i=\lambda_c+1}^{\min\{n-w,w\}}\binom{n-w}{w-i}\binom{w}{i}$ n-tuple codeword '1' positions with more than $\lambda_c$ overlaps. Moreover, each n-tuple codeword has at most n corresponding cyclically shifted codewords, and these codewords also do not represent the cross-correlation properties of optical orthogonal codes.

$$\Phi\left(n,w,\lambda_a,\lambda_c\right)\geq\frac{n^{\lambda_c}}{\frac{w!}{\left(w-\lambda_c-1\right)!}\binom{w}{\lambda_c+1}}+Low\ order\ term \tag{15}$$

Among them, the lower-order terms of the formula are negligible.
Its flow chart is as follows Figure 1.

### 2.3 Simulation Results and Analysis

If it is assumed that the code length is 43, the code weight is 4, and the autocorrelation and cross-correlation limits are 1, the result of the codeword block group simulated by this algorithm is:
{0,1,3,12}{0,4,10,17}{0,5,19,27}
When the code length is 49, the code weight is 4, and the autocorrelation limit and cross-correlation limit ae 1, the result of the codeword block group simulated by this algorithm is:
{0,1,17,26}{0,2,5,12}{0,4,15,35}{0,6,19,27)

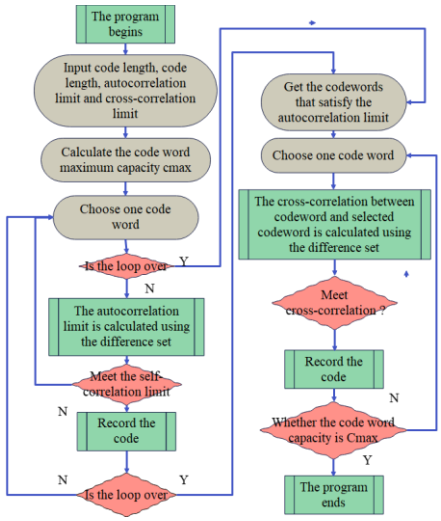


**Figure 1:** Flow chart of constructing optical orthogonal codes by greedy algorithm.

Next, we perform autocorrelation limit verification for {0, 1, 3, 12), as shown in Figure 2, and cross-correlation limit verification for {0, 1, 3, 12) and {0, 4, 10, 17}, as shown in Figure 3. It can be seen from the verification of Figure 2 and Figure 3 that the design of the algorithm is accurate. Compared with the fast greedy algorithm to be introduced in the F-face, the main advantage of this algorithm is that it can find the codeword with the largest capacity, but its time is several times or even dozens of times that of the fast greedy algorithm.

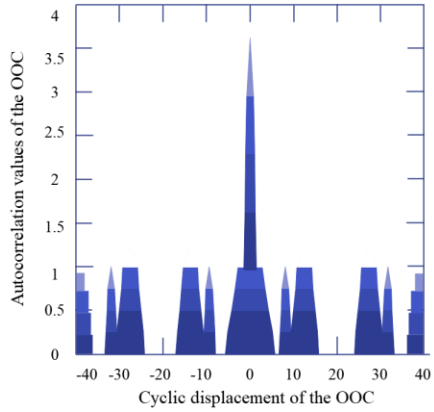


**Figure 2:** Autocorrelation function of codeword {0, 1, 3, 12}.

In practical applications, people often use the "fast and greedy algorithm". Compared with the greedy algorithm, the fast greedy algorithm is much faster, but the codeword obtained by the accelerated greedy algorithm is relatively less.

