Media resources in video information systems

Saida Beknazarova[1*], *Zamira* Ishanxadjayeva[2], and *Mexriban* Jaumitbaeva[3]

[1]Tashkent University of Information Technologies named after Muhammad Al- Khwarizmi, Tashkent, Uzbekistan
[2]National University of Uzbekistan, Tashkent, Uzbekistan
[3]Specialized Branch Tashkent State University of Law, Tashkent, Uzbekistan

Abstract. The article describes methods for increasing the contrast of images in video information systems, the main compression of the video stream is provided by eliminating inter-frame redundancy using methods for compensating the movement of image fragments of adjacent frames. However, the use of motion compensation methods requires the formation of additional data (metadata) containing information about the types of image blocks used, their movement coordinates, etc. At the same time, in order to increase the compression of the video stream without sacrificing its quality, a higher accuracy of motion compensation is required, which leads to an increase in the number of blocks and, accordingly, an increase in the volume of the video stream metadata that reduces the effectiveness of motion compensation. This is the main problem of compressing streaming video without compromising image quality. In addition, higher positioning accuracy of blocks with motion compensation dramatically reduces the speed of image processing, which is not always feasible in real-time systems.

1 Introduction

Digital image processing is a rapidly developing field of science. Research and development of methods and algorithms for processing and analyzing information presented in the form of digital images is a very urgent task.

The great contribution to the digital processing of television images was made by both domestic scientists V.T. Fisenko [3], M.L. Mestetsky [5], V.P. Dvorkovich, A.V. Dvorkovich [1-9], M.K. Chobanu [5-6], V.N. Kozlov [2-7], V.N. Gridin [1-8], V.Yu. Visilter [4], A.L. Priorov [3-4, 9], as well as by L. Shapiro [5-7], R. Gonzalez [1-7], R. Woods [17], G. Finlayson[7-8], C. Wöhler [10], R. Szeliski [6], D. Maier [5, 8].

The number of fundamental studies of Alpatov B.A., Atakishchev O.I., Bashmakov O.E., Bykov P.E., Gurevich S.B., Duda R., Hart P. and others are devoted to the development of methods for detecting and tracking moving objects, image processing and control of objects and purposeful processes. Methods of digital image processing were considered in the works of Gonzalez R., Lukyanitsa A.A., Titov B.C., Filist S.A. Issues related to the transmission of video data were investigated in the works of Zubarev Yu.B., SogdulaevYu.S., etc. Methods of recognition of static and dynamic images based on spatio-

*Corresponding author: saida.beknazarova@gmail.com

https://doi.org/10.1051/e3sconf/202340103068

temporal analysis of video images were covered in the works of Favorskaya M.N., Soifer V.A., Fisenko V.T., Foresight D., Klyuchikov I.A., etc. At the same time, the issues of recognition of moving people (dynamic images) in various situations and in conditions of changing factors (interference, lighting heterogeneity, change of angle, etc.) remain unresolved.

The analysis of the literature has shown that the systems using algorithms of applied television are of the greatest interest. Such systems use the visible part of the electromagnetic spectrum, which is convenient for practical use [2-3]. To date, applied television systems are widely used to perform various kinds of measurement work: diagnostics of the road network [1-2]; detection of pedestrians [2]; detection of obstacles on the runway [2-6]; collision prevention on railways [2-8]; detection of obstacles in front of a mobile ground object [1-5]. All the listed systems using the methods of applied television to perform their task analyze a specific type of obstacle, without solving the problem in the general case. In this regard, a system of applied television based on digital image processing is proposed to solve the problem of detecting obstacles in the room by an autonomous mobile robotic platform, which characterizes in such an obstacle, the system is the main color feature, information, which makes it possible to distinguish the types of underlying surface.

The purpose of the scientific article is to develop models, methods and algorithms for processing complex structured video data based on the use of methods to increase the contrast of television images in video information systems.

Frequency methods of image transformations are based on the idea of the Fourier transform, the meaning of which is to represent the original function as a sum of trigonometric functions of various frequencies multiplied by specified coefficients. An important property is that the function represented by the Fourier transform, after performing transformations on it, can be returned to its original form. This approach allows you to process the function in the frequency domain, and then return to the original form without loss of information. The Fourier transform can also be used to solve image filtering problems. In a practical application, the implementation of frequency approaches can be similar to spatial filtering methods.

