

## Design of smart human following on rail inspection using human pose estimation marker-less motion capture based on blazepose

### Rancang Bangun Smart Human Following Pada Inspeksi Rel Kereta Api Menggunakan Human Pose Estimation Marker-Less Motion Capture Berbasis BlazePose

Adiratna Ciptaningrum<sup>1</sup>, Wahyu Pribadi<sup>2</sup>, Dirvi Eko Juliando<sup>3</sup>, Rakhmad Gusta<sup>4</sup>, Edo Zulmi Faikhsan<sup>5</sup>

<sup>1,2,3,4,5</sup> Department of Engineering, Madiun State Polytechnic, Indonesia

<sup>1</sup>[adiratna@pnm.ac.id](mailto:adiratna@pnm.ac.id), <sup>2</sup>[why.pribadi@pnm.ac.id](mailto:why.pribadi@pnm.ac.id), <sup>3</sup>[dirvi@pnm.ac.id](mailto:dirvi@pnm.ac.id),

<sup>4</sup>[rakhmad.gusta@gmail.com](mailto:rakhmad.gusta@gmail.com), <sup>5</sup>[edozulmifaikhsan@gmail.com](mailto:edozulmifaikhsan@gmail.com)

#### ABSTRACT

Advances in artificial intelligence (AI) technology today have a significant impact in various aspects of human life. One example is the evolution of robotics that has achieved the ability to follow human movements. To achieve this, AI technology utilizes image recognition through Computer Vision and the Human Pose Estimation method with the help of the BlazePose library, which is able to recognize 33 keypoints in human body poses. Research in this area aims to develop an automatic control system that can be used on inspection carts, enabling them to follow human body movements while walking. The results showed a detection accuracy rate of 84.82% with an optimal detection distance between 4 to 8 meters from the camera, with an average detection accuracy of 89.862%. On the motor control aspect, the system is set to turn off the motor when the distance between the device and the object is in the range of 1-2 meters, and turn it on at a distance of 3-12 meters. However, it is important to note that the accuracy achieved is greatly affected by the color segmentation capabilities of the software, the lighting conditions in the environment, as well as the resolution of the camera used.

Keywords: inspection train; computer vision; human pose estimation

#### ABSTRAK

Kemajuan teknologi kecerdasan buatan (AI) saat ini memiliki dampak yang signifikan dalam berbagai aspek kehidupan manusia. Salah satu contohnya adalah evolusi robotika yang telah mencapai kemampuan untuk mengikuti gerakan manusia. Untuk mencapai ini, teknologi AI memanfaatkan *image recognition* melalui *Computer Vision* dan metode *Human Pose Estimation* dengan bantuan *BlazePose library*, yang mampu mengenali 33 *keypoints* dalam pose tubuh manusia. Penelitian dalam bidang ini bertujuan untuk mengembangkan sistem kontrol otomatis yang dapat digunakan pada kereta inspeksi, memungkinkan alat tersebut untuk mengikuti manusia saat berjalan. Hasil penelitian menunjukkan tingkat akurasi deteksi sebesar 84,82% dengan jarak optimal deteksi antara 4 hingga 8 meter dari kamera, dengan rata-rata akurasi pendeteksian mencapai 89,862%. Pada aspek pengendalian motor, sistem ini diatur untuk mematikan motor saat jarak antara alat dan objek berada pada kisaran 1-2 meter, dan menghidupkannya pada jarak 3-12 meter. Namun, penting untuk mencatat bahwa akurasi yang dicapai sangat dipengaruhi oleh kemampuan segmentasi warna dari perangkat lunak, kondisi pencahayaan di lingkungan, serta resolusi kamera yang digunakan.

Kata kunci: kereta inspeksi; visi komputer; human pose estimation.

## 1. INTRODUCTION

Motion capture has been a topic of interest in several studies. One of the applications of motion capture techniques is in human pose estimation. Before many studies adopt the marker-less approach for human pose estimation, a number of studies still used markers to generate human skeletons.

Optical motion is one of the tools used to generate human skeleton structures, with motion capture results that have a higher level of accuracy. However, optical motion has limitations in terms of practically, as it requires the intended user to wear clothing that has markers. Lately, an increasing number of researchers have been focusing on this trending subject.

In addition to optical motion, some studies have also used Kinect devices. However, sometimes in its use, the Kinect device faces challenges in obtaining the human skeleton structure (Liu et al., 2016). This arises do to self-occlusions (concealment of joints by other body parts) and also errors that may come from the Kinect sensor itself, because Kinect devices are basically intended for consumers, so the level of accuracy and reliability tends to be lower.