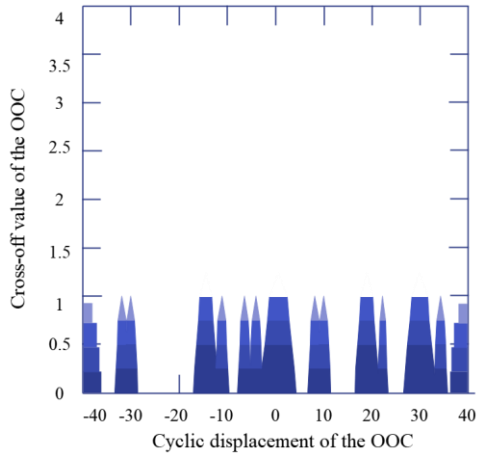


**Figure 3:** Cross-correlation function of codewords {0, 1, 3, 12} and {0, 4, 10, 17}.

Fast and greedy algorithm: This algorithm is easier to understand by discussing it in the way of set theory. When the algorithm constructs the optical orthogonal code, the elements of the code word are added one by one. It contains two loops, the outer loop is to add a codeword every time the loop is looped, and the inner loop is to add an element to the existing codeword every time it loops. At the beginning of the loop, the codeword set is an empty set, and when the inner loop cannot find more suitable codeword elements, the loop ends.

It is assumed that there are m+1 W element sets already in the codeword set, and the m-th w tuple set also has w - 1 elements that have been found. For each x (O<x<n), the algorithm counts the number of times the equation x=a+b-C holds. Among them, a, b, c, a≠c, are the elements in the m-th incomplete codeword. If the number of times to find the expression is less than $\lambda_a$, then x conforms to the autocorrelation property of the optical orthogonal code. For the found codeword $S_i$ (w-tuple set), $1 \le i \le m-1$. The algorithm also counts the number of times the equation x=a+b-c holds, where a is the element in the incomplete w-tuple to be added, and $b \ne c$ is the element in $S_i$. If for each selected codeword $S_i$, the number of times the equation holds is less than $\lambda_c$, then x conforms to the cross-correlation property of the optical orthogonal code. There are $(w-1)^2(w-2)$ kinds of expressions a+b-c for the codeword to be added, where $a,b,c,a \ne c$ is the element in the codeword to be added. For the selected codeword, there are f$(m-1)w(w-1)^2$ expressions a+b-c, $a$ is an element in the incomplete $w$-tuple to be added, and $b \ne c$ is an element in a codeword in the selected codeword. Therefore, if an element can make the codeword to be added a complete optical orthogonal code, it must satisfy:

$$\frac{(m-1)w(w-1)^2}{\lambda_c} + \frac{(w-1)^2(w-2)}{\lambda_a} + w - 1 < n \tag{16}$$

From the above equation, we can get another lower bound for optical orthogonal codes.
Theorem 3:

$$\Phi\left(n,w,\lambda_a,\lambda_c\right) \geq \frac{\lambda_c(n-w+1)-\left(\lambda_c/\lambda_a\right)(w-1)^2(w-2)}{w(w-1)^2} \tag{17}$$

The time complexity of the fast greedy algorithm is $O\left(|C|^2\ w^4\right)$ and the space complexity is O(n). Compared with the greedy algorithm, the speed of the accelerated greedy algorithm is greatly improved.

The simulation flow chart is as follows figure 4:

