

VIRTUAL MOUSE WITH HAND GESTURE RECOGNITION BASED ON HAND LANDMARK MODEL FOR POINTING DEVICE

Jeffri Dian Asmoro^{1*}, Achmad Teguh Wibowo¹, Mujib Ridwan¹

¹Information System, UIN Sunan Ampel Surabaya
e-mail: * jeffridianasmoro@gmail.com

Abstract: Technology is growing rapidly and has become one of the human needs that must be owned to solve the problems being faced. The development of touchless input devices or hand gesture recognition using a camera is a form of machine learning. Gestures can define as physical movements of the hands, arms, or body as expressive messages, besides that this hand gesture system can explain the contents of commands that have meaning. In this research, a virtual mouse system will be developed using hand gesture recognition based on the hand landmark model for pointing devices. The resulting application can be run on a desktop device using a webcam. The results of the tests carried out to analyze the implementation of the hand landmark model into the system show that the average system accuracy reaches 96% and the speed reaches 0.05 seconds.

Keywords: hand gesture recognition, hand landmark models, machine learning, virtual mouse

Abstract: Teknologi semakin pesat dan sudah menjadi salah satu kebutuhan manusia yang harus dimiliki untuk menyelesaikan permasalahan yang sedang dihadapi. Perkembangan piranti masukan tanpa sentuhan atau hand gesture recognition menggunakan kamera adalah salah satu bentuk dari machine learning. Gestur mampu mendefinisikan sebagai gerakan fisik dari tangan, lengan, maupun badan sebagai pesan yang ekspresif, selain itu sistem gestur tangan ini mampu menjelaskan isi perintah yang memiliki arti. Dalam penelitian ini akan dikembangkan sebuah sistem virtual mouse menggunakan hand gesture recognition berbasis hand landmark model untuk pointing device. Aplikasi yang dihasilkan dapat dijalankan pada perangkat desktop dengan menggunakan webcam. Hasil dari pengujian yang dilakukan untuk menganalisa penerapan hand landmark model kedalam sistem menunjukkan rata-rata akurasi sistem mencapai 96% dan kecepatan mencapai 0.05 second.

Keywords: hand gesture recognition, hand landmark models, machine learning, virtual mouse

INTRODUCTION

The rapid development of technology is of course also accompanied by changes in behavior in people who prefer to carry out activities quickly and practically. Technological progress is an unavoidable phenomenon. Every technological idea is presented to have a positive impact on life. Until now humans have enjoyed many positive benefits from the presence of technology[1]. So that technology is the discovery of objects or tools that are realized from the implementation of science and skills possessed by humans. Currently, computers are still often used in interaction with humans, for example pressing buttons on the keyboard or touching with a mouse. So that the input will produce the information as desired. The main purpose of human-computer interaction is to make it easier for users to operate computers. The working principle of a computer or technology is that there is input, process, and produces output.

Computer operation requires an input device. This tool has the function of entering and recording data in a computer system and giving orders. Common devices used for personal computers (PCs) are keyboards or mice[2]. But the mouse tends to often experience tracking, namely jumping, sometimes the mouse cannot be detected by the PC, so the mouse is not efficient, from this discussion the user needs an interface with a computer that is easier to use and efficient[3].

One solution is to use the camera as an input device in real time[2]. So if the interaction can be done efficiently then the input process will be effective. Device selection needs to pay attention to psychological factors, convenience, and daily habits with natural interactions when using technology[4].

The development of touchless input devices or hand gesture recognition uses a camera as a hand gesture detector to recognize hand movements so that the visual interpretation of hand gestures can provide a convenient effect[5]. Gestures can define as physical movements of the hands, arms, or body as expressive messages, besides that this hand gesture system can explain the contents of commands that have meaning[6].

Research[7] entitled "Vision-Based Hand Gesture Recognition" explains that the use of hand gesture recognition can use the Data-Glove or computer vision method. The difference lies in the Data-Glove method, which requires a sensor device to digitize hand movements into multiparametric data, so it is quite expensive and encounters many problems.[8]. The solution to overcome this problem is that a camera is needed for Computer Vision so that it can create natural interactions between humans and computers without additional devices[9].