Lately, an increasing number of researchers have been focusing on this trending subject of marker-less motion capture, from an algorithmic standpoint. Marker-less motion capture can be divided into two primary categories, discriminative approaches and generative approaches (Chao, 2016).

Discriminative approaches utilize data-driven machine learning techniques to transform the motion capture challenge into a regression or pose classification task (Hong, 2016), making them suitable for applications involving human-computer interaction where efficiency outweighs precision. On the other hand, in the context of generative approaches for motion capture, the ultimate objective is to determine the body's pose and shape, achieved by fitting the model to information extracted from images. These methods can generate a series of model parameters like body shape, bone lengths, and joint angles. In contrast to discriminative approaches, generative methods typically rely on temporal data and address a tracking problem. This trend is being developed continuously. Human pose estimation provides human joint information where each key point in humans can be used for robot and human interaction.

In (Cheng, 2021), built a modular interactive framework based on RGB images, which aims to overcome the problems of high dependence on depth cameras and limited distance adaptation in existing human-robot interaction frameworks. However, most of the existing mainstream methods still rely on depth cameras to obtain human joint information. Existing interaction frameworks are affected by the infrared detection distance and thus cannot properly adapt to a variety of different interaction distance.

As human pose estimation is always evolving, its scope is expanding to include approaches that utilize data-drive. Research by (Ming-Hwa et al., 2023), deep learning has undergone rapid development. The use of deep learning covers various field, one of which is human pose estimation. The need to improve accuracy in human pose estimation is also growing, and there are several challenges to overcome. Firstly, how to get the right human pose estimation considering different clothing variations, body shape, variations, and pose variations that may occur. In addition, there is a demand to obtain effective human pose estimation even when applied to many individuals at once.

As we know, BlazePose is an instant human pose detection method capable of recognizing human poses in images or videos. It functions in a single-mode setup, catering to the detection of a single human pose. In simple terms, BlazePose is a sophisticated deep learning model that permits the estimation of human pose through the identification of body segments like elbows, hips, wrists, knees, and ankles. These segments are interconnected to form a skeletal structure that depicts the pose. This model is designed to be efficient, utilizing depth-wise separable convolution to enhance network depth, minimize parameters and computational load, and enhance accuracy. BlazePose provides a comprehensive collection of 33 keypoints, covering areas ranging from the nose to the left foot index.

From some existing research, the input of this research in the form of a single view from fixed position view, and simultaneously taking the object. From the camera view it proceeds to

the 2D joint location detection process. The camera calibration process is needed to obtain the intrinsic and extrinsic features of the camera. With these features and the result of 2D joint location, we get 3D human pose estimation using BlazePose approach.

In this research area, the researchers integrated a human pose estimation system into a transport system as a supporter of the inspection process on the railway, which we often know as an inspection train. This interaction mimics the interaction between humans and robots, but the scope is at the transport level. This design system follows human movement based on human pose estimation along the railway track.

## 2. RESEARCH METHOD

The pipeline in this human pose estimation based on BlazePose system mainly includes four stages, as shown in Figure 1.

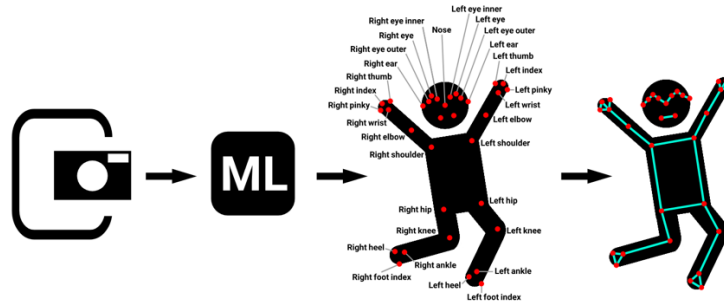


Figure 1. The pipeline in human pose estimation.

The first stage is motion capture using a camera device, where this device acts as the input and output of BlazePose-based human pose estimation. That second is the use of machine learning, which will lead to the architecture of BlazePose itself. In BlazePose there are two types of machine learning used, namely estimators and detectors. The third stage is joint detection based on 33 predefined keypoints that are marker-less motion capture referring to the Vitruvian Man model.

### A. Human Model

Our human body model consists of kinematic 'skeletons' of articulated joints controlled by angular joint parameters  $x_a$ , covered by 'flesh' built from superquadric ellipsoids with additional tapering and bending parameters. A typical model has around 33 joint parameters, plus internal proportion parameters  $x_i$  encoding the position of the hip, clavicle and skull tip joints, plus 9 deformable shape parameters for each body part, gathered into a vector  $x_d$ . The state of a complete model is thus given as a single parameter vector  $x = (x_a, x_i, x_d)$ . We note, however, that only joint parameters are typically estimated during object localization and tracking, the other parameters remaining fixed.