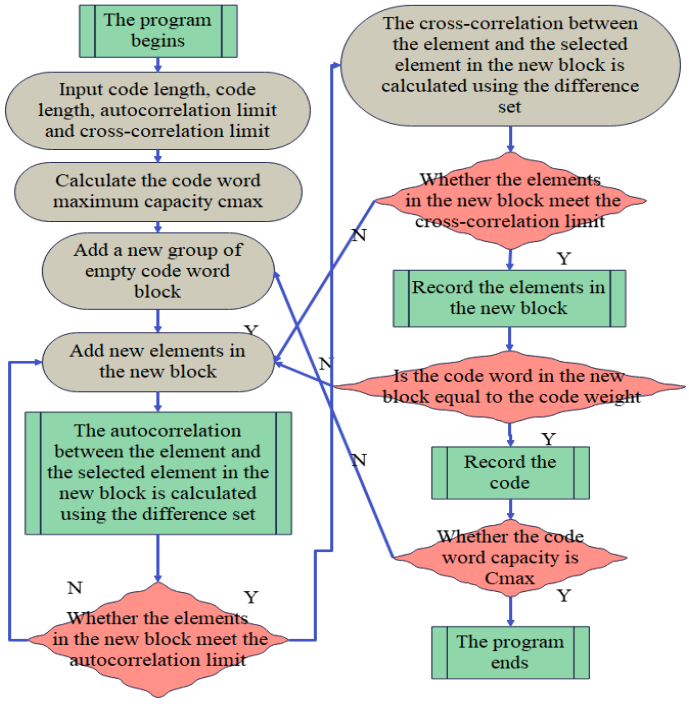


**Figure 4:** Flow chart of constructing optical orthogonal codes by fast greedy algorithm.

## 2.4 Simulation Results and Analysis

When the code length is 35, the code weight is 3, and the autocorrelation limit and cross-correlation limit are 1, the result of the codeword block group simulated by this algorithm is:
{0,1,3}{0,4,9){0,6,13}{0,8,18)
When the code length is 49, the code weight is 4, and the autocorrelation limit and cross-correlation limit are 1, the result of the codeword block group simulated by this algorithm is:
{0,1,3,7){0,5,13,22}{0,10,21,33}
Next, we perform autocorrelation limit verification for {0, 1, 3, 7), as shown in Figure 5, and cross-correlation limit verification on {0, 1, 3, 7) and {0, 5, 13, 22}, as shown in Figure 6. It can be seen that a codeword with a code length of 49, a code weight of 4, and an autocorrelation limit and a

cross-correlation limit of 1 is also generated. The codeword capacity obtained by the greedy algorithm is larger than that obtained by the fast greedy algorithm, which is the main disadvantage of the fast greedy algorithm.

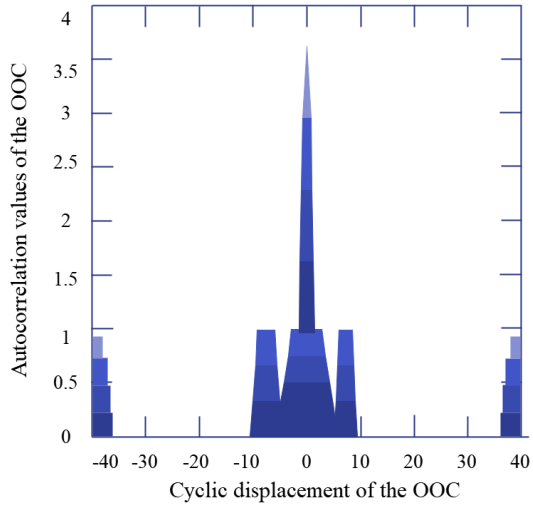


**Figure 5:** Autocorrelation function of codeword {0, 1, 3, 7).

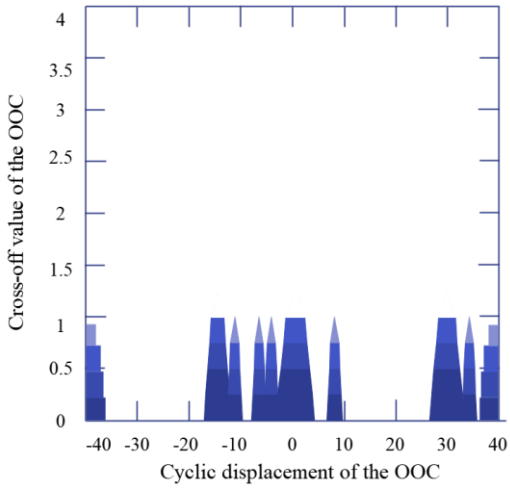


**Figure 6:** Autocorrelation function of codewords {0, 1, 3, 7) and {0, 5, 13, 22).

## 3 MULTI-PERSPECTIVE VISUAL BIOSTATISTICS SYSTEM

Visual biostatistics software consists of three modules, namely data management module, statistical analysis module and chart module. The system function module diagram is shown in Figure 7.

