IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 26, Issue 6, Ser. 1 (Nov. – Dec. 2024), PP 60-66
www.iosrjournals.org

Humanoid Head MIKE

Gonzalez Sánchez América María[1], Alcántara Montes Samuel[1]
[1](*Higher School Of Mechanical And Electrical Engineering, National Polytechnic Institute, Campus Zacatenco, Mexico)*

Abstract:
The humanoid robot MIKE is a PhD project developed in the Graduate Studies Section of the School of Mechanical and Electrical Engineering, Zacatenco campus of the National Polytechnic Institute (Instituto Politécnico Nacional).
In this article, the head of the humanoid robot called MIKE is presented, which consists of the design of a 3D printed prototype, has 2 degrees of freedom for the movement of rotation and tilt of the head through 2 servomotors respectively, has incorporated a dual stereo camera IMX219-83 that gives it artificial vision, all this is controlled by Artificial Intelligence (AI) algorithms by means of a Raspberry Pi 5 minicomputer. To identify and classify objects, people and scenarios, programs developed in Python 3.11 were created and Open CV and RPiCam algorithms were also used.
Section II presents the prototype design, head components, control system design, algorithm development and final assembly.
Section III presents the tests with different scenarios and the results obtained.
Section IV gives the conclusions obtained from the development of this work.
Key Word: *Humanoid head, Artificial Intelligence (AI), Artificial Vision, Humanoid Robot.*

Date of Submission: 29-11-2024 Date of Acceptance: 09-12-2024

I. Introduction

Artificial intelligence (AI) has had a transformative impact on the field of robotics, significantly improving the capabilities, autonomy, and efficiency of robots.

AI enables robots to make autonomous decisions based on their perception of their environment and predefined goals. AI algorithms, especially those based on machine learning and deep learning, enable robots to analyze sensor data in real time, identify patterns and make decisions without direct human intervention.

Computer vision is a field of AI that enables computers and systems to obtain meaningful information from digital images, videos and other visual inputs, and take actions or make recommendations based on that information.

Machine vision trains machines to perform various functions, and it has to do this in a very short time with built-in cameras, from which to obtain data and apply algorithms instead of retinas, optic nerves and a visual cortex. [1].

Undoubtedly one of the most remarkable and impressive early humanoids was ASIMO [2], has a tope body, can walk in a stable and smooth way, has a head with 2 degrees of freedom, as well as a fixed vision system, also the HRP-3P humanoid [3] has these characteristics.

Humanoid robotics has grown tremendously, there are more sophisticated robots such as the Karlsruhe humanoid head [4], which has seven degrees of freedom. The eyes can move independently. Each eye is equipped with two digital color cameras, one with a wide-angle lens for peripheral vision and one with a narrow-angle lens for foveal vision.

However, the design of humanoid systems capable of executing manipulation and grasping tasks, interacting with humans, and learning from human observation requires sophisticated perception systems, able to meet the associated requirements [4]. Humanoid robots with human-like heads have been developed for emotional human-robot interaction ([5], [6]) and for the study of cognitive processes ([7], [8]).

One of the most advanced robots with respect to AI is Tesla's humanoid Optimus, in the AI Day (2021) event presented by Elon Musk announces that the company Tesla will start with the construction and future commercialization of a bot (humanoid robot), there he shows the outline of what will be its future bot; one year later in September 2022, Bumblebee is introduced; in March 2023, an update is made and it is named Optimus-Gen 1; in December 2023, another update is made and it is named Optimus-Gen 2. Optimus robots have been developed to learn and adapt to different situations and environments by integrating machine learning and computer vision technologies. [9].

To develop the Optimus humanoids, the Tesla team developed the DOJO system, which is a scalable solution oriented to machine learning applications. It is based on the D1 custom computing chip, which brings together 354 independent processors, resulting in 362 TFLOPS of computation and 440 MB of internal static random access memory storage. While maintaining full programmability, DOJO emphasizes resource sharing and extremely high bandwidth interconnection, which allows it to scale from small systems to supercomputers. exaFLOP [10].

II. Material And Methods

The methodology, which was followed, included the design of the prototype, its instrumentation, the design of the control system, the development of the algorithm and the assembly of the prototype.

Design of the prototype

Designing the head of a humanoid robot is a complex task that must balance aesthetic, functional, and technological aspects. The head structure must be strong enough to support the internal components, but also light enough not to affect the mobility of the robot. The shape should allow for proper placement of sensors and cameras.

SolidWorks 2021® software was used to design the parts of this project. It should be noted that the parts were obtained with a 3D printer, which means that modifications can be made in a functional, fast, and economic way.

The parts of the humanoid presented in this article are divided into 2 modules:

Head: The head of the humanoid technology should not have a specific shape, but it should have movement and response characteristics similar to those of a human being (1). The inner part is composed of: camera module holder (2), support Raspberry pi 5, Arduino and servomotor connection 2 (5), servomotor support 2 and connector with neck (6), servomotor support 1 (7), as shown in Figure 1.