The development of algorithms and computers to simulate human visualization so that it can automatically extract information from the price of an object is called computer vision [10]. Computer vision is currently developing rapidly along with the increasing human need for technology. Today, the speed of data processing has increased significantly, so that it can assist humans in complex tasks. But for input technology, it causes new problems in performing some tasks, for example, lag, less utilization of available resources, and limits application usage expression. That way hand gesture recognition is a way to overcome it. The use of Hand Gesture Recognition uses the Hand Landmark Model. Hand Landmark Model serves to allocate 21 points on the hand[10]. The advantage of having a Hand Landmark is that it is easier to

implement so that it has a binary output, and can distinguish between the right and left hands.

In this study, several previous studies were using various methods and different applications, one of which was entitled "Hand Gesture Recognition as a Substitute for a Computer Mouse Using a Camera". The method used is the convex hull algorithm, which is a method by making the number of fingers that will later become a reference for working on mouse actions. The application of this method still has drawbacks, namely, the convex hull algorithm method still has problems distinguishing the same object with skin color and also an irregular background [11].

Furthermore, in research [12] title "Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning" using the Hand Landmark Model and Palm Detector. The Palm Detector detects hands accurately which will then be forwarded to the Hand Landmark Detector. By getting the results of an accuracy percentage of 99%. This study concludes that real-time accurate detection using the Hand Landmark Model and Palm Detector without sensors can be applied in using this technology more effectively and efficiently. Another advantage can reduce the data augmentation process, namely rotation, flipping, and scaling.

Subsequent research by [13] entitled "Convex Hull Method and Convexity Defects for Recognition of Hand Signs". The method used with convex hull and convexity defects. The convex hull method detects Hull points as fingertips and convexity defects look for finger descriptions. In its application, it still has limitations that only use the left hand and irregular backgrounds or backgrounds that have skin color are still difficult to

detect. This is also due to the excessive lighting process so that the detection becomes unstable.

In the application of a support vector machine to research (Neneng et al., 2016) with the title "Support vector machine for image classification of types of meat based on texture using feature extraction Gray Level CO Occurance Matrices (GLCM)" using the GLCM method shows evidence that texture is a descriptor strongest for detecting an image. The best distance results for image recognition with a distance of 20 cm with the best recognition rate of 87.5%.

Based on [14] entitled "Real-Time Hand Gesture Recognition Using Different Algorithms Based on American Sign Language" the method used is the combined method, namely K Convex Hul with this combined method, the recognition accuracy rate is very high, and sharp, namely 94.23%. However, this method has the disadvantage that it can only be detected by using a black background so that it can minimize errors when recognizing gestures.

Then in research [15] the title "Hand Gesture Recognition System For Multimedia Applications" used the Kinect method. The Kinect method is a method of human and computer interaction so that it can run naturally. With the research results, namely through controlling multiple applications such as pdf readers, using a web camera. However, this research has limitations, namely, the camera can only read pdf so humans cannot interact directly with the touchscreen.

After that in research [16] with the title "Hand Gesture Recognition and Voice Conversion for Hearing and Speech Aided Community" using the Support Vector Machine method, K neighbors Classifier, Logistic Regression, MLP Classifier, Naïve Bayes, Random

Forest Algorithm classifier. Using this method can design a system to help people who train the deaf to communicate with the rest of the world using sign language or hand gesture recognition techniques.

Furthermore, in research [17] entitled "Application of the Support Vector Machine Method in Diagnosing Hepatitis" applying the Support Vector Machine method. The results of the method analysis ability are known by testing using data testing through both kernels with training data of 100 positive data and 100 negative data. The test results using the linear kernel function get a percentage of 68-83% and the RBF kernel function is 70-96%.

