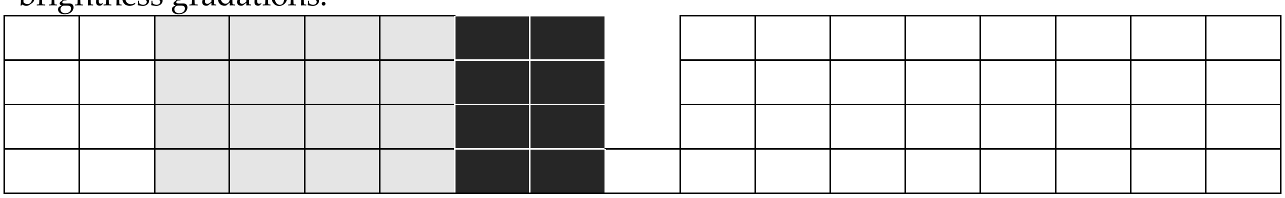
ALGORITHMS ENCODING AND COMPRESSION OF MEDIA SOURSES

Beknazarov Kamoliddin Tursunpulatovich
Independent researcher, Uzbekistan University of journalism and Mass Communications
Beknazarova Saida Safibullayevna
Professor, Uzbekistan University of journalism and Mass Communications

The visual information to be processed on a computer can be entered into a memory device from a special sensor (input device) or represent the result of some calculations. In both cases, this information is compared with a certain picture that is visually perceived by a person. This mapping is essentially the fundamental difference between visual information and numeric, logical, symbolic, or any other information that can be represented in computer memory [4].

The most simple and natural discretization is achieved using a coordinate grid formed by lines parallel to the x and y axes of the Cartesian coordinate system. In each node of such a grid, the brightness or transparency of the carrier of visually perceived information is counted, which is then quantized and presented in the computer memory. Figure 1 illustrates the encoding process using an example of an image with four brightness gradations.



a) b)

Fig. 1. Encoding of visual information: a-the original image; b - its receptor representation.

Obviously, that visual information must reflect the state of brightness or transparency of each point of the visually perceived picture with a certain degree of accuracy. For simplicity, we will consider only flat pictures in the future. To represent visual information in digital form, you need to discretize the space (plane) and quantize the brightness value at each discrete point.

The result is shown in Fig. 1, b in the form of a matrix, the elements of which are aij counts in the lattice nodes.

This representation of visual information is the receptor [3], natural [1], element-based, or matrix [2]. It deserves attention, if only because it most conveniently describes the processes of image input and output and makes it easy to establish an unambiguous correspondence between the picture and its representation in computer memory.

The wide class of sensors used for entering optical information into a computer is a set of light-sensitive elements (receptors) that convert a light signal into an electric one. During the input process, the receptors located in the nodes of the receptor grid are interrogated in a certain sequence and the signals taken from them are converted to digital form. This results in a sequence of counts at the nodes of the coordinate grid. For example, when using TV sensors, the information contained in a TV frame is read line by line, i.e. the time counts form the following sequence:

$$a_{11}, a_{12}, \dots, a_{1x}, a_{21}, a_{22}, \dots, a_{2n}, a_{m1}, a_{m2}, \dots, a_{mn}$$

Where aij is an element of the sample matrix located at the intersection of the i-th row and the j-th column; m is the number of matrix rows corresponding to the number of rows in the frame; n is the number of matrix columns corresponding to the number of elements in the row.

Where a_{ij}- is an element of the sample matrix located *i*- th row and *j*-th column; *m*- is the number of matrix rows corresponding to the number of rows in the frame; *n*- is the number of matrix columns corresponding to the number of elements in the row.

When you using other sensors (for example, solid-state sensors), the readings may be issued in a different sequence. In all cases, the piecemeal method of image representation we can use as an intermediate method for reading visual information.

This method is also convenient for displaying information on video monitoring devices. In this regard, the problem of encoding visual information we can consider as a problem of converting the natural form of representation into a form that is convenient for storing and processing in a computer.

