Evaluation of the Human Use for Sport Training Equipment based on Multicamera Video Surveillance

Nikita Bazhenov, Egor Rybin, Sergey Zavyalov, Dmitry Korzun
Petrozavodsk State University (PetrSU)
Petrozavodsk, Russia
{bazhenov, rybin}@cs.petrsu.ru,
sza123@list.ru, dkorzun@cs.karelia.ru

***Abstract*—The progress in technologies of Internet of Things (IoT) and Artificial Intelligence (AI) lead to the digitalization in various domains of human activity. In this demo, we show a digital inspector that recognizes the physical human activity with sports training equipment (mechanical simulators). We experiment with outdoor simulators produced by the MB Barbell company and located on the shore of Lake Onega in the Petrozavodsk city for free use by citizens and guests. We implemented multi-camera video surveillance system (VSS) that statistically evaluates the states of the sports equipment use such as "the simulator is occupied by a person", "the simulator is not correctly used", "the simulator is occupied but not in use", etc. The evaluation is based on well-known AI methods for image recognition, which are customized for real-time inspection. The evaluated statistics for the area with sports equipment is important for the municipality to understand the efficiency of the provided public resources for well-being and comfort life.**

I. Introduction

Sports are an important and integral part of any person who monitors their health. There are technologies of mobile medicine (mHealth [1]) and Ambient Intelligence (AmI [2]), which allow using certain electronic devices and mobile gadgets to organize autonomous monitoring of human health using sensors and video devices (CCTV). Such technologies can provide an ability to monitor patients and athletes (all those who need permanent or temporary health monitoring) in order to detect various symptoms or vital signs, detect anomalies and provide timely interventions.

Computer vision (CV) and machine processing (ML) technologies are essential for this task and can be used to solve this problem. For example, computer vision algorithms can be used to track patient movements, detect facial expressions, and monitor vital signs such as heart rate and the general condition of the athlete during the exercise. Such algorithms can be used to identify patterns in video data coming from multiple CCTV-source and predict potential health issues before they occur or make recommendations to stop current exercise.

In this work in progress, we consider a multi-camera approach for recognizing the physical activity of people on multiple simulators performing various exercises depending on the simulator. In our implemented prototype, recognition of the following states is organized:

- The algorithm simultaneously receives 3 images from different cameras, each image contains several recognizable simulators, and the same simulators can be presented at different angles;
- The received images are synchronized in time;
- There is a connection between each recognized simulator: simulator - camera weight;
- Real simulators of the embankment of the Republic of Karelia were used to recognize;
- Recognizable states:
  - Training machine is free, human either is not presented in a frame or stand away;
  - Human stands in the area near the training machine;
  - Human is in the working area of training machine, but do not necessarily do anything with them;
  - Human is actually working out, training machine is in use.

For each situation, the sum duration is measured, resulting in "a usage map" of the training machine. In our experiments, the training machine is provided by MB Barbell™ http://www.mbbarbell.com/ as well as the initial practical problem of training machine usage reporting.

The key scientific contribution of our study is an experimental extension of our previous works [3] should show that such approach [4] can be extended to several video cameras that monitor the condition of athletes from several simulators at the same time using existing algorithms and technologies in human activity recognition [5] and work in real-life conditions.

The rest of the paper is organized as follows. Section II introduces the algorithm of the service and its experimental setup. Section III demonstrates the mathematical algorithm for recognizing multiple persons in the training machine area. Section IV shows the use of existing technologies in the implemented prototype. Section V summarizes this work-in-progress R&D study.

II. Service Algorithm

The characteristics of our experimental setup are:

- IP-camera Hikvision (2.8, 4.0 mm) for pre-recordings/recordings in RT x3;
- Cameras are installed at 5-6m;

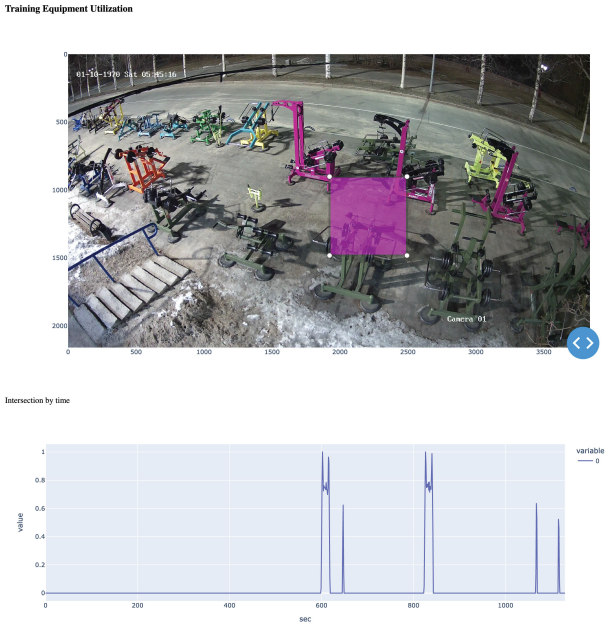

Fig. 1. The purple box shows chosen area, and the graph below shows the time when some activity was presented there, a person walked up to the training machine and did some exercises for a couple of seconds

- Human recognition occurs using the YoloV5 and YoloV8 neural networks;
- Each camera has its own zones for recognizing people on simulators;
- On most cameras there are 6 simulators with their own region on interest (ROI) for recognition;
- There is a list of coordinates of people and confidence in binding to a specific point in time and camera;
- The person's coordinates are compared with the zone coordinates and the intersection is calculated;
- A conclusion is made about the presence of a person in the area of the simulator or interaction with him.