### B. Parameter Estimation

We aim towards a probabilistic interpretation and optimal estimates of the model parameters by maximizing the total probability according to Bayes rule:

$$p(x|\bar{r}) \propto p(\bar{r}|x)p(x) = \exp\{-(e_a + e_s)\} p(x) \quad (1)$$

The above formulation reflects a Bayesian approach where our prior knowledge of the human pose parameters ( $p(x)$ ) is updated based on the observed data ( $\bar{r}$ ) to obtain a more accurate estimate ( $p(x|\bar{r})$ ). This approach provides a probabilistic framework

to approach human pose estimation and maximize the probability of fitting those parameters to the acquired data.

$p(\bar{r}|x)$  is the probability distribution of the observed data or information we have about the human pose ( $\bar{r}$ ), given the parameters ( $x$ ) used to describe the pose. It measures the degree to which those parameters match the observed data.

In this context,  $e_a$  and  $e_s$  can be interpreted concretely as activation energy and error energy, respectively. The activation energy reflects the extent of complexity or cost in determining the human pose parameters, while the error energy measures the degree to which the parameters do not match the observed data. In this framework, we are committed to optimizing the total energy  $e_a$  and  $e_s$  to fully match the available data.

### C. Observation Maximum Likelihood Estimation

In the context of parameter estimation, the likelihood is naturally viewed as function of the parameters  $\theta$ . The joint probability of a set observations, conditioned on a choice for  $\theta$ -repeated here:

$$Lik(\theta, y) \equiv P(y|\theta) \quad (2)$$

Since good predictions are better, a natural approach to parameter estimation is to choose the set of parameter values that yields the best predictions—that is, the parameter that maximize the likelihood of the observed data. This value is called the Maximum Likelihood Estimate (MLE), defined formally as:

$$\widehat{\theta}_{MLE} = \arg_{\theta} \max Lik(\theta, y) \quad (3)$$

In nearly all cases, the MLE is consistent (Cramer, 1964), and gives intuitive results. In many common cases, it is also unbiased. For estimation of multinomial probabilities, the MLE also turns out to be the relative-frequency estimate. Figure 2 visualizes an example of this. The MLE is also an intuitive and unbiased estimator for the means of normal and Poisson distributions.

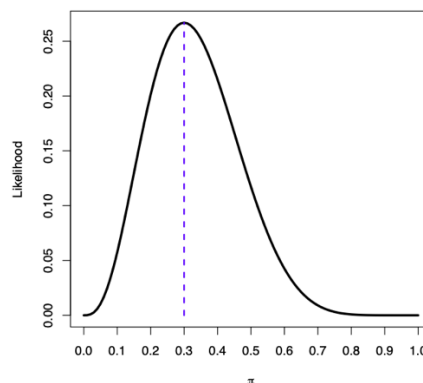


Figure 2. The likelihood function for the binomial parameter  $\pi$  for observed data where  $n = 10$  and  $m = 10$ . The MLE is the Relative Frequency Estimate (RFE) for the binomial distribution. Note that this graph is not a probability density and the area under the curve is much less than 1.

#### D. BlazePose as the Input and Output

The input is either a real-time video capture. The region in the video frame where a person has been detected. Represented as a  $256 \times 256 \times 3$  array with aligned whole human body, centred on the middle of the hip in vertical body pose and rotation distance  $(-10, 10)$ . The channel order RGB with values in  $[0, 0, 1, 0]$ . The breakdown of the process can be seen in Figure 3.

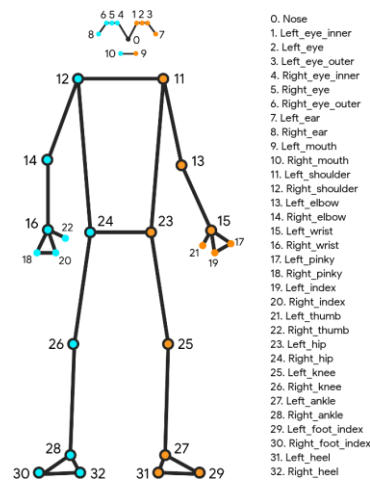


Figure 3. BlazePose keypoints.

The outputs include a  $33 \times 5$  tensor corresponding to the screen-projected keypoints ( $x, y, z$ , visibility, presence), a  $33 \times 3$  tensor corresponding the metric scale coordinates of the 3D world (world  $x$ , world  $y$ , world  $z$ ), and a scalar in the range  $[0.0 ; 1.0]$  corresponding to presence indicating the probability of a person being present on a passed image.

