TECHNOLOGY OF SEARCHING OBJECTS IN VIDEO DATA

Saida Safibullayevna Beknazarova[1*], Dildora Bakhtiyarovna Umarova[2]

Article History: Received: 17.03.2023 **Revised:** 03.05.2023 **Accepted:** 17.06.2023

Abstract

The article dedicated to video analytics is a technology that uses computer vision methods for the automated acquisition of various data based on analysis sequence of images from video cameras in real-time or from archival records. The task of discovery of dynamic objects is understood as the task of detection and selection of changing areas of the image in a sequence of frames. Accordingly, the detection of a certain object means the choice of one or more detected dynamic objects that have some similar features to a given search object. Features are selected according to the algorithm. The search process object is complicated by affine, projective distortions, overlapping object by other objects, and receiver (sensor) noise. For real practical applications, the task is to process the video sequence at the real speed of receiving the data stream.

[1*s]Journalism and Mass Communications University of Uzbekistan, Tashkent, Uzbekistan
[2]Tashkent University of Information Technologies named by Muhammad Al- Khwarizmi, Tashkent, Uzbekistan

Corresponding author: [1*]saida.beknazarova@gmail.com

DOI: 10.31838/ecb/2023.12.s3.491

1. INTRODUCTION

The widespread use of digital video surveillance systems leads to the need for the constant processing of digital signals in the form of a digital image. Modern video monitoring lines are based on the latest technical means that allow digital processing of the received video image [1]. Digital video surveillance is effective because its operation can be linked to computer systems. One of the main tasks that digital video surveillance systems solve is the task of providing video monitoring of the observed object. In video monitoring and access control systems, the processing and identification of digital images should be as fast as possible. The autonomy of objects of digital video surveillance systems imposes restrictions on the size of such devices and their energy efficiency. For this reason, there is a continuous search for new, better, and more reliable means of digital signal processing (DSP), as well as the improvement of known processing methods [2-4].

On the other hand, video surveillance systems that work directly with digital signals in the form of an image must solve three main tasks that arise during the operation of the system. These tasks consist of processing, storing and analyzing the digital signal received at the input [3]. The task of signal processing requires the use of effective digital filters in the objects of digital systems, which in turn change the image in accordance with the processing algorithm used. The analysis of the digital signal in video surveillance systems consists of the recognition (identification) of the received signal in the form of an image [7, 8]. Thus, solving the problems of image processing and its analysis are closely related, since the use of advanced processing methods will improve recognition results [5,13].

In turn, the solution of the designated tasks facing the objects of digital video surveillance systems impose certain conditions on their development. There is a contradiction between the need to design the main computational structures of video surveillance objects, allowing, on the one hand, to best transform the processed image for further pattern recognition, and on the other hand, to minimize the time spent on their processing in real-time [3, 4]. The solution to this contradiction is the parallel organization of computing, which is a promising tool in solving the problem of minimizing the time required for processing. One of the ways to ensure the parallel execution of operations at the arithmetic-logical level is the use of non-positional number systems instead of the traditional binary number system, which is a positional number system[9,10].

The purpose of the research is to develop mathematical models, methods and algorithms for improving the speed of digital image processing filters of video surveillance systems based on the integration of a system of residual classes and analyze an adaptation of the brightness control mechanism for processing dynamic image sequences.

A significant scientific contribution to the theory of modular computing and its applications were made by domestic and foreign researchers: I.Y. Akushsky, D.I. Yuditsky, V.M. Amerbayev, A.A. Kolyada, A.I. Galushkin, I.T. Pak, M.V. Sinkov, V.A. Torgashev, N.I. Chervyakov, I.A. Kalmykov, O.A. Finko, D. Svoboda, N. Szabo, M. Valach, H.L. Garner, A.S. Fraenkel, A. Huang, B. Purhami, W. Ienkns, H. Krisha, A. Omondi, A. Premkumar, I. Ramires, A. Curcik, L.Yang, D. Zhang, P Steffan, G. Pirlo, L. Sousa and others.

Additionally, the quality of images is indirectly affected by the amount of compression of audio signals, since both video and sound are transmitted in a single stream at a speed of 2 Mbit/s. At the same time, the volume of audio information, depending on the sound quality, can reach 10-20% of the video. Therefore, we must strive to increase the compression of audio signals. In this regard, the works aimed at improving the methods and algorithms of compression of streaming video and sound, to increase the compression coefficients without a noticeable decrease in the quality of the restored images and sound are of great scientific and practical importance.