After that, the last previous research was "SSd: Single Shot Multibox detector" with the authors [18] using the SSD method, PASCAL Dataset, VOC, COCO, and ILSVRC. With the simple SSD method and requiring further objects, there are stages of resampling pixels and summarizing features in one network so that this method can reduce the occurrence of detection errors. Compared to other single-stage methods, SSD has much better accuracy even with smaller input image sizes.

From several previous research references, it can be concluded that this research uses SSD because it is easier to implement and uses the Hand Landmark Model and Palm Detector so that system processes are resolved and used more quickly, besides that it does not require training so it is more efficient.

METHOD

Preparing tools and materials such as PCs and Python software, the next step is regarding the implementation of the Hand

Gesture Recognition system flow.

In short, the way this software works is by using the camera media from a laptop/pc, the camera functions as an identification of the palm which will later be extracted using the SSD method into numeric numbers. Once a hand is detected, it proceeds to localization landmarks to precisely localize the 21 3D knuckle coordinates (i.e., x, y, z axes) within the detected hand region. Then it is continued in the Normalizing and Training & Validation process which functions as a process of normalizing the x and y coordinates to suit the system. data files are then prepared to be broken down into training and validation sets. 80% of the data is reserved for training the model with various optimization and loss functions, while 20% of the data is reserved for validating the model. Then proceed to the trigger that was created.

The first step is to install Python. Python is used because it has the advantage of being easy to apply to technology development according to [19].

The second is to install the package. At this stage, install the packages needed for the needs of the Hand Gesture Recognition system in the form of OpenCV, Mediapipe, and AutoPY. OpenCV is used with a special purpose for image processing [20]. While the MediaPipe has a function to be able to recognize hand gesture objects [21]. MediaPipe has a large collection of who was trained human body detection and tracking models on the very large and most diverse Google data set. AutoPY is a simple cross-platform GUI automation library in python, that functions to be able to control the keyboard and mouse, find color and bitmap layers, and will display alerts. Doing Application Coding At the final stage of the manufacturing process, coding is carried out into the

IDLE text editor tool that has been provided by Python using version 3.8 because the library is compatible with support, in contrast to using version 3.9 there are constraints on the Tensorflow library which cannot support

In the final stage, testing is carried out after the development of hand gesture recognition technology has been completed and can be run, which will determine the results of this research. Furthermore, in conducting testing based on [21] do the test using the distance parameter. This distance parameter is carried out to see whether the distance can affect program or system performance. After testing the distance, it will then be calculated using the confusion matrix calculation [12]. The confusion matrix test aims to see the results of the accuracy of the hand gesture recognition system. Tests on distances starting from 1 meter to 2.5 meters.

RESULTS AND DISCUSSION

Hand Gesture Recognition (HGR) technology was developed based on the results of a needs analysis in the form of a system architectural design, as shown in Image 1.

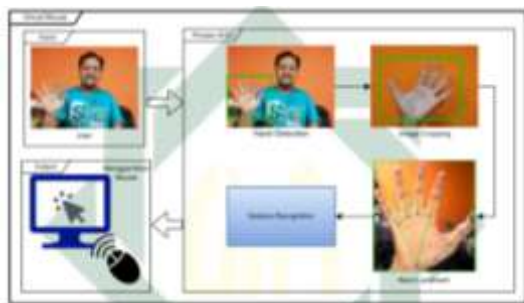


Image.1 System Architecture Design

In the system architecture design, there are 3 main parts, namely initial conditions, data processing, and output.

The initial part is to detect the image. After the image can be detected it is forwarded to the process section, namely first-hand detection to detect hands, when the hand can be detected it will only crop the hand. If the crop process has been carried out, the process will continue at the stage of giving landmark points on the hands. Meanwhile, the function of gesture recognition is to make the functions of each finger execute mouse commands that have been adapted to the program created. After going through several processes, the final step is to be able to output, namely by being able to use HGR technology which is used to replace the mouse as a pointing device.