#### E. BlazePose Architecture

BlazePose consist of two machine learning models, a detector and an estimator. The detector cuts out the human region from the input image, while the estimator takes a  $256 \times 256$  resolution image of the detected person as input and outputs the keypoints. The detector is an Single-Shot Detector (SSD) based architecture. There are two ways to use the detector. In box mode, the bounding box is determined from its position ( $x, y$ ) and size ( $w, h$ ). In alignment mode, the scale and angle are determined from  $(kp1x, kp1y)$  and  $(kp2x, kp2y)$ , and bounding box including rotation can be predicted. The estimator uses heatmap for training, but computes keypoints directly without using heatmap for faster inference.

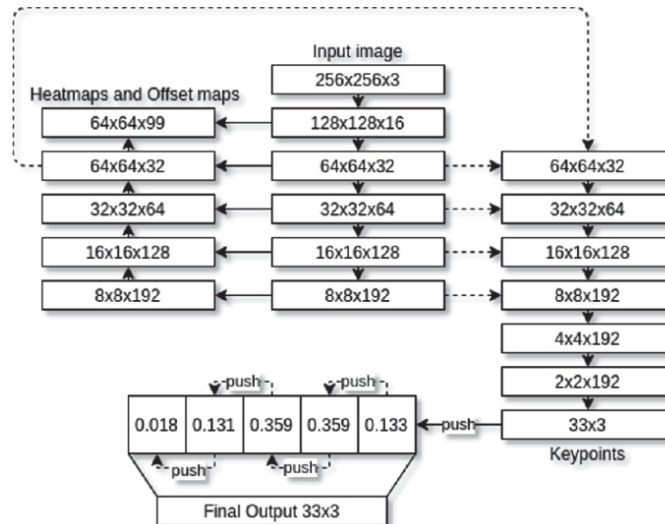


Figure 4. Tracking network architecture of regression with heatmap supervision.

The first output estimator is (1, 195) landmarks, the second output is (1, 1) flags. The landmarks are made of 165 elements for the (x, y, z, visibility, presence) for every 33 keypoints. The z-values are based on the person's hips, with keypoints being between the hips and the camera then the value is negative, and behind the hips when the value is positive.

The visibility and presence are stored in the range of [min\_float, max\_float] and are converted to probability by applying a sigmoid function. The visibility returns the probability of keypoints that exist in the frame.

## F. Rail Inspection System

The design of this system uses distance detection and human pose estimation captured by a webcam camera then programming is carried out using the BlazePose method and forwarded to the ESP32 microcontroller, after which it is connected to the PLC using Modbus RS485 communication as TX and RX inputs, while the Rotary Encoder which is a speed sensor is connected to the PLC at input address S8 which has a High-Speed Counter function.

The two inputs will be processed using a PID controller which is output in the form of a PWM signal. The PWM signal issued by the PLC at R7 is then converted by the PWM to Voltage module and forwarded to the BLDC Motor Controller. The voltage entering the BLDC Motor Controller will be processed and forwarded to the BLDC Motor to adjust the motor speed. The 48V 10Ah Lithium Ion Battery is used as the main power source of the BLDC Motor Controller and BLDC Motor which enters the battery socket on the BLDC Motor Controller. The system can be seen in Figure 5.

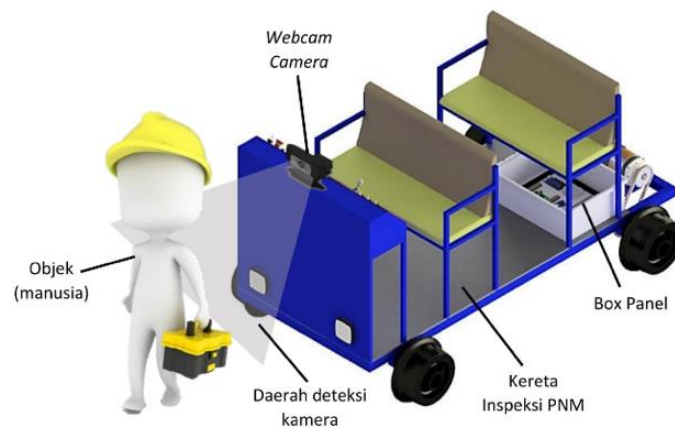


Figure 5. Integrating system to rail inspection.

### 3. RESULTS AND DISCUSSION

From this research, the results of estimating the human body using BlazePose are obtained. However, the results of this research will be divided into two categories. The first category is the result of human pose estimation using BlazePose by validating the keypoints value obtained. And the second, bread on the scope of integration of the human pose estimation system for rail inspection.