2. METHODOLOGY

2.1 Definition of requirements

Based on the formulated goal and the conducted analytical review, it is possible to determine the requirements for the object search system being developed in the video stream. These requirements can be divided into two groups: basic (functional) and additional (structural). The main requirements can be attributed to the search technology and algorithms of the system, and additional requirements can be attributed to the system's structure.

The main requirements for the object search system in the video stream are formulated based on the results of an analytical review and the needs of practical use:

1. The system's algorithm should be based on the search for key points of the object.
2. The computational complexity of the search method should be minimally achievable for use in solving real-time problems.
3. The system's algorithm should not contain a step of preliminary training.
4. The system should allow you to control the way the result is processed: save the result as text information, as a set of images with a selected found object, or display a set of frames with selected objects on the screen.

Based on the requirements of the video stream processing speed and ease of implementation, the structural requirements can be reduced to the following: the system structure model should have a minimum number of static links. It follows from this that the desired system structure should be static and networked.

Functional model

According to the formulated requirements, the system's algorithm should be based on the search for key points. It follows from the review of the technical literature that the methods of searching for key points have the following features that must be taken into account when using:

1. The key point may not be part of the image of only one object; it may also occur in images of other objects since the key point is a local area in the image of a relatively small size and may have "similar" duplicates in the image due to low information content.
2. The key points may not be located on the entire area of the object image, but only in some areas, i.e., it is impossible to judge the location of the entire object image by the found key points, therefore determining the location of the object position in the image based only on the positions of the key points is incorrect.

Local features do not guarantee reliable identification due to the possibility of repetition of features on other objects and the possibility of distortion. In this regard, to increase the reliability of identification, it is necessary to expand the vector of local features with one of the global ones that allows identifying an object. The color histogram should be distinguished from the global features since the color histogram is calculated quickly and does not depend on the rotation and scale of the image.

The model of representation of an object image in the form of a set of integral and local features of the object image is proposed:

$$F(x,y) \rightarrow \begin{cases} K^{ref} = \{K_0^{ref}, \dots, K_{m-1}^{ref}\} \\ H^{ref} = \{H_0^{ref}, \dots, H_{n-1}^{ref}\} \end{cases}$$

where $F(x, y)$ is rectangular matrix of pixel intensity values, K^{ref} is vector of key points of the image of an object of dimension m, H^{ref} is color histogram of an object image consisting of n elements.

To improve performance, it is proposed to perform the following actions before starting the analysis of the video stream:

1. Create a set of images with projective distortions of the sample.
2. Find key points on each created image.
3. Calculate the descriptors of each key point.

The described steps for calculating key point descriptors on pre-projectively distorted sample images allow using algorithms to search for and describe key points that are unstable to projective transformations.

Thus, the general scheme of the search algorithm can be presented in the form of two stages [1]:

1. Quickly search candidate areas with an object image using key points.
2. Identify an object in an image based on a global feature – a color histogram.

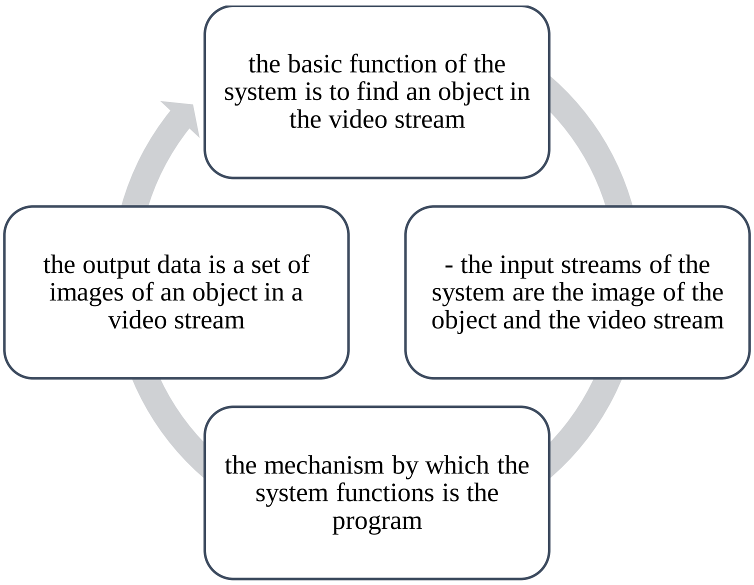
Based on this scheme, the formulated requirements for the system, and the proposed method for calculating descriptors on projectively distorted images of the sample, it is possible to determine the set of functions that the system should have, to carry out a functional design. The functional model is described based on the IDEF0 notation [2].