The server for data processing has the following specification.

- Intel Core i5-9400F 2.9GHz.
- Nvidia RTX 2080 8GB.
- 32 GB RAM.

III. Approach

The main script runs a neural network model based on YoLoV5(8). As a result, the model returns the coordinates of the rectangle: x_1, y_1, x_2, y_2 Graphics are shown in Fig 1.

At first bounding boxes for training equipment must be specified.

To determine if the training machine is free or not (if a person is presented near it), the intersection between the person and training equipment bounding boxes is calculated. The area of intersection over a person's bounding box should exceed some threshold value for it to be considered that person is near equipment.

$$\text{Intersection: } \frac{|\text{Detected Box} \cap \text{Chosen Area}|}{|\text{Detected Box}|}$$

Such an approach is error-prone to people walking in the background far behind the training equipment because their bounding box would be small and completely included in the area of training equipment, which could happen when the camera angle is low. In this situation, information from multiple cameras can be combined to confirm a person's location.

To detect if a person presented in the area combined information used from confidence(DC) of the detection, intersection(Intersection, IoB) with training equipment, camera weight(CW).

If the majority (at least 2 out of 3 cameras) agrees that the person presented then it gets counted, otherwise ignored.

$$\sum_{i=0}^{\text{num of cameras}} CW[i] \cdot DC[i] \cdot (IoB[i] > threshold)$$

Most frequent location (average bounding boxes) (Light green boxes in Fig. 2) calculated as average coordinates for each bounding box for people that intersect with chosen area for at least some part (threshold value). This can be used to determine if a person standing near the training machine or inside the "working area".

$$\frac{1}{\text{num of records}} \cdot \sum_{i=0}^{\text{num of records}} (x_{1_i}, y_{1_i}, x_{2_i}, y_{2_i})$$

To determine if training equipment is actually used we try to detect the movement of the training equipment itself, not just the person. This could be done by specifying the region on the image where some parts of the training equipment should be moving while it in use and not moving by themself otherwise. Then it is possible to detect such movement by comparing the current and previous images.

IV. Software Implementation

The following existing technologies were used:

- PyTorch;
- YoLov5(8);
- OpenCV;
- FFmpeg;
- GPU-based recognition.

Let us summarize the existing technologies and underlying recognition algorithms.

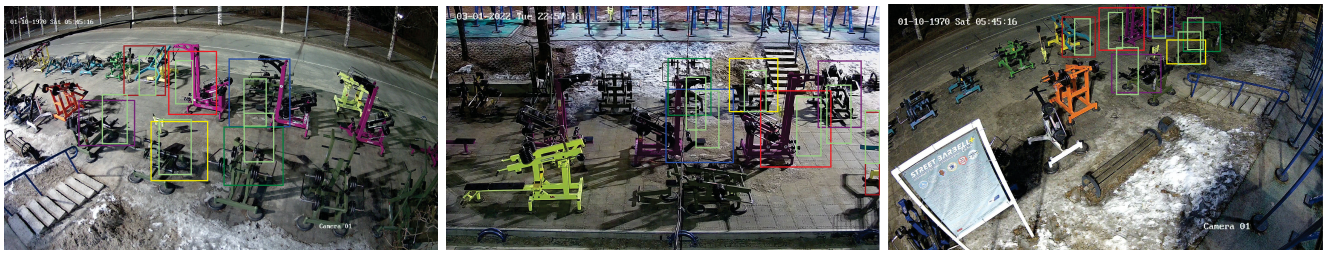

Fig. 2. Example of images from each camera feed, with zones selected. Larger Red, Blue, Green, Yellow, Purple, and Brown squares are manually chosen areas of 6 training machines, with colors synced between images. Smaller Light green squares inside are zones for each box where humans are most likely to be (calculated as average coordinates of detected humans)(working area of the machine, presumably)

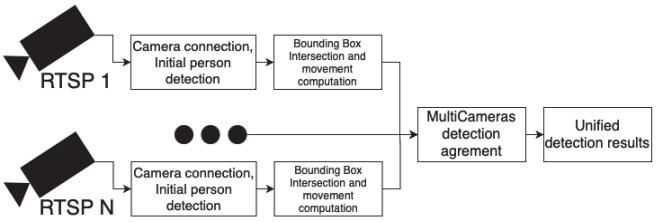

Fig. 3. Modules of developed system and data flow with multiple cameras connected to services