To evaluate model, researchers checked the coordinates of 33 keypoints (x, y, z). The result of the 3D projection can be seen in Table 1.

**Table 1: The result of 3D projection.**

ID	Joints	Pixel	Image Projection		
			x	y	z
0.	Nose	564	0.537945	0.228714	-0.7999144
1.	Left eye (inner)	565	0.469518	0.190116	-0.530800
2.	Left eye	566	0.463484	0.178852	-0.564803
3.	Left eye (outer)	567	0.450156	0.173961	-0.478437
4.	Right eye (inner)	568	0.441001	0.172332	-0.396517
5.	Right eye	569	0.413222	0.157255	-0.340979
6.	Right eye (inner)	570	0.410556	0.152512	-0.292190
7.	Left ear (inner)	571	0.404946	0.151231	-0.302297
8.	Right ear (inner)	572	0.404004	0.152843	-0.282213
9.	Mouth (left)	573	0.398666	0.149637	-0.257025
10.	Mouth (right)	574	0.397440	0.148363	-0.261147
11.	Left shoulder	575	0.398961	0.145511	-0.322533
12.	Right shoulder	576	0.401024	0.141420	-0.334692
13.	Left elbow	577	0.404488	0.134825	-0.329674
14.	Right elbow	578	0.404458	0.132088	-0.335903
15.	Left wrist	579	0.404410	0.130728	-0.332628
16.	Right wrist	580	0.405184	0.128445	-0.369504
17.	Left pinky	581	0.405475	0.127273	-0.379874
18.	Right pinky	582	0.406179	0.126494	-0.323348



19.	Left index	583	0.406843	0.126528	-0.297850
20.	Right index	584	0.407295	0.126410	-0.276154
21.	Left thumb	585	0.407344	0.125873	-0.213827
22.	Right thumb	586	0.407805	0.125554	-0.253524
23.	Left hip	587	0.411940	0.123076	-0.293951
24.	Right hip	588	0.415726	0.121805	-0.323780
25.	Left knee	589	0.418729	0.121242	-0.372628
26.	Right knee	590	0.420105	0.120559	-0.362168
27.	Left ankle	591	0.421403	0.118802	-0.326854
28.	Right ankle	592	0.421945	0.117704	-0.312582
29.	Left heel	593	0.423139	0.115538	-0.365676
30.	Right heel	594	0.423489	0.114297	-0.372545
31.	Left foot index	595	0.424299	0.113465	-0.376875
32.	Right foot index	596	0.425115	0.112725	-0.372456

The coordinate results from the 3D projection are then validated to determine whether the point in the human pose estimation is in accordance with the joint keypoint. This can be seen in Figure 6.

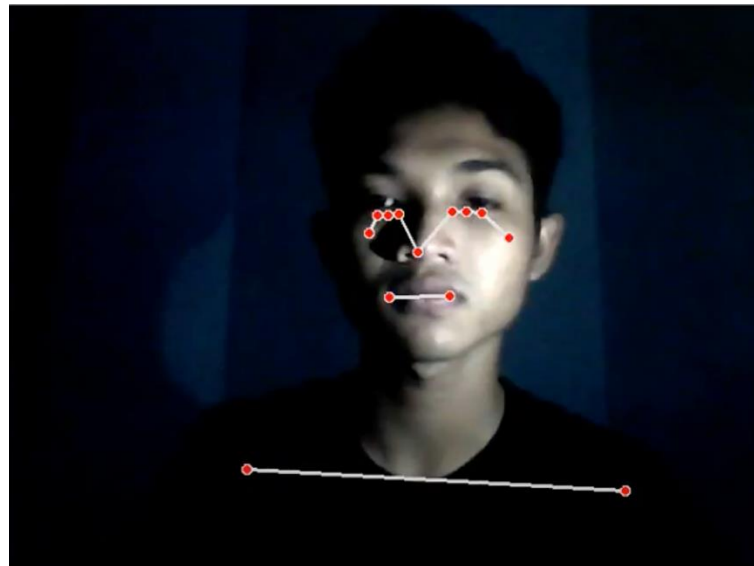


Figure 6. Result of 2D human pose estimation keypoints.

It can be seen that the keypoints in Figure 6 include several points in the face and chest area. The results of the taping can be seen in Table 2. These results are stated to be in accordance with the reference keypoints index.

