DEMO: Smart Assistance Video Monitoring Services in Industrial Systems

Nikita Bazhenov, Sergey Marchenkov, Artur Harkovchuk, Vsevolod Averkov, Dmitry Korzun
Petrozavodsk State University (PetrSU)
Petrozavodsk, Russia
{bazhenov, marchenkov, harkovchuk, averkov}@cs.petrsu.ru
, dkorzun@cs.karelia.ru

***Abstract*—The demo is part of a system for multi-parameter monitoring of production machinery (technical state, operating conditions, personnel actions). We show a pilot implementation of the smart assistance video monitoring services. They are used in manufacturing processes based on simultaneous video data analysis from multiple video cameras. Surveillance cameras observe the workplace of production machinery. Mechanical components (machinery units) are analyzed to detect deviations in normal state in machine operations. Operator presence is analyzed in the area to control production processes. Text from CNC display monitor is recognized to detect displayed error messages and their sequences.**

Services implementation

Video cameras located in the surveillance area provide an overview of objects at the enterprise. Moreover, this survey can be organized both for living and non-living objects. In our case, by living objects, we mean operators, recognizing their silhouettes, faces, equipment and determining the distance. Non-living objects are understood as machinery presented in the form of machine tools for performing certain industrial tasks.

As an example, Fig. 1 shows service is used to recognize people and calculate the distance to a person and equipment. The first video camera, installed on the top of the engine room, contains a video stream with machine operators. The second camera, which is installed directly opposite the equipment (approximately at the human level), contains a video stream with multiple operators and equipment (the difference between the cameras is in their presentation in three-dimensional space). First, a connection to the database is initialized, from where the current information about the presence of personnel on the site, as well as the distance to the equipment, is taken. Cameras are connected to a router and transmit information to video processing modules. Video analytics modules (human recognition and distance calculation module) determine the presence of a person in the frame and also determine the distance to him by the person's silhouette. Then events are generated using the monitor and sent to the MongoDB database, as well as a message broker the RabbitMQ. Further, the information is converted into information necessary for users (construction of graphs and diagrams) and is displayed in video services on the network on end devices.

Services demonstration based on main tasks of video monitoring in industrial systems:

- Monitoring mechanical components to detect deviations in machine operations;
- Operator presence in the area to control production processes;
- Screen image text analysis from CNC display monitor to detect errors.

Operator recognition and distance calculation

This video service is an improvement over our previous demo. Operator recognition is represented by the YOLOv3 neural network [1]. Video streams from multiple cameras are processed in real time. In the first video stream, the search for people in the frame is organized. The service can count their number, coordinates (in the form of a rectangle), as well as determine the distance to a human silhouette.

Since the processing power is quite limited due to the simultaneous use of multiple services in real time, the YOLOv3 model uses the configuration file and the lightweight yolo-tiny weights. For the greatest optimization, only 1 frame per second is processed, with an image size of .800 pixels.

Further, the service recognizes a machine object in the image, determining its distance in automatic mode (or taking initial data from a database), determines the distance to a person and concludes about the proximity of a person and a machine. In the event that the operator is within the boundaries of the equipment, this probably means that he interacts with him in some way.

The operator recognition service determines the presence of a person on the time axis, storing the image of the operator's entry / exit from the frame, as well as the video with his presence. It is possible to render and save video in the user interface.

Helmet recognition on the operator's head

The YOLO model (version 3) was used and trained [1] to determine the presence of a safety helmet. A dataset was assembled and marked on the photographs of which helmets are shown. The markup tool is LabelImg.

YOLO has 24 convolutional layers followed by 2 fully connected layers. Some convolutional layers use 1×1 shrink layers as an alternative to reduce the depth of feature maps. For the last convolutional layer, it outputs a tensor with the shape (7, 7, 1024). Then the tensor is aligned. Using 2 fully

