УДК 031.023:031.25

PROCESSING OBJECT IN A VIDEO STREAM USING BIG DATA

S.S. Beknazarova
Professor of the Department of Television and Media Technologies of M.Khorazmiy TUIT, Doctor of Technical Sciences, Professor
saida.beknazarova@gmail.com

С.С. Бекназарова
Окончила Ташкентский университет информационных технологий. Область научных интересов связана с разработкой методов и алгоритмов построения информационно-компьютерных систем, обработкой медиаресурсов: видео, аудио, организацией учебного и научно-исследовательского процессов в техническом университете.

Аннотация. The article is devoted to the study of methods for processing objects in a video stream using big data technologies. With the rapid growth of video data volumes and the need for operational analysis, traditional processing methods are becoming insufficient to effectively work with such information arrays. Approaches involving the use of distributed computing, machine learning algorithms, and real-time analytics to extract meaningful information from a video stream are considered. Special attention is paid to the issues of data collection, storage and processing, as well as the problems that arise when integrating big data technologies into video surveillance systems. The article discusses both technical and practical aspects, including the possibilities of using these technologies in various fields such as security, smart cities, and behavior monitoring. The results of the study may be useful for video surveillance system developers, data analysts, and researchers in the field of artificial intelligence.

Ключевые слова: big data, advanced analytics, video analytics.

Introduction. Modern video surveillance and video data processing technologies are rapidly developing, providing new opportunities for analyzing large amounts of information. One of such promising areas is the processing of objects in a video stream using big data analysis methods. Video streams are a series of images that can contain important information about the movement, behavior, and interaction of objects in various fields, from security and monitoring to medicine and commerce.

Video stream processing requires high-performance computing and efficient algorithms to extract and interpret data in real time. At the same time, as the amount of data generated by various video cameras increases, traditional processing methods become insufficient for adequate processing and analysis of such data. In this regard, the use of big Data technologies such as distributed computing, machine learning and real-time analytics opens up new horizons in the field of video stream processing.

The current stage of development of information and computing systems is characterized by the widespread introduction of multimedia technologies, and therefore there is a need to develop methods and algorithms for compressing digital video images and video streams based on processing groups of images with information redundancy. The ever-increasing computational complexity of video sequence processing algorithms and high data storage costs require significant computing resources. Video compression not only makes it possible to use digital video in a transmission medium that does not support the original (uncompressed) video images, but also increases the efficiency of using high-speed communication channels for transmitting high-resolution video streams, including simultaneous transmission of multiple streams of high-quality video data[1-3].

Many other prominent scientists and specialists have also made a significant contribution to the theory of video data compression.

Existing video data processing systems use lossy compression methods due to the transmission of reference (I-frames) compressed by spatial coordinates, as well as reference (P- and B-frames) compressed by spatial-temporal coordinates, the number of which determines the degree of compression of the video sequence. At the same time, the peculiarities of the perception of this type of data by a person are taken into account. To date, the problem of eliminating the redundancy of reference frames has been solved to a certain extent, whereas this problem is relevant for reference frames [4-6].

The basis for encoding reference frames is the construction of motion vectors of individual pixels or groups of them. Also, the efficiency of algorithms for eliminating time redundancy is improved by increasing the efficiency of encoding algorithms by transformation and interpolation. This dissertation work is devoted to these problems.

The aim of the article is to improve existing and create new methods for eliminating information redundancy of video sequences, characterized by reduced computational costs with high compression ratio and quality of the restored video sequence.

Methodology

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 structure of the system [7].

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 algorithm of the system 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 algorithm of the system 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 be with 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 algorithm of the system 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.

The presence of local features does not guarantee reliable identification due to the possibility of repetition of features on other objects and the possibility of distortion. In this regard, in order 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 [8].

The model of representation of an object image in the form of a set of integral and local features of an 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 $F(x,y)$ – rectangular matrix of pixel intensity values, K^{ref} – vector of key points of the image of an object of dimension m, H^{ref} – 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 [9].

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

1 Quick search of candidate areas with an object image using key points.

2 Identification of 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 functional design. The functional model is described based on the IDEF0 notation [10].

Realization of the concept

Top-level diagram

When considering the system from the point of view of the top - level function, the following provisions are highlighted (see Figure 1):

– 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.