**Table 2: Keypoints validation of 2D human pose estimation keypoints.**

ID	Joints	Pixel
0.	Nose	564
1.	Left eye (inner)	565
2.	Left eye	566
3.	Left eye (outer)	567
4.	Right eye (inner)	568
5.	Right eye	569
6.	Right eye (inner)	570
9.	Mouth (left)	573



After the design system has been matched with the keypoints, the human pose estimation is then integrated with the rail inspection system through the motor, which further regulates the range distance, so that the system will get information on the optimum joint distance. The system will manage the rail inspection in which direction to move through the human movement on the rail track. If a good estimation of the human body is detected (in standing condition) then the motor condition is on, and landmarks are detected. If the inspection direction is backward, the direction will be reverse. Conversely, if the inspection direction is forward, the direction will be forward.

**Table 3: Integrating motor to system based on poses.**

Motor Condition	All Landmarks	Direction
ON	Detected	Reverse
OFF	Undetected	Reverse
ON	Detected	Forward
OFF	Undetected	Forward



Figure 7. Motor condition is ON, overall landmark is DETECTED, direction is REVERSE.



Figure 8. Motor condition is OFF, overall landmark is UNDETECTED, direction is REVERSE.



Figure 9. Motor condition is ON, overall landmark is DETECTED, direction is FORWARD.



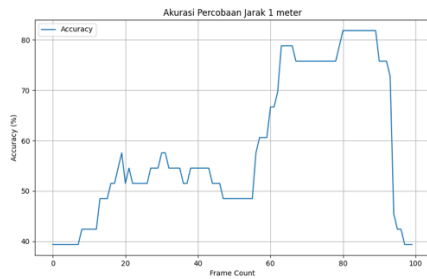
Figure 10. Motor condition is OFF, overall landmark is UNDETECTED, direction is FORWARD.

The results of the integration provide information on the optimal coverage distance for rail inspection of objects (human poses). In this condition, the researcher uses a distance of 1-12 meters, where this distance is the optimal range based on triangulation (image projection). In addition, this distance can estimate the object detection range which is useful for activating the motor status.

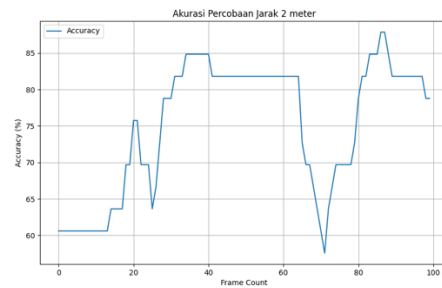
**Table 4: Optimal distance accuracy and motor activation conditions.**

Distance (Meters)	Accuracy (%)	MSE Loss	Motor Status
1	58.82	491.438	OFF
2	75.09	208.707	OFF
3	86.39	448.716	ON
4	93.85	171.128	ON
5	90.52	544.836	ON
6	78.42	812.601	ON
7	92.64	394.725	ON
8	93.88	249.118	ON
9	89.3	346.738	ON
10	86.33	407.657	ON
11	83.27	234.755	ON
12	89.33	340.036	ON

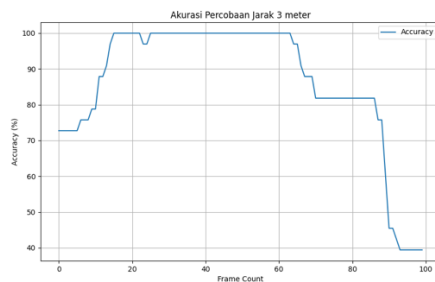
In Table 4, it can be seen and analysed that the ideal distance for the system to detect and state the motor ON is at a distance of 4, 5, 7, and 8 metres. It indicates that the graph with the fourth cluster of distances condition shows a good initialisation between frame count and accuracy compared to the other distances. However, in the initialisation condition or at a distance of 1 meter and 2 meters the motor status is OFF. It happens that the detection of human pose estimation is too close to the distance of the camera, and the range of the camera still cannot see the position of the rail. The results of the distance accuracy graph from 1 metre to 12 metres can be seen in the following figure.



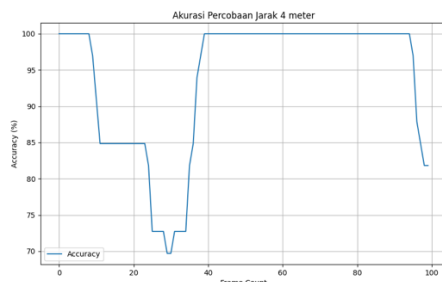
(a). Accuracy graph with a distance of 1 metre.



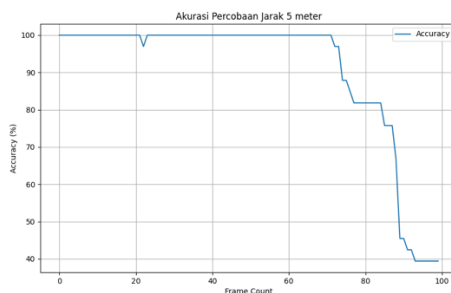
(b). Accuracy graph with a distance of 2 metre.