Spatial image enhancement techniques are applied to raster images represented as two-dimensional matrices. The principle of spatial algorithms is to apply special operators to each point of the original image. Rectangular or square matrices called masks, kernels or windows act as operators. Most often, the mask is a small two-dimensional array, and improvement methods based on this approach are often called mask processing or mask filtering.

The existing methods of isolating (filtering) significant characteristics of individual image components, some periodic image structures are not optimal from the point of view of Fourier approximation in the specified frequency intervals in which filtering is carried out. Therefore, an urgent problem is the creation of mathematical models and filtering methods that allow for adequate consideration of the energy characteristics of images in selected frequency intervals. The paper develops and theoretically substantiates a method of optimal linear image filtering based on frequency representations, which is optimal in the sense that the spectrum of the image obtained as a result of filtering has the smallest standard deviation from the spectrum of the filtered image in a given two-dimensional frequency subinterval, and outside this subinterval has the smallest deviation from zero.

Usually images are distorted under the influence of various kinds of interference. This complicates both their visual analysis by a specialist and automatic processing using computer technology. Attenuation of the interference effect can be achieved using various image filtering methods. When filtering, the brightness of each point of the source image distorted by interference is replaced by some other brightness value, which is assumed to be distorted to a lesser extent. Such a decision can be made based on the following considerations. The image is represented by a two-dimensional function of spatial coordinates, the values of which change more slowly when moving from point to point of the image than the values of a two-dimensional function describing interference. This allows, when evaluating the value of the useful signal at each point of the image, to take into account a certain set of neighboring points, taking advantage of a certain degree of similarity of the useful signal at these points.

2 Methods

Many of the developed methods and algorithms are associated with a wide variety of plots that have to be described using various mathematical models. In addition, the application of various optimality criteria also leads to the development of various filtration methods. Finally, in many cases, very often, due to mathematical difficulties, it is not possible to find the optimal image filtering procedure. The complexity of finding exact solutions leads to the generation of various variants of approximate methods of digital image analysis.

The image processing process consists of a number of stages, among which one of the most important is the pre-processing of images. Preprocessing and contour extraction on digital images have a wide range of applications in various fields, ranging from image preparation to recognition, image improvement in various recording devices by low-frequency filtering and equalization of brightness histograms - photo and video cameras, scanners, sonar, images obtained using ultrasound, X-ray, radio location, astronomical photographs, electron microscopy, etc.

During preprocessing, an image analysis is performed that determines various statistical characteristics of the image, such as mathematical expectation and standard deviation of brightness, contrast, construction of a histogram of brightness and contrast, selection of the most suitable model and parameters of digital noise [4]. At the stage of preprocessing, low-frequency filtering is performed, which removes digital noise in the image [5].

As a rule, after low-frequency filtering, the contrast of the image decreases and, therefore, it needs to be corrected. To correct the contrast, the contours of the image are calculated. As a result of summing the brightness of the pixels of the image with the brightness of the calculated contours, the contrast correction of the image is carried out [14,15].

The general block diagram of image preprocessing is shown in Figure1. The original raw image is taken as input data A, for which the brightness histogram is calculated H such statistical characteristics of image pixel brightness as mathematical expectation are determined μ, standard deviation σ and the median $\mu_{1/2}$,and also the mathematical expectation is calculated μ_n and the standard deviation σ_n digital noise.

To remove digital noise and enhance the contours of the image A processed by low-frequency filters. The result of low-frequency filtering is an image A', on which the contours are calculated D.