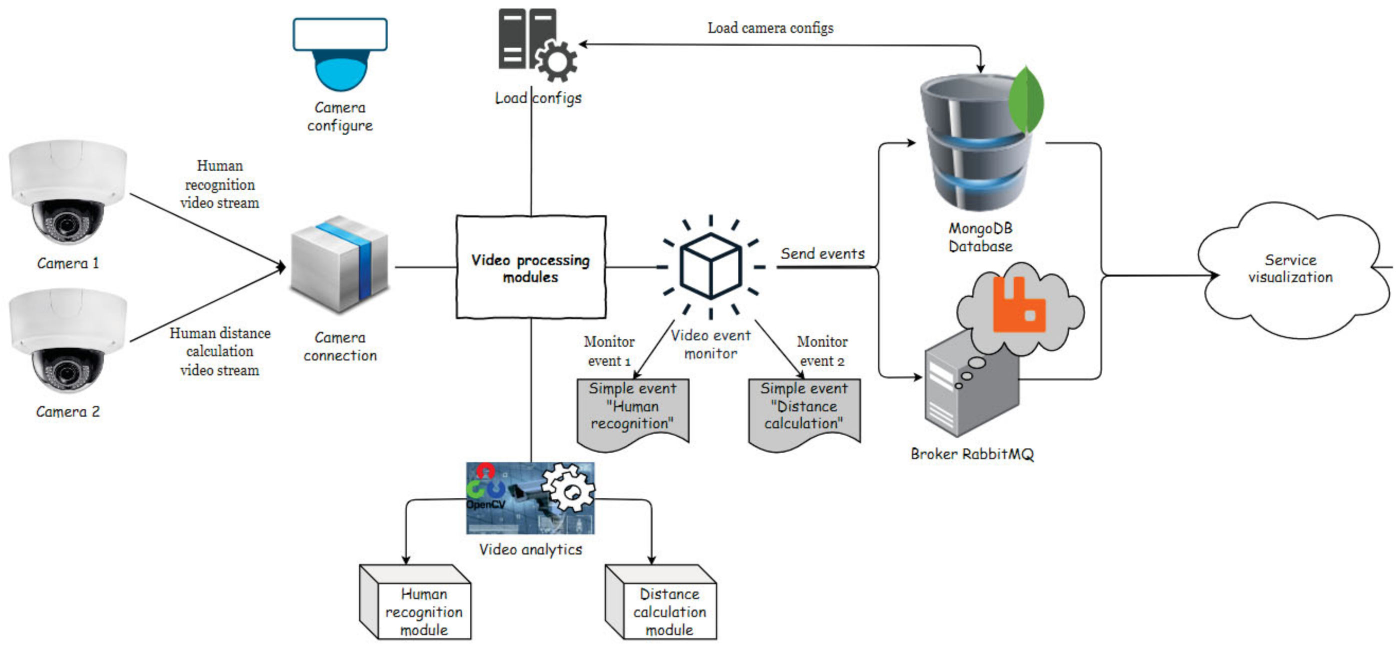

Fig. 1. Video Event Representation Model

connected layers as the linear regression form, it outputs $7 \times 7 \times 30$ parameters and then converts them to (7, 7, 30), that is, 2 prediction bounding boxes for each location. A faster but less accurate version of YOLO called Fast YOLO uses only 9 convolutional layers with less functional maps.

The result of detecting or not detecting a helmet will be the result of the operation of 2 services, one detects a person and the other searches for a helmet on the detected person, if the second service does not find the object we need in the area of human detection, then we will assume that he is without a protective helmet.

Error recognition on the CNC monitor

The service for recognizing an error code from the CNC machine screen allows to receive CNC error codes, analyze and provide information about errors [2]. The use allows receiving data from the screen without using a direct connection to the CNC. To receive video, a webcam with autofocus is used, stirred with a bracket on the side of the CNC screen.

The service finds the area on the monitor screen in which the error code is published. The search is carried out by the ”findContours” method provided by the OpenCV library, finding a closed polyline consisting of four vertices. The broken line found is confirmed by the color scheme using the ”moments” method in the OpenCV library, which should be red. The region found is converted using affine transformations, filtered, and the Tesseract tool [3] is used to recognize the numeric error code.

The user is presented with an error code and description for an accurate understanding of the error type and remedies. For accurate analysis, the user is presented with five common errors and five errors that are on the screen for the longest time. The user is also provided with a timeline, which indicates when errors were published on the screen.

As part of the demonstration we used the following technical devices and records:

- 2 IP-cameras with different focal length: Hikvision DS-2CD2123G0-IS (4 mm), Hikvision DS-2CD2143G0-IS (2.8 mm) for pre-recordings;
- 1 Web-camera Logitech C615 with bracket attachment;
- Video processing with OpenCV face detection [4] using Haar cascade filter [5] and algorithms to calculate distance to the face of the operator;
- Pre-trained neural network based on TensorFlow [6] and YOLO [1] for recognizing a person by its silhouette and its helmet;
- Web technologies: php, laravel (backend), gentelella, bootstrap (frontend), nodejs server based on socket.io and express;
- Database MongoDB for monitoring events;
- Message protocol MQTT for message exchange;
- Python 3.8 programming language with libraries for the implementation of the basic video processing modules and interaction with above technologies.

The service is part of the Opti-Repair monitoring system developed in Petrozavodsk State university (petrsu.ru) and Opti-Soft company (www.opti-soft.ru).

Acknowledgment

This research is implemented in Petrozavodsk State University (PetrSU) with financial support by the Ministry of Science and Higher Education of Russia within Agreement no. 075-11-2019-088 of 20.12.2019 on the topic "Creating the high-tech production of mobile microprocessor computing modules based on SiP and PoP technology for smart data collection, mining, and interaction with surrounding sources".

References

[1] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv*, 2018.

[2] Z. Yang, D. Zhu, H. Tian, J. Guo, and S. Li, "Reliability modelling of cnc machine tools based on the improved maximum likelihood estimation method," *Mathematical Problems in Engineering*, vol. 2018, pp. 1–11, 05 2018.

[3] R. Smith, "An overview of the tesseract ocr engine," in *Ninth international conference on document analysis and recognition (ICDAR 2007)*, vol. 2. IEEE, 2007, pp. 629–633.

[4] K. Pulli, A. Baksheev, K. Kornyakov, and V. Eruhimov, "Real-time computer vision with opencv," *Communications of the ACM*, vol. 55, no. 6, pp. 61–69, 2012.

[5] S. Soo, "Object detection using haar-cascade classifier," *Institute of Computer Science, University of Tartu*, pp. 1–12, 2014.

[6] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.