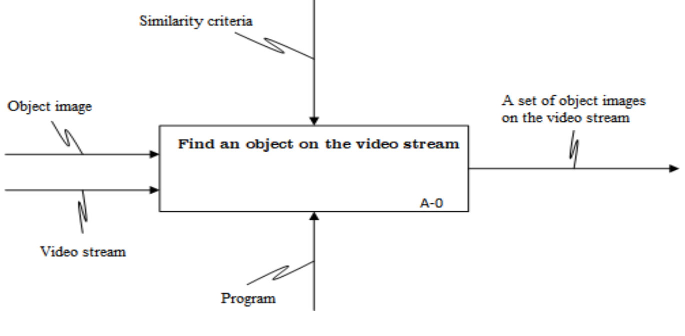

Figure 1. Top-level context diagram

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.

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 3), 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.

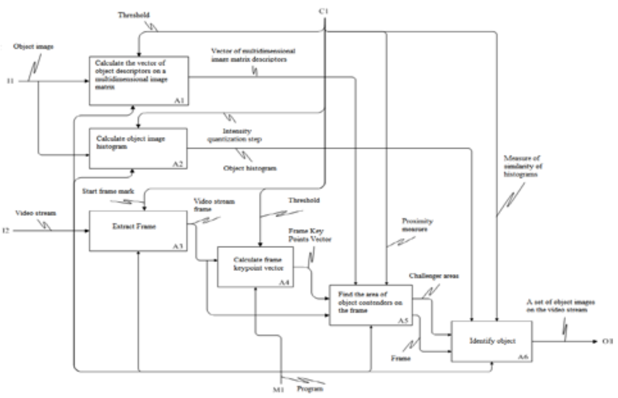

Figure 2. Context diagram of the first level

3 Generation of camera tilt changes (A13) according to the tilt parameters. This stage creates an image matrix from a vector of scaled images by changing the camera tilt [10].

4 Generating a change in the rotation of the object 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 output of the function 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 output of the function 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 coordinates of the center and the radius of the circle. 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.

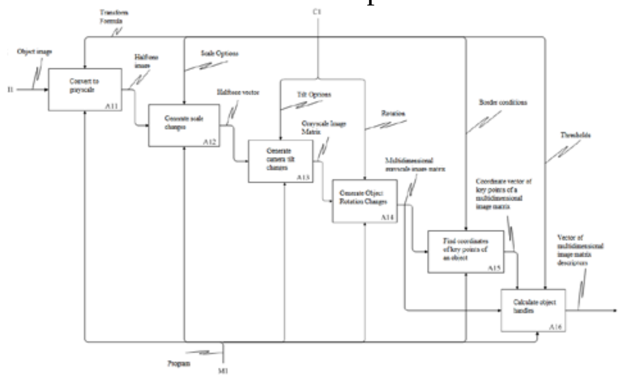

Figure 3. Diagram of the function for calculating the vector of object image descriptors (A1)

Discussion of results. Just as for the (4), ordering the values produced c th 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 4. 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. At this stage, according to the frame parameters, a frame is extracted from the decoded video stream.

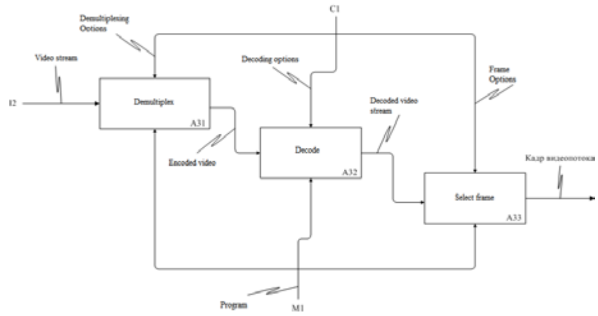

Figure 4. Diagram of the image extraction function from the video stream (A3)

The function of calculating the vector of key points of the A4 frame is decomposed into the following stages (see Figure 5):

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 output of the function is a vector of key points with information about the frame number, coordinates and descriptor for each key point.