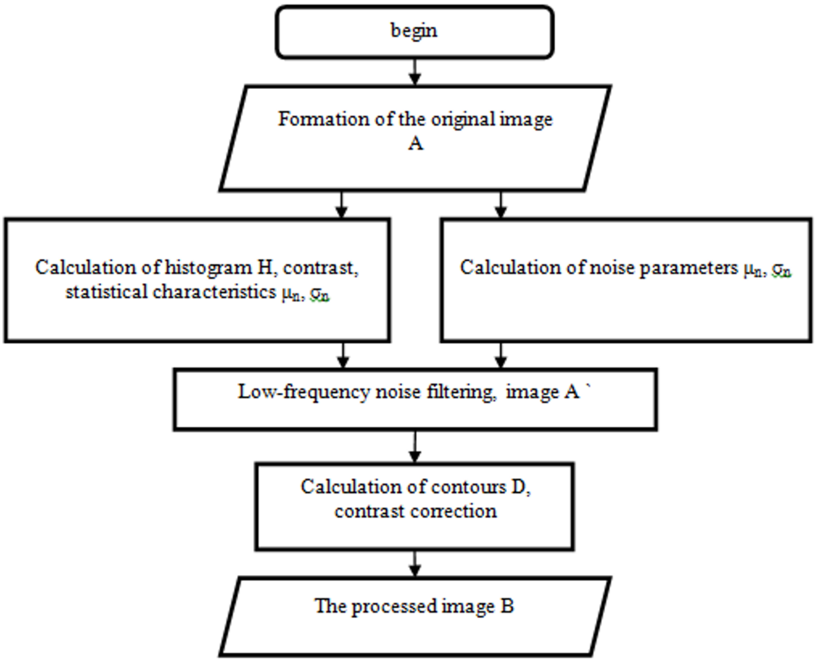

Fig. 1. Block diagram of image preprocessing

Also, the result of preprocessing is an image B about balanced contrast and suppressed low-frequency noise.

For image preprocessing, statistical characteristics, contrast ratio, and noise are calculated on the original image. Then low-frequency noise filtering is performed, removing the noise component of the image. At the final stage, the image contours are searched and contrast correction is performed using the calculated contours.

When forming a digital image, it uses three main definitions of contrast to measure contrast. The Weber contrast is determined by the ratio [2-3]: $C = (I - I_b)/I_b$, where I - the brightness of a single image element at which the contrast is estimated, I_b- background brightness (the brightness of neighboring image elements averaged in some way). Usually, when small details are present in the image against the background of large objects that differ little in color, Weber contrast is used. The disadvantage of such a contrast detection mechanism is a decrease in the calculated value with an increase in background brightness, therefore, such a determination mechanism for calculating the contrast of light images is unacceptable.

For images where the number of dark and light areas is approximately the same, the Michelson contrast is determined by the ratio: $C = (I_{max} - I_{min})/(I_{max} + I_{min})$, where I_{min} and I_{max} accordingly, the minimum and maximum brightness values in the analyzed area of the image, and the denominator is twice the value of the average brightness.

The most common mechanism for determining contrast is the RMS contrast, which is applied to all types of images and is determined by the formula: $C = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(I_i - \bar{I})^2}, \bar{I} = \frac{1}{n}\sum_{i=1}^{n} I_i$, where$I_i$, - brightness of i - the pixel of the area for which the contrast is estimated. The main disadvantage of this definition is its low performance compared to the contrast of Weber and Michelson. The less common way to determine contrast is the formula of V.F. Nesteruk and N.N. Porfiriev: $= (I^{2\gamma} - I_b^{2\gamma})/(I^{2\gamma} + I_b^{2\gamma})$, where I - the brightness of the image element for which the contrast is estimated,I_b- background brightness, γ- parameter that characterizes the physiological properties of a particular object. The disadvantage of this detection mechanism is a large number of conditions under which the contrast value reaches a maximum.

To determine the overall contrast of a digital image using one of the contrast definitions (Weber, Michelson or RMS), local contrast values are calculated in all pixels of the image or in certain groups of pixels, after which the values obtained are somehow averaged. The resulting value is the overall contrast of the image [15].

The R.A. Vorobel method is an alternative method for determining how contrasting a digital image is:$C_L = (I_1 - I_2)/I_{max}$, where I_1, and I_2 - brightness of image elements, I_{max}- the maximum brightness value of the image elements. Thus, the maximum value of local contrast is achieved with the minimum brightness value of one of the elements and the maximum brightness value of the other, and the minimum - with equal brightness of the compared elements. Vorobel proposed a method based on a linear description of local contrasts to calculate the total contrast:

$$C_G = \frac{1}{2I_{max}} \int_0^{\infty} \left|2(I - \bar{I}) + I_{max} - |2(I - \bar{I}) - I_{max}|\right| * h(I)dI$$

where $h(I)$- histogram of the brightness of the analyzed image.

The above-mentioned standard method for determining the contrast of a digital image has a characteristic disadvantage that they provide a qualitative assessment of the contrast of the image. In order to use them to determine how contrasting the image is relative to an image with balanced contrast, it is necessary to evaluate the contrast of any reference image and compare the resulting value with the value calculated for the analyzed image.