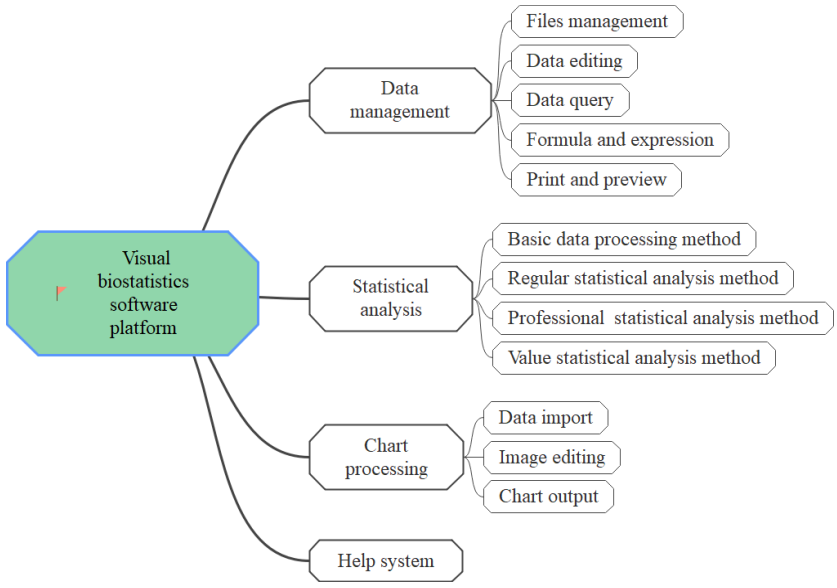


**Figure 7:** System function block diagram.

This section uses the system architecture diagram to describe the functions of each module and the relationship between the modules in the multi-source biological pathway data visualization system. As shown in Figure 8, the framework is divided into three layers, namely presentation layer, application service layer and data layer. The presentation layer exchanges data with the application service layer through the visualization processor of the application service layer, and the data layer exchanges data with the application service layer through the remote access processor of the application service layer.

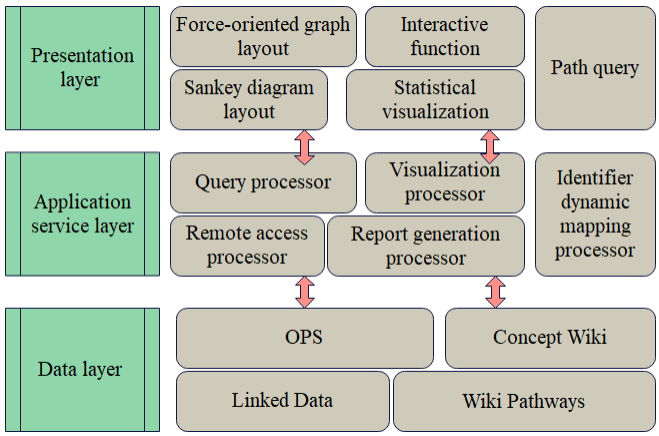


**Figure 8:** Data visualization system framework for multi-source biological pathways.

The effect of the above model is verified, the application effect of the multi-perspective visual biostatistics system is calculated, and the application effect of the system in this paper is calculated

and evaluated through multiple groups of simulations. Finally, the test results shown in Figure 9 are obtained.

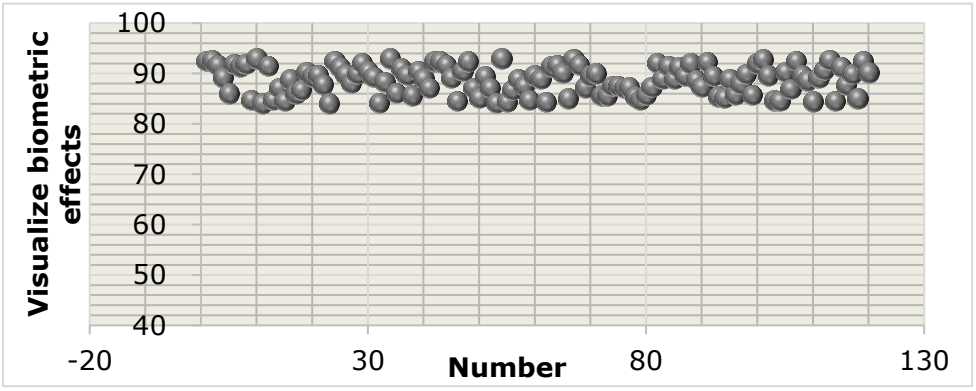


**Figure 9:** Application effect of multi-perspective visual biostatistics system.

The simulation study shows that the multi-perspective visual biostatistics system has a good bioinformatics statistical effect.