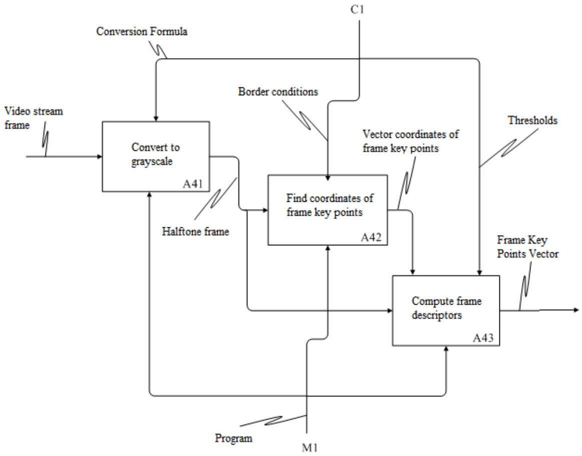

Figure 5. Diagram of the function for calculating the vector of key points of the frame (A4)

The function of finding the candidate areas of the object on the frame (A5) is decomposed into components (see Figure 6):

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 depends 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.

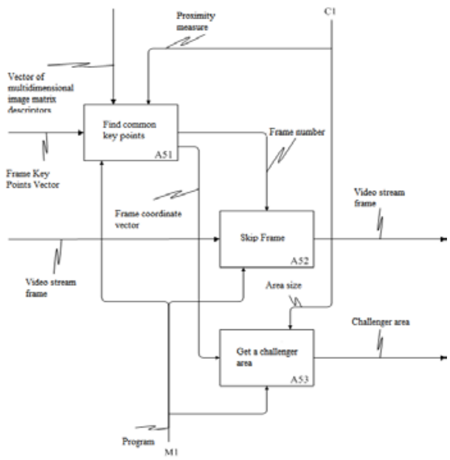

Figure 6. Diagram of the function of finding the areas of the object's applicants on the frame (A5)

The practical value of the results

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 value of the compression ratio compared to existing methods.

2 The developed fast algorithms of multidimensional Hartley transformations and PREP allowed to reduce the amount of 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 the use of video technologies are widely used in various spheres of human activity. The development and use of production technology management systems, security

Modern video stream processing methods using big data technologies open up new possibilities for analyzing and interpreting information extracted from video images. However, the use of such technologies is associated with a number of both technical and practical challenges that must be considered for the successful implementation of effective video processing systems.

One of the main advantages of using big data is the ability to work with huge amounts of video data that can come in real time from multiple cameras. In traditional video surveillance systems, processing such volumes of information would become impossible without the use of distributed computing, such as cloud computing and the use of cluster systems. This allows you to efficiently distribute the load across multiple nodes, reducing processing time and ensuring system scalability.

In addition, the use of machine learning and deep learning techniques significantly improves the accuracy and speed of extracting meaningful information from a video stream. For example, algorithms for object recognition, motion analysis, and behavior prediction can work in real time, automatically classifying and tracking objects, which significantly improves the efficiency of video surveillance. This is especially true in areas such as security and monitoring of public places, where high accuracy and speed of data processing are required.

However, despite the obvious advantages, the use of big data technologies in video stream processing faces several challenges. One of them is the high demands on computing power and data storage. Real-time video processing requires significant resources, which can be difficult if there are limitations on computing power or a limited budget. For such systems to work effectively, it is also necessary to ensure a high level of algorithm optimization, which becomes even more difficult when it comes to dynamic and diverse video streams.

Another challenge is ensuring confidentiality and data security. When using video surveillance, it is necessary to take into account the risks of data leaks and ensure the protection of personal information. Therefore, in addition to technical solutions, it is necessary to pay attention to the legal and ethical aspects related to the use of video surveillance and big data analysis. Data storage and processing issues should be regulated accordingly in order to avoid violating user privacy.

The issue of data accuracy and quality is also important. Video streams can be subject to various interferences, such as changes in illumination, noise, object blocking, or distortion, making it difficult to accurately extract information. This poses the challenge for researchers to develop more resilient and adaptive algorithms that can work with incorrect or incomplete data.

In addition, it is important to note that the use of big data technologies in video stream processing opens up broad prospects for various industries. For example, within the framework of smart cities, such systems can be used to monitor traffic, analyze human flows, as well as to prevent offenses and optimize the operation of urban infrastructures. In the medical field, such technologies can be used to analyze video from medical devices or in the field of telemedicine to monitor the condition of patients.