The direct implementation of digital mask filters allows sequentially processing blocks of 3, 9, 25, etc. (depending on the dimension of the filter mask) pixels of the input image. The speed of operation, in comparison with software filters, is quite high (hundredths of a second for processing an image with a size of 200 * 200 pixels). This fact suggests that the use of a systolic structure for image processing is justified, and it becomes possible to use digital filters to process a video sequence in real time. The proposed filter option is shown in Fig.2.

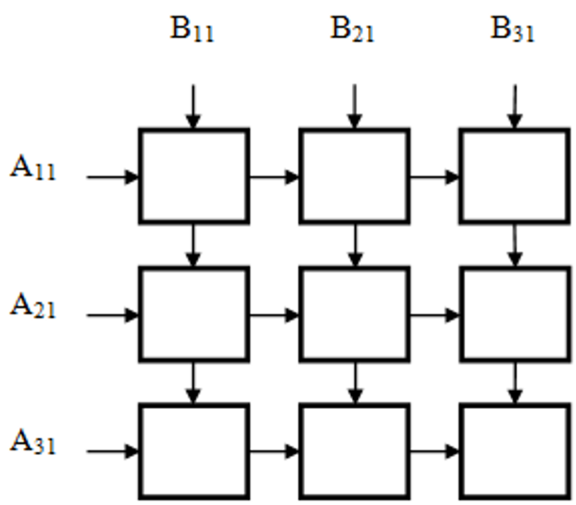

Fig. 2. Systolic structure of the digital filter for the mask *m=3*.

3 Results

Let's take a closer look at the operation of the filter. The task of digital filtering is reduced to multiplication of matrices. Let 's write down the classical multiplication of a row by a column:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} * \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} & a_{11}b_{13} + a_{12}b_{23} + a_{13}b_{33} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} & a_{21}b_{13} + a_{22}b_{23} + a_{23}b_{33} \\ a_{31}b_{11} + a_{32}b_{21} + a_{33}b_{33} & a_{31}b_{12} + a_{32}b_{22} + a_{33}b_{32} & a_{31}b_{13} + a_{32}b_{23} + a_{33}b_{33} \end{bmatrix}$$

With this approach, we have extra terms, because to calculate convolution (mask filtering), we need to multiply the matrices element by element and then add up the results. If we use the transposed matrix B (filter mask), then as a result we get the necessary result by adding the main diagonal of the resulting matrix:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} * \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}b_{11} + a_{12}b_{12} + a_{13}b_{13} & & \\ & a_{21}b_{21} + a_{22}b_{22} + a_{23}b_{23} & \\ & & a_{31}b_{31} + a_{32}b_{32} + a_{33}b_{33} \end{bmatrix}$$

$$z = a_{11}b_{11} + a_{12}b_{12} + a_{13}b_{13} + a_{21}b_{21} + a_{22}b_{22} + a_{23}b_{23}a_{31}b_{31} + a_{32}b_{32} + a_{33}b_{33}$$

Let's consider a variant of pipelining Figure 3 with a bidirectional linear processor array.

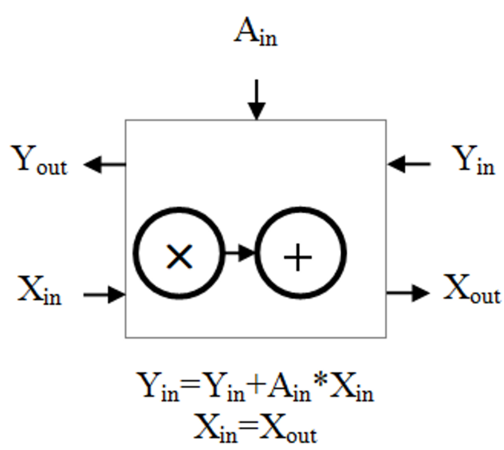

Fig. 3. Conveyor processing

If we model this algorithm, then multiplying matrices requires 5 elements. Use of simple operations and elements, a hardware-oriented image filtering algorithm on neural-like elements will allow for real-time image filtering.

To take advantage of this method of image formation, a block diagram of the sensor is proposed (Fig. 4). In the right part there is a matrix of pixels $M * L$each of which has its own photodiode with a reset transistor, a comparator and a logical control unit. The data memory block for strings is designed to record the counter values at the current time. A frame memory buffer of the same dimension as the pixel matrix is needed to use parallel recording of illumination values and increase the reading bandwidth. For the correct recording of time data, there is a row selection scheme that ensures that the address of the only active row is correctly supplied at each time. From the pixel matrix, signals of all active pixels are sent to memory in parallel to write to the desired columns. With the help of the control unit, the signals of the beginning of reading are transmitted to the digital counter and the pixel matrix, as well as the control of filtering parameters for processing the resulting image [11].

