# Human Activity Recognition for Sport Training Machines

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*—Modern video surveillance systems (VSS) can recognize a person and her/his activity. A particular example of such a VSS system is the monitoring and support of human health and wellbeing during exercise with sports equipment. In this work-in-progress paper, a prototype of the VSS system is demonstrated for monitoring the movement of a person around the sport training machine. For the recognition problem we consider the following situations: a) training machine is free (no person is close to the machine); b) a person is in the area near the training machine (but not in the working area); c) a person is in the working area of training machine (but the training machine is not used); d) a person is working on the training machine. Our early experimental study evaluates the feasibility of such a VSS system.**

## I. Introduction

Development of Video Surveillance Systems (VSS) is progressing towards Ambient Intelligence (AmI) services in Internet of Things (IoT) environments [1]. An interesting application domain for VSS is mobile healthcare (mHealth), well-being, sport, and various at-home situations [2]. A person is tracked in real-time with provision of a services as a sequence of recognized events and recommendations. Human activity recognition also supports digital services for Ambient Assisted Living (AAL), especially important for elderlies and patients with chronic diseases.

In this work-in-progress paper, we consider human physical activity recognition in respect to sport exercises with a training machine. The service collects statistics on using the training machine. On every video frame a person is recognized (as human). The following situations are implemented in our early prototype.

1) The training machine is free, i.e., no person is recognized in the frame or a recognized person stands away.
2) A person stands in the area near the training machine.
3) A person is in the working area of the training machine, but the training machine is not in use.
4) A person is actually working out, the 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 experimental proof that such a VSS service can be developed with existing recognition algorithms and technologies. The implemented recognition of movement (human or machine) in video data can achieve satisfactory results in real-life settings.

The rest of the paper is organized as follows. Section II introduces the service and experimental setup. Section III demonstrates the algorithm for recognizing a person in training machine area. Section IV demonstrates the algorithm for recognizing human posture. Section V demonstrates the algorithm for recognizing training machine pitching. Section VI discusses the use of existing technologies in the implemented prototype. Section VII summarizes this work-in-progress R&D study.

## II. Service for Reporting on Training Machine Usage

Our basic scenario for the service is shown in Fig. 1. The imoplementation is based on the following three modules.

- Recognition of a person in the area of the training machine.
- Human posture recognition;
- Recognition of the training machine pitching.

The modules implement existing recognition algorithms based on neural networks. The following outputs are produced.

- Training machine is free.
- Human stands away.
- Human is on the bench, preparing for the exercise.
- Human is working out.
- Human is working out, started exercise.
- Human is working out, exercise is in progress.
- Human is working out, +1 repetition.

Our experimental setup is shown in Fig. 2. The characteristics are as follows.

- A single IP-camera Hikvision DS-2CD2143G0-IS (2.8 mm) for pre-recordings/recordings in RT;
- Camera is installed at 2.4 m.
- Training equipment is training machine (bench press).