In general, despite the existing challenges, the integration of big data technologies into video stream processing has enormous potential and opens up new opportunities for creating efficient and intelligent real-time data monitoring and analysis systems.

Conclusion. Processing objects in a video stream using big data technologies is a promising and highly efficient field that continues to evolve taking into account modern requirements for speed and accuracy of analysis. As a result of the integration of machine learning methods, distributed computing and real-time analytics, it becomes possible to efficiently extract meaningful information from huge amounts of video data, which significantly expands the possibilities of using such systems in various fields such as security, monitoring, smart cities and medicine.

However, despite the clear advantages, the technology faces a number of challenges. High demands on computing power, difficulties in ensuring data confidentiality, as well as problems with the quality of the video stream require the development of new approaches and algorithms to increase the stability and accuracy of processing. These challenges highlight the need for further research and improvement of existing technologies.

Given the rapid growth of video data volumes and the need for its operational processing, the future of processing video streams using big data looks promising. This opens up new horizons for the development of intelligent systems capable of not only efficiently analyzing data, but also predicting events, increasing security and improving the quality of life in various fields.

In conclusion, it can be argued that the use of big data technologies for video stream processing has significant potential, which requires further efforts to optimize and adapt to specific tasks and conditions.

Reference

[1] Davenport, T. H. (2014). Big data at work: Dispelling the myths, uncovering the opportunities. Harvard Business Review Press.

[2] Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS quarterly, 36(4), 1165-1188.

[3] Madaan, R., & Belyaeva, Z. (2016). Predictive analytics in education using big data. Procedia Computer Science, 98, 461-466.

[4] Kalyuga, S., & Nastashkin, M. (2018). Implementing big data analytics in education sector to improve student outcomes. In Proceedings of the International Scientific Conference "Digitalization of Modern Society" (pp. 241-245).

[5] Kizilcec, R. F., Papadopoulos, K., & Sritanyaratana, L. (2014). Showing face in video instruction: Effects on compliance, emotion, and learning. In Proceedings of the first ACM conference on Learning @ scale conference (pp. 189-194).

[6] Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. American educational research journal, 38(4), 915-945.

[7] Soundararajan, A., Alfayly, A., Ravindran, P., & Shah, V. (2016). Predicting academic performance of students using big data analytics. In 2016 IEEE international congress on big data (pp. 313-320). IEEE.

[8] West, E., & Tomlinson, C. A. (2016). Learning personalization in the digital age. Educational Leadership, 73(6), 34-39.

[9] Beknazarova, S., Abdullayev, S., Abdullayeva, O., Abdullayev, Z. Machine Learning Method for Predicting Human Movements/AIP Conference Proceedings 3244(1)

[10] Beknazarova, S., Abdullayeva, O., Abdullayev, S., Abdullayev, Z. Images processing: Face recognition by neural networks/E3S Web of Conferences 587.

Авторский вклад

Бекназарова Саида Сафибуллаевна – исследования по внедрению Big Data and Advanced Analytics в образовательный процесс.

ОБРАБОТКА ОБЪЕКТА В ВИДЕОПОТОКЕ С ИСПОЛЬЗОВАНИЕМ БОЛЬШИХ ДАННЫХ

С.С. Бекназарова
Профессор кафедры
телевизионные и
медиатехнологии ТУИТ им.
М.Хоразмий, доктор
технических наук, профессор

Аннотация. Статья посвящена изучению методов обработки объектов в видеопотоке с применением технологий больших данных. В условиях стремительного роста объемов видеоданных и необходимости их оперативного анализа, традиционные способы обработки становятся недостаточными для эффективной работы с такими массивами информации. Рассматриваются подходы, включающие использование распределенных вычислений, алгоритмов машинного обучения и аналитики в реальном времени для извлечения значимой информации из видеопотока. Особое внимание уделяется вопросам сбора, хранения и обработки данных, а также проблемам, возникающим при интеграции технологий больших данных в системы видеонаблюдения. В статье рассматриваются как технические, так и практические аспекты, включая возможности применения данных технологий в различных областях, таких как безопасность, умные города и мониторинг поведения. Результаты исследования могут быть полезны для разработчиков систем видеонаблюдения, специалистов по аналитике данных и исследователей в области искусственного интеллекта.

Ключевые слова: большие данные, продвинутая аналитика, видеоаналитика.