Given that any processing of information can be considered as its recoding, we will distinguish primary, or absolute, code descriptions from secondary, or relative [1]. Without giving strict definitions, we will point out only two significant differences in such code descriptions. First, primary methods of representation, as opposed to secondary ones, must be universal, i.e. allow the possibility of reconstructing the corresponding image with accuracy determined by the parameters of the receptor lattice and the number K quantization levels, and do not take into account the specifics of the input images or further processing algorithms. Secondly, the primary representation can be obtained quite simply from the natural one so that the recoding is performed by relatively simple devices in the process of entering information into the computer. The example of a primary description of visual information is its element-by-element encoding. Secondary ways of presenting information can differ significantly from the primary (or natural) ones and are obtained as a result of fairly complex algorithms for processing visual information.

For example, you can offer very compact ways to represent basic electrical circuits, based on a list of elements that occur in the scheme, indicating the locations of the drawing of the elements themselves and their connecting lines. As an example of a secondary description, you can give a description of an image in a certain feature space that is later used for image recognition.

In order to further formalize the encoding task, it is convenient to more clearly identify the structural units of visual information. Let's call all the visual information to be recorded in the computer a picture or a scene. We will assume that it is presented in a natural form. There are no restrictions on the size of the picture. In a particular case, its size may exceed the size of the receptor field of the optical sensor used for entering information into the computer. In this case, the picture is divided into rectangular areas called rasters or frames (Fig. 2).

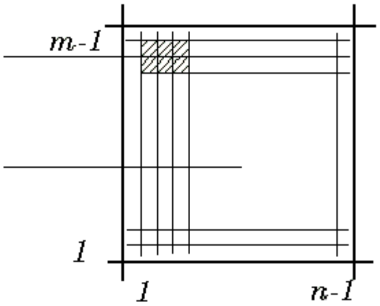


Fig. 2. The structural units for the machine representation of visual information

Each raster size corresponds to the sensor's receptor field and contains *mxn* samples forming a rectangular matrix consisting of m rows and n columns. Each count is characterized by a number of K levels of quantization of the video signal. We usually assume $K=2k$, where k –is an anger. Sometimes there is another structural unit of information-the field. A field is a part of a frame formed by several neighboring counts.

Sometimes there is another structural unit of information-the field. A field is a part of a frame formed by several neighboring counts. Parameters m, n and k are selected depending on the required accuracy of information representation, taking into account the capabilities of the optical sensor. For example, if such a sensor uses a TV camera operating in standard mode, the number of rows and columns of the raster usually varies between 256 and 512, and the number of distinguishable gray levels does not exceed 256 ($k \leq 8$).

In general, recording a frame in increments requires memory containing $W=mnk$ бит информации. If we evaluate W, we can make sure that a significant amount of memory is required for the piecemeal representation of the frame. For example, let $m=n=256$ and $k=6$. In this case $W=6 \cdot 216$. When determining how to place information in computer memory, it is necessary to take into account the fact that computer memory has a dictionary structure, i.e. there is a minimum addressable word of a fixed length s. In general s is not equal to or a multiple k. here we can offer two approaches to solving the problem of placing counts in computer memory words.

In the first of them, all bits of all words in the selected array are used to represent visual information.

In the first of them, all bits of all words in the selected array are used to represent visual information. In this case, the view is more compact and requires less memory. However, this placement is inconvenient for processing information, since some counts are placed in several words. In another approach, each count is represented as a whole by some machine word. However, in General, not all bits of machine words are used, which leads to an increase in the required amount of memory. For example, if you take an approach where each count is placed in a separate word, you will need *mn* to represent a single frame.

Of course, to formulate the coding problem as a problem of obtaining such a representation of visual information that takes up the minimum amount of computer memory and at the same time is convenient for processing. The practical significance of minimizing the amount of memory becomes more obvious when you consider that in many cases it is necessary to analyze large images containing hundreds or even thousands of rasters. To get some numerical estimates, consider a real example in which the size of the processed picture is 200x300 mm and each count is represented by a word containing eight binary digits. If the step of the receptor grid is equal to 0.1 mm, then memory of the order of 6 MB is required.

Let's estimate the quality or efficiency of the code by the average number of bits per count: $l = \frac{W_{mn}}{mn}$,where *Wmn* - is the amount of memory required to represent a single frame size *mxn*.

If you do not take into account possible losses due to the dictionary organization of memory, for the element-by-element encoding method in accordance with (3) $l=k$. Assuming that the amount of memory required for piecemeal encoding can be reduced, let's assume $l_{max}=k$. Let's introduce the reduction coefficient into consideration $a = \frac{l_{max}}{l}$ characterizes a reduction in memory size compared to piecemeal encoding.