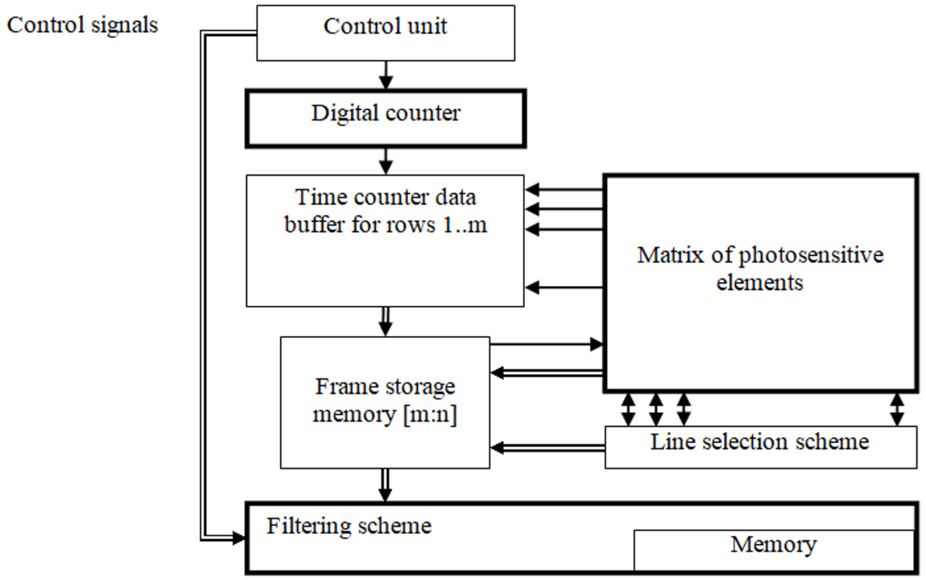

Fig. 4. Block diagram of the image receiving and processing device

4 Discussion

The recording of the digital value of the counter into memory can be considered as a process of analog-to-digital conversion of the time interval.

The filtering scheme is implemented on the basis of a neural-like two-layer network with a linear activation function for the output neuron and a sigmoidal one for the neurons of the hidden layers. Frame processing is performed line by line.

The input signal matrix (input layer) (discrete brightness samples for each pixel of the input image, the size of which) will be written as:

$$Y = \begin{bmatrix} y(1,1) & y(1,1) & y(1,l) \\ y(2,1) & y(2,1) & y(2,l) \\ y(m,1) & y(m,1) & y(m,l) \end{bmatrix}$$

The image matrix and the matrix of neurons are represented as vectors:

$$Y = \left[\begin{pmatrix} y(1,1) & \cdots & \\ \vdots & \ddots & \vdots \\ y(m,l) & \cdots & \end{pmatrix}\right]$$

$$Q1 = \left[\begin{pmatrix} q(1,1) & \cdots & \\ \vdots & \ddots & \vdots \\ q(m,l) & \cdots & \end{pmatrix}\right]$$

Element w_{ij} – that is, the element i – lines j – column - represents the value of the weight of the link connecting j – neuron with i – entrance.

By varying these numbers for different conversion levels and different quadrants, you can control the degree of video data loss in the image, thereby changing the compression ratio and the quality of the restored images. At the same time, to ensure the constancy of the bitrate of the compressed video stream, an adaptive change in the values of the quantization coefficients is used, which maintains the constancy of the frame compression ratio when excessive information changes in them. At the same time, the quantization coefficients calculated in the compressor are stored in the output array for proper operation of the decompress or. However, an increase in video data compression leads to an increase in irreversible data loss, which affects the visual quality of the restored images. Therefore, determining the optimal values of the quantizer is a rather difficult task and requires further research.

One of the most urgent tasks in the field of audio-video data processing is the improvement of audio-video data compression methods, taking into account the elimination of temporary redundancy of TV images and audio accompaniment. This problem is very relevant in conditions of limited frequency resources. In addition, it becomes possible to significantly reduce the preparation time for television reports to be broadcast directly from the event sites by transmitting signals from TV cameras directly to the installation hardware of television centers over cellular networks. At the same time, there is no need to use expensive and not always available broadband communication channels.

As a result of the conducted research, TV images have code, intra-frame statistical, psych visual, structural, temporal or inter-frame redundancy, when eliminated, image information reduction or video data compression is achieved.