3. REALIZATION OF THE CONCEPT

3.1 Top-level diagram

When considering the system from the point of view of the top-level function, the following provisions are highlighted

- the basic function of the system is to find an object in the video stream;
- the input streams of the system are the image of the object and the video stream;
- the mechanism by which the system functions is the program;
- the output data is a set of images of an object in a video stream.



3.2 First level diagram

As a result of the primary decomposition of the top-level diagram A-0, the diagram A0 is obtained (see Figure 2).

The decomposition of the top-level function of the system reveals the following main functions:

- the function of calculating the vector of descriptors from a multidimensional matrix of projectively distorted images of an object (A1);

- the function of calculating the histogram of the object image (A2);
- the function of extracting a frame from a video stream (A3);
- function for calculating the vector of key points of a video stream frame (A4);
- the function of finding the areas of applicants of the object on the frame (A5);
- the function of identifying an object in the found areas of applicants based on the calculation of color histograms (A6).

Diagram A0 highlights the main information flows:

1. A stream containing an image of the desired object.
2. Video stream. This stream includes a set of frames on which the object is searched.

The identification function for the selected areas of the frame contains the area with the closest match of the histogram of the image of the desired object.

3.3 Second-level charts

The function of calculating the vector of descriptors of the multidimensional matrix of images of object A1 is decomposed into stages (see Figure 2), the mechanism of which is the program:

1. Converting the image to grayscale (A11). The control is carried out by the formula for converting the image into a halftone.
2. Generation of scale changes (A12) according to the control–scale parameters. At this stage, a vector of scaled images of the object is created.

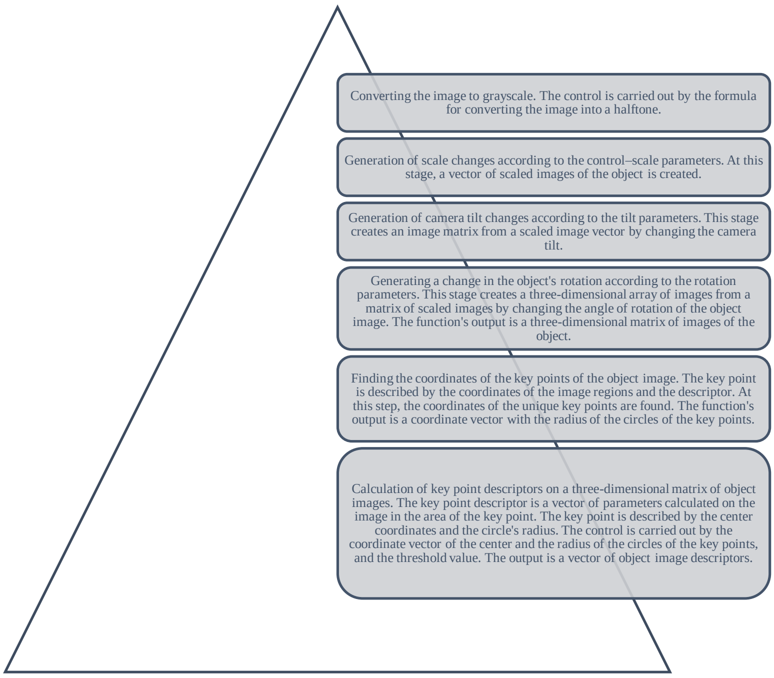


Fig. 2. Context diagram of the first level

3. Generation of camera tilt changes (A13) according to the tilt parameters. This stage creates an image matrix from a scaled image vector by changing the camera tilt [10].
4. Generating a change in the object's rotation according to the rotation parameters (A14). This stage creates a three-dimensional array of images from a matrix of scaled images by changing the angle of rotation of the object image. The function's output is a three-dimensional matrix of images of the object.
5. Finding the coordinates of the key points of the object image (A15). The key point is described by the coordinates of the image regions and the descriptor. At this step, the coordinates of the unique key

points are found. The function's output is a coordinate vector with the radius of the circles of the key points.

6. Calculation of key point descriptors (A16) on a three-dimensional matrix of object images. The key point descriptor is a vector of parameters calculated on the image in the area of the key point. The key point is described by the center coordinates and the circle's radius. The control is carried out by the coordinate vector of the center and the radius of the circles of the key points, and the threshold value. The output is a vector of object image descriptors.