## 4 CONCLUSION

Biological resources are the basis for human survival, and a clear understanding of its structure, distribution, quantity, species and other characteristics is essential for better protection and utilization of resources. From small microorganisms to human beings, biological resources constitute an organic world. Moreover, population surveys, population numbers, distribution, growth laws, etc., the protection, development and utilization of forest resources, the occurrence trend and control of agricultural pests and diseases must be clear, so as to change from passive to active and achieve targeted. However, if a comprehensive investigation is carried out, not only a lot of human, material and financial resources will be invested, but some of them will not work at all. This paper uses the intelligent visualization algorithm to construct a multi-perspective visual biostatistics system from multiple perspectives, and conducts biostatistics through intelligent methods. The simulation study shows that the multi-perspective visual biostatistics system has a good bioinformatics statistical effect.

*Han Zeng,* https://orcid.org/0000-0003-4448-0440

**Funding Statement**

This work was supported by Simon Fraser University.

## REFERENCES

[1] Ben Itzhak, N.; Vancleef, K.; Franki, I.; Laenen, A.; Wagemans, J.; Ortibus, E.: Visuoperceptual profiles of children using the Flemish cerebral visual impairment questionnaire, Developmental Medicine & Child Neurology, 62(8), 2020, 969-976. https://doi.org/10.1111/dmcn.14448

[2] Chen, C. C.; Huang, C.; Gribbins, M.; Swan, K.: Gamify Online Courses with Tools Built into Your Learning Management System (LMS) to Enhance Self-Determined and Active Learning, Online Learning, 22(3), 2018, 41-54. https://doi.org/10.24059/olj.v22i3.1466

[3] de Oliveira, P. V. A. G.; de Moura, D. T. H.; Ribeiro, I. B.; Bazarbashi, A. N.; Franzini, T. A. P.; Dos Santos, M. E. L.; de Moura, E. G. H.: Efficacy of digital single-operator cholangioscopy in the visual interpretation of indeterminate biliary strictures: a systematic review and meta-analysis, Surgical Endoscopy, 34(8), 2020, 3321-3329. https://doi.org/10.1007/s00464-020-07583-8

[4] Demarest, S. T.; Olson, H. E.; Moss, A.; Pestana-Knight, E.; Zhang, X.; Parikh, S.; Benke, T. A.: CDKL5 deficiency disorder: Relationship between genotype, epilepsy, cortical visual impairment, and development, Epilepsia, 60(8), 2019, 1733-1742. https://doi.org/10.1111/epi.16285

[5] Dingen, D.; van't Veer, M.; Houthuizen, P.; Mestrom, E. H.; Korsten, E. H.; Bouwman, A. R.; Van Wijk, J.: RegressionExplorer: Interactive exploration of logistic regression models with subgroup analysis, IEEE transactions on visualization and computer graphics, 25(1), 2018 , 246-255. https://doi.org/10.1109/TVCG.2018.2865043

[6] Huiliang Zhao, Jian Lyu, Xiang Liu, and Zhenghong Liu. "Customization-oriented product flexible manufacturing experience system design based on VR." In IOP Conference Series: Materials Science and Engineering, 561(1), 012098, IOP Publishing, 2019. https://doi.org/10.1088/1757-899X/561/1/012098

[7] Hamedani, A. G.; Abraham, D. S.; Maguire, M. G.; Willis, A. W.: Visual impairment is more common in Parkinson's disease and is a risk factor for poor health outcomes, Movement Disorders, 35(9), 2020 , 1542-1549. https://doi.org/10.1002/mds.28182