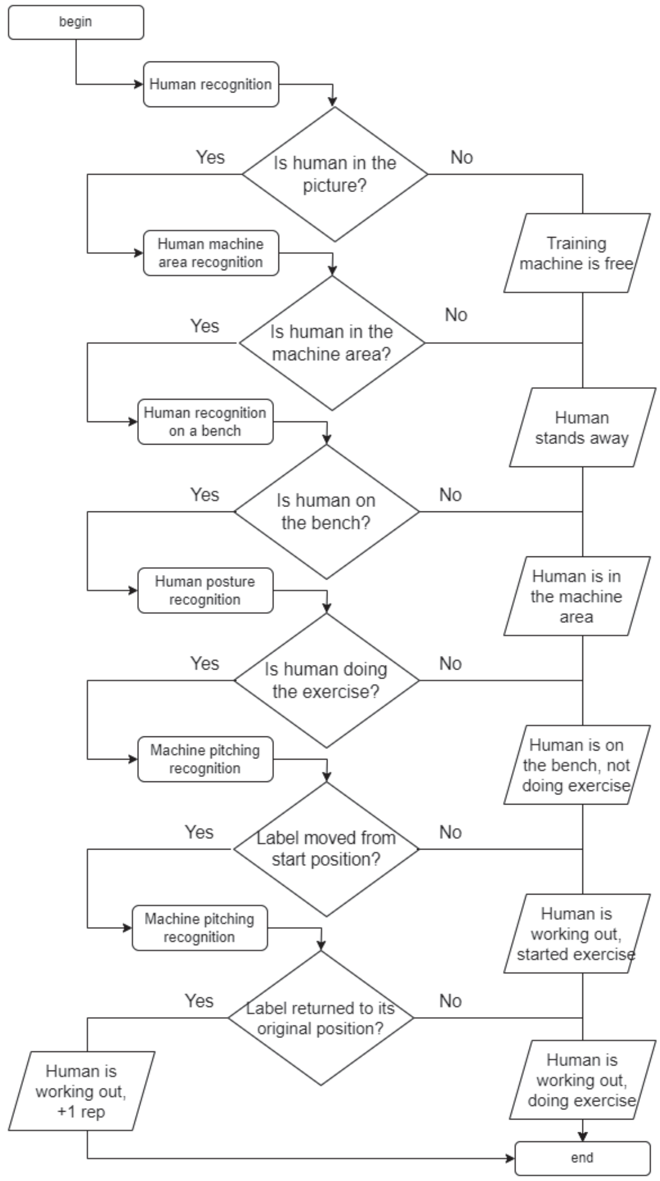


Fig. 1. Service scenario


Fig. 2. Experimental setup

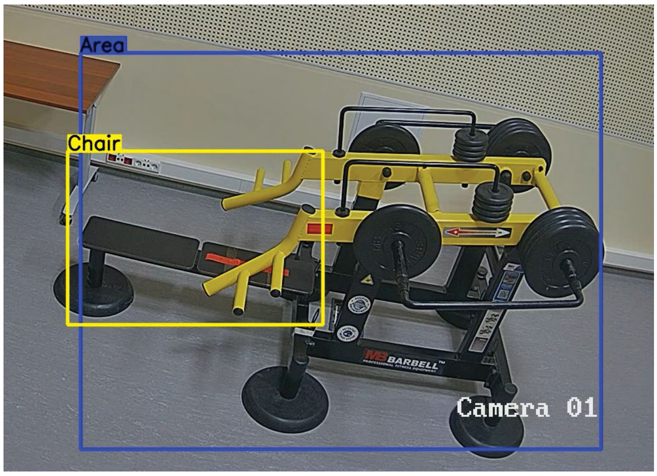


Fig. 3. Running the algorithm for recognizing a person in training machine area

The server for data processing has the following specification.

- Intel Core i9-9900K.
- Nvidia RTX 2060.
- 32 GB RAM.

Video stream is captured from the camera using FFmpeg command and forwarded to services via ZeroMQ library.

## III. Recognizing a Person in Training Machine Area

Results of running the recognition algorithm is demonstrated in Fig. 3. The following existing technologies are used.

- PyTorch;
- YoLov5;
- OpenCV;
- FFmpeg;
- GPU-based recognition.

The main script runs a neural network model based on YoLov5 [3]. As a result, the model returns the coordinates of the rectangle: $x_1$, $y_1$, $x_2$, $y_2$ (upper left and lower right coordinates) with a person recognized in the frame. The coordinates of the location of the simulator and the bench can be determined automatically using a neural network, but this method is not always accurate, since the simulator is not always determined if a part of the simulator is blocked by a person or other foreign objects.

Next, the coordinates of the person are compared with the area of the simulator (the intersection of rectangles). If a person is in the simulator area, the coordinates of the person are compared with the bench area. If the area of intersection of the bench and the person (the area of intersection of the rectangles) exceeds a certain threshold (for example, 50%), it is considered that the person is on the bench. In addition, there are additional checks in the algorithm. For example, the average height of a


Fig. 4. Running the algorithm for recognizing human posture


Fig. 5. Running the algorithm for recognizing training machine pitching

person is taken into account. If the height is large, then the person is very close to the camera and “overlaps” the bench area, and is not directly on it.

The neural network YoLov5 is a medium sized network with 640 input. It is sufficient for the task, considering additional information that come from human posture recognition and training machine pitching recognition.

The neural network detects person and returns bounding box as result. For this bounding box interception with predefined zones calculated. If interception exceed threshold value, then we consider person to be in this zone.

In the example image (Fig. 3), blue box indicates the area around training-equipment and yellow box indicates working area where person supposes to be while working with the training machine.

## IV. Recognizing Human Posture

Results of running the recognition algorithm is demonstrated in Fig. 4. The following existing technologies are used.