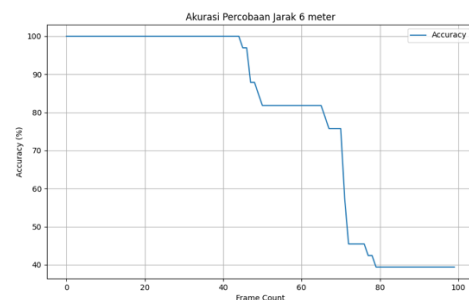
(c). Accuracy graph with a distance of 3 metre.



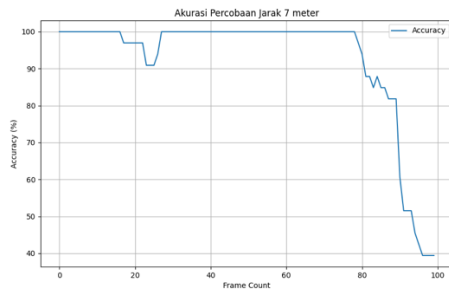
(d). Accuracy graph with a distance of 4 metre.



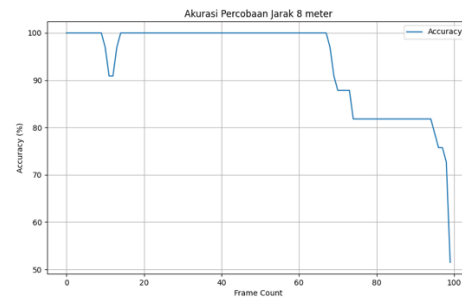
(e). Accuracy graph with a distance of 5 metre.



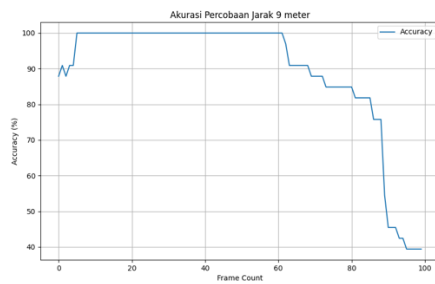
(f). Accuracy graph with a distance of 6 metre.



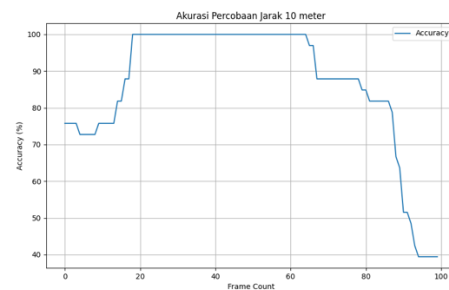
(g). Accuracy graph with a distance of 7 metre.



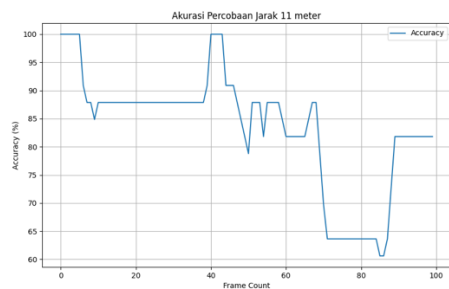
(h). Accuracy graph with a distance of 8 metre.



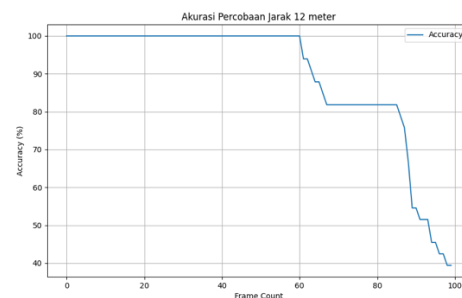
(i). Accuracy graph with a distance of 9 metre.



(j). Accuracy graph with a distance of 10 metre.



(k). Accuracy graph with a distance of 11 metre.



(l). Accuracy graph with a distance of 12 metre.

Figure 9. Graphical results of accuracy against distance.

#### 4. CONCLUSION

The technique of using computer vision as an automatic control system of the BLDC Motor-driven inspection cart is able to generate numerical values of the pixel values of the keypoints, the accuracy level, and also the coordinates of the keypoints.

The use of human pose estimation using BlazePose is used to be able to detect human optical motion gestures so that decisions can be obtained when humans are walking and humans are not walking which will be recognized through validation of the keypoints values obtained which will then be integrated with the control system of the inspection train where if the human is identified as walking then the system will turn on the motor so that the motor is ON, while if the human is identified as silent then the motor is OFF or does not turn on.