4. RESULTS AND DISCUSSION

Just as for (4), ordering the values produced cth feature. The described steps A11, A12, A13, A14, A15, and A16 are the steps of the method for calculating descriptors on projectively distorted images of the sample.

The result of the decomposition of the frame extraction function from the A3 video stream is shown in Figure 3. During the execution of the function, the video stream goes through the following stages:

1. Demultiplication. This step is necessary to extract the encoded video signal from the container (A 31).

2. Decoding. The stage is necessary to represent the encoded video signal into a set of images that are convenient for processing in the object search function in the video stream (A32) [12].

3. Frame selection. According to the frame parameters, frame is extracted from the decoded video stream.

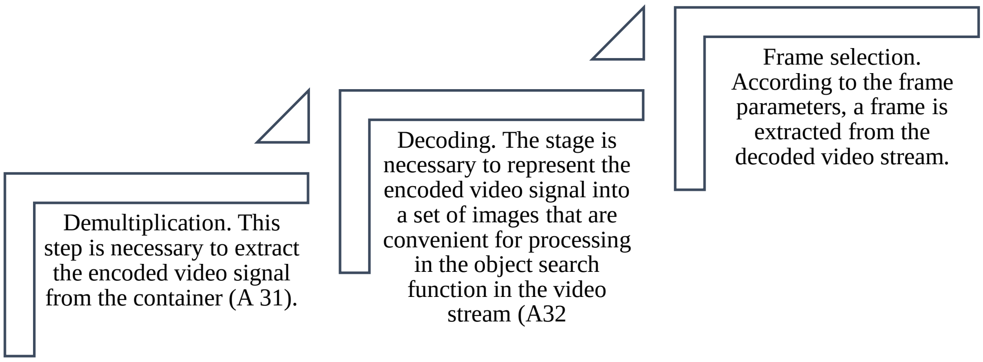


Fig. 3. Diagram of the image extraction function from the video stream (A3)

The function of calculating the vector of key points of the A4 frame is decomposed into the following stages (see Figure 4):

1. Conversion to halftone (A41);

2. Finding the coordinates of the key points of the frame (A42);

3. Calculation of descriptors in the found frame coordinates (A43).

The coordinate vector of the key points is needed to calculate the descriptors (A43). The function output is a vector of key points with information about the frame number, coordinates, and descriptor for each key point.

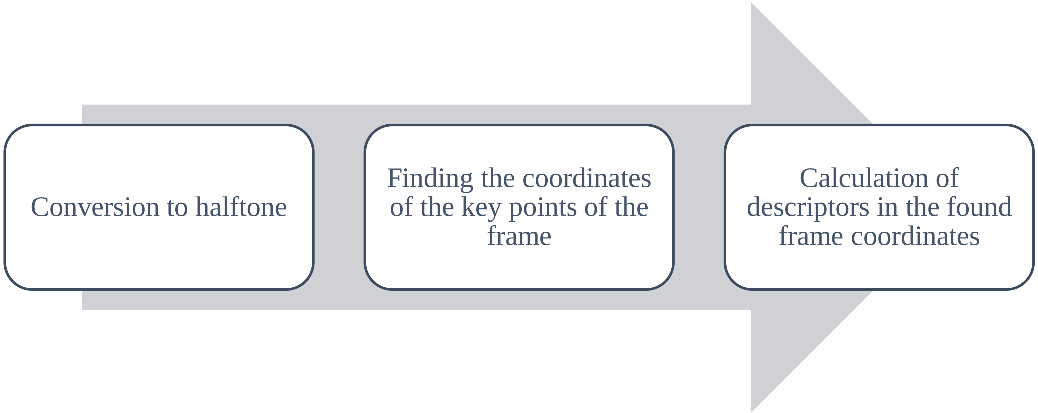


Fig. 4. Diagram of the function for calculating the vector of key points of the frame (A4)

The function of finding the candidate areas of the object on the frame (A5) is decomposed into components (see Figure 5):

1. Finding common key points (A51). The input receives a vector of key points of the frame. The control is carried out by the vector of object descriptors. Using the proximity measure, the key points are determined, the descriptors of which are considered the same as the descriptors of the object image, i.e., the images of the areas indicated by the key points are considered "similar". If a non-empty set is obtained after the intersection, common key points and information about the frame number are output [11].

2. Frame skipping (A52). According to the information about the frame number, a decision is made to skip the input frame of the video stream for further processing.

3. Getting the applicant areas (A53). The input vector of the coordinates of the key points of the frame is combined according to the size of the window in the areas that are fed to the output of the function.