- MediaPipe (BlazePose);
- 12 key points;
- 264 features;
- k-NN algorithm.

Human posture recognition uses BlazePose [4], which is a neural network from the MediaPipe library. The neural network detects 32 keypoints on person, but for our posture recognition methods only 12 of them needed: wrist, elbow, shoulder, hip, knee, ankle (from both sides of person) as shown in example in Fig. 4.

Then keypoints are processed to produce 264-feature vector by taking difference in x and y coordinates between every keypoint, i.e., evaluating $|x_i - x_j|$ and $|y_i - y_j|$ for each $i, j \in 0, \ldots, 11$ and $i \neq j$.

Templates for specific postures recorded before method started. This could be either video samples or set of images. Template feature vectors computed for templates in the same way as for images obtained in runtime.

Vectors obtained from people during runtime compared with templates with $k$-Nearest Neighbour algorithm. So closest match will indicate that person is in corresponding posture.

In case of human activity recognition for Sport Training-machine two postures was chosen: human is standing or walking somewhere around sport training-machine and human is working with training-machine. In Fig. 4, the person conduct standing around training-machine.

## V. Recognizing Training Machine Pitching

Results of running the recognition algorithm is demonstrated in Fig. 5. Reflective red tape is used on the arm of training-machine for recognition. The following existing technologies are used.

- Machine pitching recognition with a pre-installed label.
- Red color saturated pixel value search.

During runtime image that we get from camera transformed from RGB format to HSV, because color transformation is easier in HSV. Mask created choosing color in Hue-Saturation-Value format between 0–70–70 and 3–255–255 values, thus separating red color from image.

Position of the mark detected as place with highest amount of red color on image. For this window goes though mask on x and y axis and calculate coordinates where window gets highest value. The result is presented in Fig. 5 where green dot represents detected coordinates of the marker.

If mark moved further away from any one of last 60 frames than some threshold value, we assume it was due to arm movement that happens when one operates with training-machine normally.

When mark moves away from the start position and return back later, we assume that this was due to repetition as somebody was using the training-machine, and can count plus one repetition.

## VI. Use of Existing Technologies

Let us summarize the existing technologies and underlying recognition algorithms.

- IP-camera Hikvision DS-2CD2143G0-IS (2.8 mm) for pre-recordings/recordings in RT.
- Pre-trained neural network based on PyTorch [5] and YOLOv5 [3] for recognizing a person by its silhouette.
- Web technologies: flask, OpenCV [6], and OvenMedia [7].
- Message protocol ZeroMQ [8] for message exchange.
- Python 3.8 programming language with libraries for the implementation of the basic video processing modules and interaction with above technologies.

Our experiments show that the technologies (and implemented algorithms) provide satisfactory accuracy and performance for recognition of human activity. The statistical metrics on the use of a training machine are evaluated, so solving the problem of training machine usage reporting.

## VII. Conclusion

The presented prototype is a demo that shows a possible solution to recognition of situation of human-machine interaction. The demo service can estimate statistics on using the training machine by a person. The following situations are implemented in our early prototype.

1) The training machine is free.
2) A person stands in the area near the training machine
3) A person is in the working area of the training machine.
4) A person is working out, so using the training machine.

The list of accounted situations (variants of using a training machine) is subject to extension in the next versions of the service. We showed that the existing recognition algorithms and technologies can be used in development of such VSS services in practice.

## 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] D. Korzun, E. Balandina, A. Kashevnik, S. Balandin, and F. Viola, *Ambient Intelligence Services in IoT Environments: Emerging Research and Opportunities*. IGI Global, 2019.

[2] A. Y. Meigal, D. G. Korzun, L. I. Gerasimova-Meigal, A. V. Borodin, and Y. V. Zavyalova, "Ambient intelligence at-home laboratory for human everyday life support," *International Journal of Embedded and Real-Time Communication Systems (IJERTCS)*, vol. 10, no. 2, 2019, in Press.

[3] 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

[4] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "BlazePose: On-device Real-time Body Pose tracking," *arXiv:2006.10204 [cs]*, Jun. 2020, arXiv: 2006.10204. [Online]. Available: http://arxiv.org/abs/2006.10204

[5] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," 2017, tutorial. [Online]. Available: https://pytorch.org/tutorials/beginner/basics/autogradqs_tutorial.html

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

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

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