[8] Han, G.; Han, J.; Han, K.; Youn, J.; Chung, T. Y.; Lim, D. H.: Visual acuity and development of Parkinson's disease: a nationwide cohort study, Movement Disorders, 35(9), 2020, 1532-1541. https://doi.org/10.1002/mds.28184

[9] Jadidi, K.; Naderi, M.; Mosavi, S. A.; Nejat, F.; Aghamolaei, H.; Serahati, S.: Pre-operative factors influencing post-operative outcomes from MyoRing implantation in keratoconus, Clinical and Experimental Optometry, 102(4), 2019 , 394-398. https://doi.org/10.1111/cxo.12859

[10] Liu, L.; Wang, L.; Yang, W.; Ni, W.; Jin, L.; Liu, J.; Ren, A.: Gestational hypertension and pre-eclampsia and risk of spontaneous premature rupture of membranes: A population-based cohort study, International Journal of Gynecology & Obstetrics, 147(2), 2019, 195-201. https://doi.org/10.1002/ijgo.12943

[11] Marin, I.; Wagner, B. D.; Manoharan, N.; Palestine, A.; Poppelaars, F.; Patnaik, J. L.; Lynch, A. M.: AREDS Supplementation and Complement System Levels in Patients with Intermediate Age-Related Macular Degeneration, Investigative Ophthalmology & Visual Science, 63(7), 2022 , 2531-A0100.

[12] Mark, P. B.; Papworth, R.; Ramparsad, N.; Tomlinson, L. A.; Sawhney, S.; Black, C.; McCowan, C.: Risk factors associated with biochemically detected and hospitalised acute kidney injury in patients prescribed renin angiotensin system inhibitors, British journal of clinical pharmacology, 86(1), 2020, 121-131. https://doi.org/10.1111/bcp.14141

[13] Messer, L. H.; Berget, C.; Vigers, T.; Pyle, L.; Geno, C.; Wadwa, R. P.; Forlenza, G. P. Real world hybrid closed-loop discontinuation: predictors and perceptions of youth discontinuing the 670G system in the first 6 months, Pediatric Diabetes, 21(2), 2020,319-327. https://doi.org/10.1111/pedi.12971

[14] Pucchio, A.; Eisenhauer, E. A.; Moraes, F. Y.: Medical students need artificial intelligence and machine learning training, Nature Biotechnology, 39(3), 2021, 388-389. https://doi.org/10.1038/s41587-021-00846-2

[15] Sotirchos, E. S.; Gonzalez Caldito, N.; Filippatou, A.; Fitzgerald, K. C.; Murphy, O. C.; Lambe, J.; International Multiple Sclerosis Visual System (IMSVISUAL) Consortium.: Progressive

multiple sclerosis is associated with faster and specific retinal layer atrophy, Annals of neurology, 87(6), 2020,885-896. https://doi.org/10.1002/ana.25738

[16] Subbaswamy, A.; Saria, S.: From development to deployment: dataset shift, causality, and shift-stable models in health AI, Biostatistics, 21(2), 2020, 345-352. https://doi.org/10.1093/biostatistics/kxz041

[17] Venkatesan, S.; Saji, S.: Graphic medicine and the limits of biostatistics, AMA Journal of Ethics, 20(9), 2018 , 897-901. https://doi.org/10.1001/amajethics.2018.897

[18] Wentzel, A.; Hanula, P.; Luciani, T.; Elgohari, B.; Elhalawani, H.; Canahuate, G.; Marai, G. E.: Cohort-based T-SSIM visual computing for radiation therapy prediction and exploration, IEEE transactions on visualization and computer graphics, 26(1), 2019, 949-959. https://doi.org/10.1109/TVCG.2019.2934546

[19] Ying, G. S.; Maguire, M. G.; Glynn, R. J.; Rosner, B.: Tutorial on biostatistics: longitudinal analysis of correlated continuous eye data, Ophthalmic epidemiology, 28(1), 2021, 3-20. https://doi.org/10.1080/09286586.2020.1786590