The central functions of the system are the functions of finding the key points of the object image (A15) and frame (A42), calculating the descriptors of the key points of the object image (A16) and frame (A43), finding common key points by descriptors (A51) and identifying the frame for the presence of an object image (A6). The quality and speed of object detection in the video stream depend on the result of performing these functions. Further, algorithms for finding key points together with the calculation of descriptors, algorithms for finding the intersection of descriptors, and algorithms for identifying the image of an object in the frame area by calculating the histogram are considered in more detail.

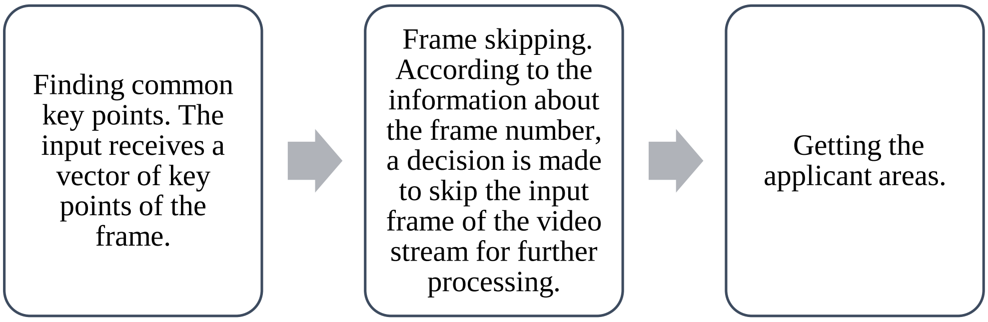


Fig. 5. Diagram of the function of finding the areas of the object's applicants on the frame (A5)

5. CONCLUSIONS

1. The proposed method, as shown by the results of the study of its effectiveness, provides good quality of the restored video sequence with a higher compression ratio value than existing methods.

2. The developed fast algorithms of multidimensional Hartley transformations and PREP allowed to reduce computational costs.

3. Software and algorithmic tools have been developed that implement the proposed codec model, provide encoding and decoding of video files with the possibility of changing encoding parameters, as well as illustrating the process of codec functioning in the form of intermediate results of its operation in combination with a user-friendly interface.

Security systems based on video technologies are widely used in various spheres of human activity. The development and use of production technology management systems, security video surveillance, access control and management prove that video technologies can also solve fire safety problems.

6. REFERENCES

1. Yu G., and Morel J. M. ASIFT: An algorithm for fully affine invariant comparison. Image Processing On Line, 1, pp. 11-38, (2011).
2. Viola P., and Jones M. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, Vol. 1, pp. I-I. (2001).
3. Lowe D. G. Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60, pp. 91-110, (2004).
4. Tvoroshenko I., and Zarivchatskyi R. Analysis of existing methods for searching object in the video stream. (2020).
5. Dawson-Howe K. A practical introduction to computer vision with opencv. John Wiley and Sons. (2014).
6. Bernatovich A. S. An active experiment in the identification of functional systems for the operational implementation of simulation-type models. Cybernetics. No. 1. pp. 99-104. (1983).
7. Ayvazyan S. A., Bukhstaber V. M., Enyukov I. S., Meshalkin L. D. Applied statistics. Classification and dimension reduction. Moscow: Finance and Statistics, p. 608, (1989).
8. Anderson D. Tannehill G. Pletcher R. Vychislitelnya gidromehanika i petloobmen, Moscow, (1990)
9. Sedova N., Sedov V., Bazhenov R., Karavka A., and Beknazarova S. Automated Stationary Obstacle Avoidance When Navigating a Marine Craft. In 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), pp. 0491-0495. (2019).
10. Beknazarova S., Mukhamadiyev A.Sh. Jaumitbayeva M.K. Processing color images, brightness and color conversion. In International Conference on Information Science and Communications Technologies ICISCT 2019 Applications, Trends and Opportunities. (2019)
11. Beknazarova S., Insu M. A. S. P., and Adbullayev S. The Mask Of Objects In Intellectual Irrigation Systems. In International Conference on Information Science and Communications Technologies ICISCT 2020 Applications, Trends and Opportunities. (2020).
12. Beknazarova S., Sadullaeva S., Abdurakhmanov K., and Beknazarov, K. Nonlinear cross-systems of numerical simulation of

diffusion processes. In International Conference on Information Science and Communications Technologies ICISCT 2020 Applications, Trends and Opportunities. (2020).