TABLE I. PROTOTYPE RESULTS USING 1 FULL DAY

Machine	Not present (hours)	Present (minutes)	Taken/Used (seconds)
1 (red)	20.66	200.3	1344
2 (blue)	22.88	67.3	560
3 (green)	23.89	6.6	56
4 (yellow)	23.51	29.4	261
5 (purple)	23.92	5.1	5
6 (brown)	22.69	78.3	1143

- IP-cameras for pre-recordings and recordings in real-time:
  - Hikvision DS-2CD2T83G2-4I 2.8mm, 4K/30fps;
  - Hikvision DS-2CD2143G2-IU 2.8mm, 1080p/25fps;
  - Hikvision DS-DE4225IW-DE(S5), 1520p/30fps;
- Pre-trained neural network based on PyTorch and YOLOv5(8) [6] for recognizing a person by its silhouette.
- Web technologies: flask [7], OpenCV [8], and OvenMedia [9].
- Message protocol ZeroMQ [10] for message exchange.
- Python 3.10 [11] programming language with libraries for the implementation of the basic video processing modules and interaction with above technologies.
- Julia [12] programming language for implementation of areas recognition.

In Table I results from processing a video lasting 1 day (weekday) are presented. It can be seen that the training machines with red (1344 seconds) and brown (1143 seconds) colors were the most popular. In particular, the greatest presence of people was noticed on the red simulator (200.3 minutes). This was probably due to the fact that people passed through this area. On some machines, it was difficult to describe if a person had taken the machine or had already started using it. Moreover, in some camera areas, one simulator overlapped the other, which made it even more difficult to organize proper recognition.

V. CONCLUSION

The implemented prototype can estimate statistics on using multiple training machines by a person using a multi-camera video surveillance system (VSS) using well-known AI, CV, ML algorithms and methods. The next following 4 situations were implemented using 3 cameras and 6 training machines:

1) The training machine is free.
2) A person is in the area near the training machine.
3) The training machine is occupied but not in use.
4) The training machine is used by a person.

The list of accounted situations (variants of using a training machine) is subject to extension in the next versions of our prototype.

ACKNOWLEDGMENT

The implementation of this demo is supported by MB Barbell™(http://www.mbbarbell.com/). The scientific results of this research study are supported by Russian Science Foundation, project no. 22-11-20040 (https://rscf.ru/en/project/22-11-20040/) jointly with Republic of Karelia and Venture Investment Fund of Republic of Karelia (VIF RK). The work is in collaboration with the Artificial Intelligence Center of PetrSU.

REFERENCES

[1] S. Almotiri, M. Khan, and M. Alghamdi, "Mobile health (m-health) system in the context of iot," 08 2016, pp. 39–42.

[2] H. Sabit, P. H. Joo Chong, and J. Kilby, "Ambient intelligence for smart home using the internet of things," in *2019 29th International Telecommunication Networks and Applications Conference (ITNAC)*, 2019, pp. 1–3.

[3] N. Bazhenov, E. Rybin, and D. Korzun, "An event-driven approach to the recognition problem in video surveillance system development," in *2022 32nd Conference of Open Innovations Association (FRUCT)*, 2022, pp. 65–74.

[4] N. Bazhenov, E. Rybin, S. Zavyalov, and D. Korzun, "Human activity recognition for sport training machines," in *2022 32nd Conference of Open Innovations Association (FRUCT)*, 2022, pp. 328–331.

[5] J. A. Patiño-Saucedo, P. P. Ariza-Colpas, S. Butt-Aziz, M. A. Piñeres-Melo, J. L. López-Ruiz, R. C. Morales-Ortega, and E. De-la-hoz Franco, "Predictive model for human activity recognition based on machine learning and feature selection techniques," *International Journal of Environmental Research and Public Health*, vol. 19, no. 19, 2022. [Online]. Available: https://www.mdpi.com/1660-4601/19/19/12272

[6] G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, NanoCode012, Y. Kwon, TaoXie, J. Fang, Imyhxy, K. Michael, Lorna, A. V, D. Montes, J. Nadar, Laughing, Tkianai, YxNONG, P. Skalski, Z. Wang, A. Hogan, C. Fati, L. Mammana, AlexWang1900, D. Patel, D. Yiwei, F. You, J. Hajek, L. Diaconu, and M. T. Minh, "ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference," Feb. 2022. [Online]. Available: https://zenodo.org/record/6222936

[7] "Flask Dcoumentation." [Online]. Available: https://flask.palletsprojects.com/

[8] "Home - OpenCV." [Online]. Available: https://opencv.org/

[9] "OvenMediaEngine | Open-Source Projects." [Online]. Available: https://www.ovenmediaengine.com/

[10] "ZeroMQ." [Online]. Available: https://zeromq.org/

[11] "Home - Python." [Online]. Available: https://www.python.org/

[12] "Home - Julia." [Online]. Available: https://julialang.org/