Software for generation of video files metadata

M Shakurov$^{1}$, D Korobkin$^{1}$, S Fomenkov$^{1}$ and A Golovanchikov$^{1}$

$^{1}$Volgograd State Technical University, 28, Lenina ave., Volgograd, 400005, Russia

E-mail: dkorobkin80@mail.ru

Abstract. In today's world, information is one of the most important resources. Its large turnover in the media makes companies look for ways to automate processes related to labeling and storage of data. But often these processes are left on the shoulders of archive workers who have to review the footage in real-time and create metadata based on what they see, which will be transferred into the archive. This process is often very time-consuming. As a result of the analysis of approaches to video data processing, comparative analysis of the efficiency and availability of computer vision services, a software for searching faces, and forming metadata in the video archive of the VGTRK company was designed and implemented.

1. Introduction

At present, the main function of the archivist of a branch of VGTRK "Volgograd-TRV" is a "manual" review of all video files. Every day, on average, more than 20 hours of finished news stories, the source material for topical and news programs are shot. The archivist generates metadata for the video file, which contains information about specific people who appear in the file, locations, the actions performed.

The automation of the process of identifying specific persons together with the timecodes in the video files of the archive is an urgent task that helps to intensify the work of the archivist. In this regard, the purpose of this article is to introduce the reader to the developed software for finding important persons and forming metadata.

The subject area of this paper is closely related to the business processes taking place in the VGTRK company. Video material that goes to the video archive goes through several stages before the final reservation on the servers, most of the processes are automated with the help of the "Dalet" software installed on the company servers, but video material description is still done in real-time, which takes much time and distracts employee from other, more important things.

Currently, no software on the Russian market that provides an interface and automatic analysis software for TV companies, thereby allowing them to fully automate this process. In this connection, a task was set to create software to solve this problem.

The main task of the archivist is to get the footage into a long-term tape repository. In the process, the archivist will need to describe the archived material in detail (location, personalities, keywords, etc.). The video description is the main and most important task of the archivist, as the possibility of further use of the archived video depends on this information. Next, let us consider the existing algorithm of the VGTRK company archivist "GTRK "Volgograd-TRV".

Figure 1 shows the work algorithm of the archivist.
As can be seen from the presented algorithm, the main part of the archivist's time takes the process of viewing and selecting meta-information from the video material. In the current implementation, this action completely occupies the attention of an employee and does not allow you to integrate into the workflow other tasks.

2. Description of the automated process

When using the face search and metadata generation software, the archivist can work in parallel with the program, thereby increasing the speed of work, or set the system a queue of tasks for processing for long time intervals, for example, on weekends. When evaluating the metadata received from the software, it should be taken into account that the program can determine a person with insufficient accuracy or with a large error, which ultimately leads to distortion of the archive data. When transferring data to the Dalet software, the archivist must take into account the accuracy, with which a particular person would be found. If the accuracy is low, it makes sense to double-check the specified time interval in the video file in "manual" mode.

Figure 2 shows a diagram of the archivist's automated work process.
As you can see, when using the automatic face search and metadata generation software, the archivist can combine the process of analyzing video material and other work processes, which leads to a qualitative increase in labor productivity.

3. Technical solutions

The analysis affects only the key I frames of the video material. To determine their index in the array of all frames it was decided to use the auxiliary software ffprobe.

Face recognition \[1,2\] consists of many interrelated subtasks: one must find all the faces in the image, recognize each face even if it is rotated or poorly lit, identify the unique facial features that distinguish one person from others (such as eye size, face shape, and so on), then compare the identified unique features of that face with all the people the software already knows to understand who is in the image.

The software uses a method invented in 2005, the directional gradient histogram \[3\].

To find faces in an image, you first have to make it black and white, since no color data is needed to find faces. Then consider each pixel. The software will not be interested in the pixel itself, but the pixels surrounding it. The goal is to find out how dark the current pixel is compared to its neighbors. Then you need to replace the pixel with an arrow indicating in which direction the image becomes darker. By following this procedure for each pixel, each pixel will be replaced by an arrow. These arrows are gradients, they show the direction from light pixels to dark pixels throughout the image.

To find a face in the image, you need to identify the part of the image that most closely resembles a known histogram pattern of directional gradients, derived from many other faces during training.

To identify the person in the video, the process of transforming the image into a set of parameters that describe a particular person must be done. Measurements that seem obvious to people (such as eye color) do not make sense to software that looks at individual image pixels. Research has shown that the most accurate approach is to let the computer measure what it needs itself. Deep learning determines which parts of the face need to be measured with greater accuracy.

The solution is to create a convolutional deep learning neural network \[4,5\]. But instead of training the network to recognize objects in the image, it creates 128 measurements for each face.

During training, the network analyzes three faces simultaneously: a training image of a known person's face, another photo of the same known person, and an image of a completely different person. The algorithm then looks at the measurements it makes for each of these three images, tweaking the neural network a bit to make sure that the measurements created for images 1 and 2 are more similar, and the measurements for images 2 and 3 are less similar.