Thus, to date, existing image processing algorithms can only achieve 130-150 times compression of the video stream due to a noticeable deterioration in their visual quality. Therefore, to ensure good quality of TV images with a frame size of 8-10 kbytes, it is necessary to develop new effective methods for processing video streams that significantly minimize the amount of metadata (no more than 500 bytes per frame), or do not use motion compensation at all [14].

5 Conclusions

Experimental evaluation of the efficiency of audio file compression by fractal and fractal-spectral codecs. To evaluate the effectiveness of the proposed method of audio signal compression, based on the elimination of temporal redundancy of audio frames, an experimental study was made of compressing audio files of various genres with various audio frame identification errors.

Currently, Haar wavelets are widely used for information compression, which are easy to implement, since they have only 2 coefficients. However, the Haar wavelet is not very suitable for compressing audio signals, since it does not provide a high degree of compression of the SL, since when a large number of transformation coefficients are discarded, distortions occur in the form of extraneous noise, crackling and rumbling. To eliminate this shortcoming, higher-order wavelets can be used, for example, Daubeschi-4th order, which has 4 coefficients and Daubeschi-10, which has 10 coefficients [4,14]. At the same time, higher-order wavelet functions have a “smoother” shape, which allows you to increase the compression ratio while maintaining sound quality. Therefore, to implement the compression algorithm, it is most expedient to use the 10th order Dobshi wavelet, since, on the one hand, it provides greater transformation accuracy, and, on the other hand, it does not significantly reduce the processing speed.

References

1. Artyushenko V.M., Shelukhin O.I., AfoninM.Yu. "Digital compression of video information and sound" I.: Moscow 2003, p. 430
2. D. Vatolin, A. Ratushnyak, M. Smirnov, V. Yukin "Data compression methods" http://compression .graphicon.ru/;
3. D.Salomon. "Data, image and sound compression". // Moscow: Technosphere, 2004 (translated from the English by V. V. Chepyzhov).
4. V.I.Vorobyov, V.G.Gribunin. "Theory and practice of the wavelet transform". //St. Petersburg: VUS, 1999
5. Konstantin Glasman. Video compression. «625» 7/1997. http://rus.625-net.ru/625/1997/07/6.htm
6. Otto S. E. Modeling of a digital video signal compression system in real time based on wavelet transformations. Article in the collection of reports of the Republican Scientific and Technical Conference of Young scientists, researchers, undergraduates and students "Problems of information technologies and telecommunications”, Tashkent, March 15-16, 2012. I vol. pp.82-83.
7. R.Gonzalez, R.Woods Digital image processing / Transl. from English-M.: Technosphere, 2006 — - 1070 p.
8. Guzman et al. Digital image processing in information systems-Novosibirsk: NSTU, 2000. - 156 p.
9. F. Durand and J. Dorsey Fast Bilateral Filtering for the Display of High-Dynamic-Range Images. In John Hughes, editor, SIGGRAPH 2002 Conference Graphics Proceedings, Annual Conference Series // ACM Press/ACM SIGGRAPH 2002 — P. 257—265.
10. Hirschmiiller, Gehrig S. Stereo Matching in the Presence of Sub-Pixel Calibration Errors//IEEE Conf. on Computer Vision and Pattern Recognition. Miami, Florida, USA. - 2009. - P. 437-444.
11. H. Hirschmiiller Accurate and efficient stereo processing by semi-global matching and mutual information. IEEE Conf. on Computer Vision and Pattern Recognition. San Diego, California, USA. - 2005. - V. 2. - pp. 807-814
12. G. Bastin, J. M. Coron and B. d'Andrea Novel, Networks and Heterogeneous Media 4(2), 177-187 (2009).
13. M. Krstic and A. Smyshlyaev, Advances in Design and Control. Society for Industrial and Applied Mathematics (SIAM) 16, 13-22 (2008).
14. Beknazarova, S., Latipova, N., Maxmudova, .J., Alekseeva, V.S. Filtering of digital images by the convolution method/ International Journal of Mechanical Engineering, 2022, 7(1), p. 1182–1191
15. Beknazarova, S., Muhamadiyev A. Sh., Iterative methods for solving managed problems of the television image processing / International Conference on IT in Business and Industry (ITBI 2021). Journal of Physics: Conference Series. 2032 (2021) 012002. Omsk-2021.-P.1-10