In testing, it will produce a test consisting of 7, namely s (test scenario), ss (test sub scenario), r1 (average of 3 tests for each test sub scenario), r2 (average of r1), success (if 1 means the detection is correct, and 0 means false), and j (the number of detections is correct). All numbers contained in the column are the time needed by the system to detect. It was found that all tests were successful by achieving an accuracy rate of 95% with a total of 216 tests with an average working speed of 0.056 seconds.

Based on each aspect of the test, both from the distance test and the light test, above-average results have been obtained as follows. In the distance test with 144 tests, the average speed of the system in detecting hands was 0.05 Seconds, then from the recall results it got 96.5%, Precision got 100%, and Accuracy got 96.5%. Furthermore, in the light test with a total of 216 tests, the average speed of the system in detecting hands was 0.06 seconds, this was much different from the results of research based on [22] which produces a difference of 0.5 seconds by 10x faster, then from the recall results get 95.8%, Precision gets

100%, and Accuracy gets 95.8% results. So it can be said that the system created meets the real-time criteria which runs 10x faster and are running well.

CONCLUSION

The design of hand gesture recognition technology can be identified. First, in making hand gesture recognition technology using hand detection to be able to detect hands, hand landmarks to provide points to the palms that have been detected so that they can be used to run the process of pointing devices to replace the mouse, the use of the palm can be explained in detail, namely the little finger to click right, index finger left click, middle finger and ring finger to double click and all fingers to move the cursor or mouse pointing device. Second, by implementing the Hand Landmark Model on Virtual Mouse as a Pointing Device, it has a high level of accuracy, which is above 90%, and has a low failure rate, which is below 10%. Third, The application of the hand landmark model for the virtual mouse can be said well with evidence. In the distance test with a number of tests of 144, the average system speed results in detecting hands are 0.05 Seconds, then the recall results get 96.5%, and precision gets 100%. , and Accuracy results were obtained at 96.5 %. Furthermore, in the light test with the number of tests of 216, the average speed of the system in detecting hands was 0.06 Second, then from the recall results obtained 95.8%, Precision obtained 100%, and Accuracy obtained results of 95.8%. So that it can be said that the system indicates the load can run well by the Hand Gesture Recognition system, the placement of the Pointing Device is also appropriate because the

success and readiness of the system can run well.

BIBLIOGRAPHY

- [1] M. Ngafifi, "Kemajuan Teknologi Dan Pola Hidup Manusia Dalam Perspektif Sosial Budaya," *J. Pembang. Pendidik. Fondasi dan Apl.*, vol. 2, no. 1, pp. 33–47, 2014, doi: 10.21831/jppfa.v2i1.2616.
- [2] S. N. Supal, M. Hayati, and M. Yatim, "Kemahiran meta kognitif kanak-kanak dalam penggunaan perisian dan peranti input tanpa sentuhan leap motion," *J. ICT Educ.*, vol. 7, no. 2, pp. 1–13, 2020.
- [3] Murthy;Jadon, "A Review of Vision Based Dynamic Hand Gestures Recognition," *Comput. Sci. Appl.*, vol. 10, no. 05, pp. 990–1001, 2009, doi: 10.12677/csa.2020.105102.
- [4] Falabiba, "Modul Interaksi Manusia dan Komputer," 2019.
- [5] A. Sunyoto and A. Harjoko, "Review Teknik, Teknologi, Metodologi dan Implementasi Pengenalan Gestur Tangan Berbasis Visi," *Semin. Nas. Apl. Teknol. Inf.*, p. H-7, 2014, [Online]. Available: <http://jurnal.uui.ac.id/Snati/article/view/3290>
- [6] S. Anwar, S. K. Sinha, S. Vivek, and V. Ashank, "Hand gesture recognition: A survey," *Lect. Notes Electr. Eng.*, vol. 511, no. 3, pp. 365–371, 2019, doi: 10.1007/978-981-13-0776-8_33.
- [7] P. Garg, N. Aggarwal, and S. Sofat, "Vision based hand gesture recognition," *Proc. Int. Conf. Serv.*