The distance accuracy rate obtained from this detection is the optimal detection distance of 12 meters from the camera with the most optimal accuracy in the range of 4-8

meters from the camera with an average detection accuracy rate of 89.862%. While the lowest level of accuracy is found at a distance of 1-2 meters from the camera so that in this condition results in decision making The motor does not turn on.

From all the tests carried out so that the results obtained the total detection accuracy rate of 84.82% and the motor will OFF in the range of 1-2 meters then the motor will ON in the range of 3-12 meters. This level of accuracy is highly dependent on the color segmentation capabilities of the program that has been created, room lighting conditions, and the level of camera resolution used. The level of accuracy is also influenced by how many test samples are carried out so that the system is tested in detail.

## REFERENCES

- Agustin, Y. H., & Sugiharto, A. (2017). Implementasi Visi Komputer Untuk Mengidentifikasi Mobilitas Kendaraan Pada Citra Jalan Raya. *Jurnal VOI (Voice Of Informatics)*, 6(2).
- Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., & Grundmann, M. (2020). BlazePose: On-device real-time body pose tracking. *arXiv preprint arXiv:2006.10204*.
- Cheng, Y., Yi, P., Liu, R., Dong, J., Zhou, D., & Zhang, Q. (2021, May). Human-robot interaction method combining human pose estimation and motion intention recognition. In *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)* (pp. 958-963). IEEE.
- Fitriyah, H. (2020). Pengukuran Panjang-Berat Ikan dan Sayuran pada Budikdamber (Budi Daya Ikan dalam Ember) Menggunakan Visi Komputer dan Regresi Linier. *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, 4(1), 8-14.
- Grishchenko, I., Bazarevsky, V., Zhanfir, A., Bazavan, E. G., Zhanfir, M., Yee, R., ... & Sminchisescu, C. (2022). BlazePose ghum holistic: Real-time 3d human landmarks and pose estimation. *arXiv preprint arXiv:2206.11678*.
- Irfan, I., Kartika, K., & Meliala, S. M. S. (2023). PENGIRAAN POSE MODEL MANUSIA PADA REPETISI KEBUGARAN AI PEMOGGRAMAN PYTHON BERBASIS KOMPUTERISASI. *INFOTECH journal*, 9(1), 11-19.
- Kharka, K. B., Bhutia, T. R. W., Chettri, L., Subba, R., Giri, E., & Lepcha, S. (2021). Human Following Robot Using Arduino Uno. *Open Access*, 3(07), 5.
- Liu, M., Han, S., & Lee, S. (2016). Tracking-based 3D human skeleton extraction from stereo video camera toward an on-site safety and ergonomic analysis. *Construction Innovation*, 16(3), 348-367.
- Marpaung, F., Aulia, F., & Nabila, R. C. (2022). COMPUTER VISION DAN PENGOLAHAN CITRA DIGITAL.
- Raihan, M. J., Hasan, M. T., & Nahid, A. A. (2022). SMART HUMAN FOLLOWING BABY STROLLER USING COMPUTER VISION. *Khulna University Studies*, 8-18.
- Shah, M., Gandhi, K., Pandhi, B. M., Padhiyar, P., & Degadwala, S. (2023, May). Computer Vision & Deep Learning based Realtime and Pre-Recorded Human Pose Estimation. In *2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* (pp. 313-319). IEEE.
- Shuai, L., Li, C., Guo, X., Prabhakaran, B., & Chai, J. (2016). Motion capture with ellipsoidal skeleton using multiple depth cameras. *IEEE transactions on visualization and computer graphics*, 23(2), 1085-1098.
- Steven, G., Liliana, L., & Purbowo, A. N. (2022). Penerapan 3D Human Pose Estimation Indoor Area untuk Motion Capture dengan Menggunakan YOLOv4-Tiny, EfficientNet Simple Baseline, dan VideoPose3D. *Jurnal Infra*, 10(2), 198-204.
- Toshev, A., & Szegedy, C. (2014). DeepPose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1653-1660).
- Wijaya, A. A., & Prayudi, Y. (2010, June). Implementasi Visi Komputer Dan Segmentasi Citra Untuk Klasifikasi Bobot Telur Ayam Ras. In *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*.
- Zhang, C., Kang, K., Li, H., Wang, X., Xie, R., & Yang, X. (2016). Data-driven crowd understanding: A baseline for a large-scale crowd dataset. *IEEE Transactions on Multimedia*, 18(6), 1048-1061.
- Zhao, B., Zhu, W., Hao, S., Hua, M., Liao, Q., Jing, Y., ... & Gu, X. (2023). Prediction heavy metals accumulation risk in rice using machine learning and mapping pollution risk. *Journal of Hazardous Materials*, 448, 130879.