In the last step, all that needs to be done is to find the person in the database of known people whose measurements are closest to those obtained for this purpose the work uses a linear SVM classifier.

4. Developed algorithms

4.1 Extracting keyframes from video footage

A video recording is a sequence of rapidly changing frames. If you decompose it into frames and compare two neighboring frames, you will notice that they do not differ much from each other.

When compressing video the program records only the first frame (I-frame) completely. The next frame (P-frame) is not recorded completely. Instead, it records information referring to the previous frame as well as information about changed areas.

Taking into account the abovementioned it was decided to process only keyframes of the video clip; that would not only increase the percentage of recognized faces but also the speed of video clip processing. Figure 3 shows the algorithm for extracting keyframes from the video footage.
4.2 Searching for all faces and comparing with the faces in the database

Taking into account the considered algorithms, it becomes clear that the database stores the decoded representations of the faces to be found, and with the help of auxiliary software the keyframes are obtained. All these actions are a preparatory stage for the main function of the software - search for important persons of the Volgograd region and the formation of metadata [6].

During the analysis, the frames are transmitted for processing in a batch. The size of the packet is determined depending on the number of keyframes in the video. This approach allows to speed up the analysis process and to make this process even faster if there is hardware acceleration on the server. Figure 4 presents an algorithm for batch search of faces and metadata formation for video footage.

5. Results

The software was implemented in the python programming language. Django framework was used to implement the web application.

The architecture of the software is shown in Figure 5.
Figure 4. Algorithm of batch search of faces and metadata formation

Figure 5. Architecture

The following functions are realized in the software [7,8]:

- authorization on the site;
- division of roles on the user and the administrator (the role is defined at authorization);
- creation of models of significant persons;
- finding persons in the video stream;
- identification of persons found;
- generating a report on the analysis performed;
- creation of a history.

Input data of the software: a completed database with the calculated characteristics of models of persons and history of searches by employees, and a video material in m2v format, which is stored locally on the user's computer.

The output data of the software: the name of the founded person, the time interval where the software detected the person. The output data is written to the application database for later use.

Figure 6 shows the main window of the software.

![Software main window](image)

**Figure 6. Software main window**

6. **Conclusion**

In today's world, information is one of the most important resources. Its large turnover in the media makes companies look for ways to automate processes related to labeling and storage of data. But often these processes are left on the shoulders of archive workers who have to review the footage in real-time and create metadata based on what they see, which will be transferred into the archive. This process is often very time-consuming.

That is why it was necessary to develop software package for collecting, storing, and analyzing the data that would allow using modern technologies to perform metadata formation in parallel with the other work of the archivist.

As a result of the analysis of approaches to video data processing, comparative analysis of the efficiency and availability of computer vision services, a software for searching faces, and forming metadata in the video archive of the VGTRK company was designed and implemented.

The application area of the developed software is the automation of the archivist's work when describing video files for preservation in the archive.

The main direction of improvement and further development of the developed software is to add the function of recognizing other objects that require description, such as monuments, famous places, etc., and also to introduce the hardware acceleration of video file analysis by paralleling the frame processing process within the stream.

7. **Acknowledgments**

The reported study was funded by RFBR and the Administration of the Volgograd region according to the research projects 19-47-340007, 19-41-340016.

**References**

[1] Waldemar W, Konrad G and Muhtar J 2016. *Face Recognition: Issues, Methods and Alternative Applications*. [Online]. Available: https://www.intechopen.com/books/face-recognition-semisupervised-classification-subspace-projection-and-evaluation-methods/face-recognition-issues-methods-and-alternative-applications
[2] Tolokin D S, Korobkin D M, Fomenkov S A, Kolesnikov S G 2021 The analysis of the graphical representation of mathematical formulas for extending descriptions of physical effects Journal of Physics: Conference Series 1801 012016

[3] Allibhaic E Building a Convolutional Neural Network in Keras. [Online]. Available: https://towardsdatascience.com/building-a-convolutional-neural-network-enn-in-keras-329fbbadc5f5

[4] Dwivedi D Face Recognition for Beginners. Towards data science. [Online]. Available: https://towardsdatascience.com/face-recognition-for-beginners-a7a9bd5eb5c2

[5] Facial recognition technology, Tadveser. 2021. [Online]. Available: https://www.tadviser.ru/index.php(Facial_recognition)#:~:text=Facial%20recognition%20technology

[6] Kochura O A, Korobkin D M, Fomenkov S A and Kolesnikov S G 2021 Development of the patent array analysis module based on the "problem-Solution" model Journal of Physics: Conference Series 1801 012014

[7] OpenCV, Opencv.org, 2020. [Online]. Available: https://opencv.org/

[8] TensorFlow, Tensorflow.org, 2021.[Online]. Available: https://tensorflow.org/