- Sci. ICSS*, pp. 260–265, 2009, doi: 10.1109/ICSS.2013.40.
- [8] A. Mulder, “Hand gestures for HCI,” *Hand Centered Stud. Hum. Mov. Proj. ...*, no. February, pp. 1–19, 1996, [Online]. Available: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Hand+Gestures+for+HCI#3%5Chttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Hand+gestures+for+HCI#3>
- [9] Y. Velaskar, A. Dulam, S. Sureliya, S. Shenoy, and C. Chouhan, “Computer Vision based Hand Gesture Interfaces,” *Int. J. Eng. Res. Technol.*, vol. 05, no. 01, pp. 1–4, 2017, [Online]. Available: www.ijert.org
- [10] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, “Hand keypoint detection in single images using multiview bootstrapping,” *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 4645–4653, 2017, doi: 10.1109/CVPR.2017.494.
- [11] H. Yunita and E. Setyati, “Hand Gesture Recognition Sebagai Pengganti Mouse Komputer Menggunakan Kamera,” *J. ELTIKOM*, vol. 3, no. 2, pp. 64–76, 2019, doi: 10.31961/eltikom.v3i2.114.
- [12] A. Halder and A. Tayade, “Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Network,” *Int. J. Res. Publ. Rev.*, no. 2, pp. 9–17, 2021.
- [13] N. Yuliana, K. Ratri, and R. Wardani, “Metode Convex Hull dan Convexity Defects untuk Pengenalan Isyarat Tangan,” *J. Telemat.*, vol. 11, no. 2, pp. 81–88, 2016.
- [14] M. M. Islam, S. Siddiqua, and J. Afnan, “Real time Hand Gesture Recognition using different algorithms based on American Sign Language,” *2017 IEEE Int. Conf. Imaging, Vis. Pattern Recognition, icIVPR 2017*, 2017, doi: 10.1109/ICIVPR.2017.7890854.
- [15] N. S. Rokade, H. R. Jadhav, S. A. Pathan, and U. Annamalai, “Hand Gesture Recognition System For Multimedia Applications,” *Int. Res. J. Eng. Technol.*, pp. 1894–1899, 2016, [Online]. Available: www.irjet.net
- [16] D. K V, H. E, N. Jain D, and N. Reddy B, “Hand Gesture Recognition and Voice Conversion for Hearing and Speech Aided Community,” *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 6, no. 3, pp. 223–228, 2020, doi: 10.32628/cseit206346.
- [17] R. Munawarah, O. Soesanto, and M. R. Faisal, “Penerapan Metode Support Vector Machine,” *Kumpul. J. Ilmu Komput.*, vol. 04, no. 01, pp. 103–113, 2016.
- [18] W. Liu *et al.*, “SSD: Single shot multibox detector,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9905 LNCS, pp. 21–37, 2016, doi: 10.1007/978-3-319-46448-0_2.
- [19] T. M. Kadarina and M. H. Ibnu Fajar, “Pengenalan Bahasa Pemrograman Python Menggunakan Aplikasi Games Untuk Siswa/I Di Wilayah Kembangan Utara,” *J. Abdi Masy.*, vol. 5, no. 1, p. 11, 2019, doi: 10.22441/jam.2019.v5.i1.003.
- [20] T. C. A.-S. Zulkhaidi, E. Maria,

- and Y. Yulianto, “Pengenalan Pola Bentuk Wajah dengan OpenCV,” *J. Rekayasa Teknol. Inf.*, vol. 3, no. 2, p. 181, 2020, doi: 10.30872/jurti.v3i2.4033.
- [21] S. S. Makahaube, A. M. Sambul, and S. R. U. Sompie, “Implementation of Gesture Recognition Technology for Automated Education Service Kiosk,” vol. 16, no. 4, pp. 465–472, 2021.
- [22] Z. Ren, J. Meng, and J. Yuan, “Depth Camera Based Hand Gesture Recognition and its Applications in Human-Computer-Interaction,” pp. 1–5, 2011.