Neck: It is a mechanism that has two axes of movement, for the rotation and inclination of the mechanism. Consists of neck (3) and connector between the neck and the back (Torso) (4).

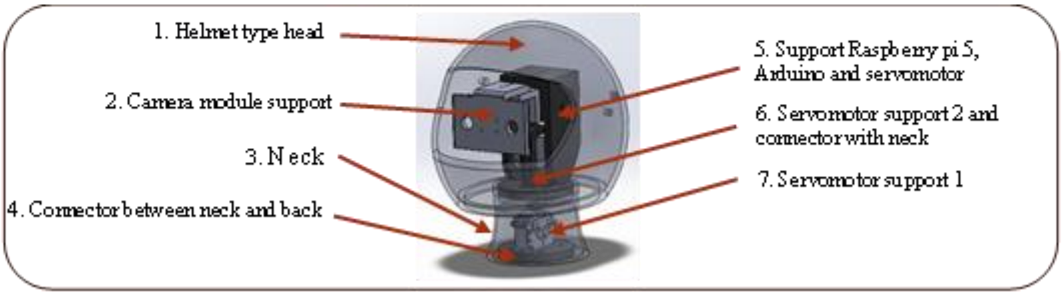

Figure 1. Parts of the humanoid head MIKE

Instrumentation of the prototype

It has two degrees of freedom to perform head movement such as nodding or turning, 2 servomotors are incorporated to allow rotation and tilt of the head. The servomotors used are 6.8V and 40kg-cm, high torque and relatively small size.

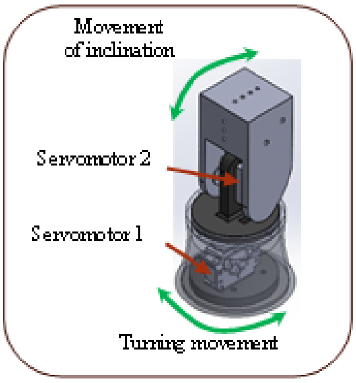

Figure 2. Humanoid head tilt and pan mechanism MIKE

Vision system

To give it the ability to "see" and interpret the world around it, the humanoid has a binocular camera module that has 2 dual 8-megapixel IMX219 stereo cameras with a resolution of 3280 × 2464 pixels per camera, in addition to having an accelerometer, gyroscope and magnetometer incorporated in the same module, giving it stereoscopic vision and depth vision.

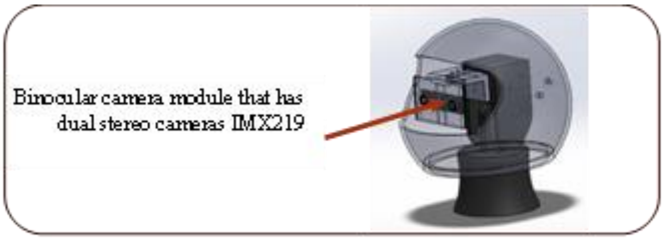

Figure 3. Binocular camera module used for humanoid vision MIKE

Design of the control system

To provide the humanoid with AI to perform the necessary processing for vision, voice recognition and motion control, a "brain" is needed, i.e. in this case a Raspberry Pi 5 minicomputer, It features a 64-bit, quad-core Arm Cortex A76 microcontroller at 2.4 GHz, supports two cameras independently, through which artificial vision is incorporated into the humanoid.; also features WiFi which allows you to control the MIKE robot remotely (from anywhere).

For the turning and tilting movement of the humanoid's head, 2 servomotors are used, the angular control of these is done through an Arduino nano development board, which communicates with the Raspberry Pi 5 through a serial port that is created when a USB is connected between the devices, as shown in Figure 4.

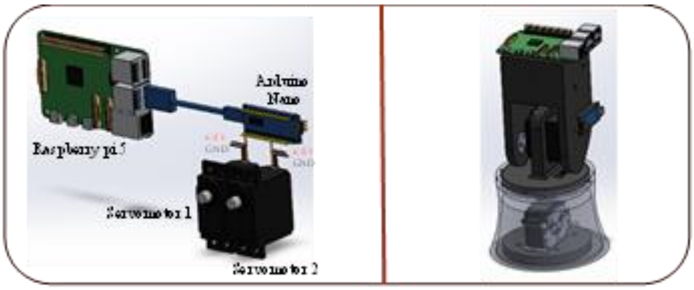

Figure 4. On the left connection Raspberry Pi 5 with Arduino Nano through a serial port and connection with the servomotors, on the right assembly of the system on the humanoid's head.

Development of the algorithm

For the implementation of humanoid vision, facial and/or object recognition algorithms are required to enable robots to identify and classify objects, people and scenarios, which is essential for tasks such as manipulating objects in various environments or navigating the humanoid around obstacles.

As mentioned in section II, a Raspberry Pi 5 minicomputer and a dual stereo camera module are being used, so the Picamera2 library is required to control the cameras, which are Python-based algorithms specifically for the cameras used by the Raspberry Pi. To detect people and/or objects, one camera is used and for stereoscopic and depth vision, both cameras are used.

To implement face recognition, 3 phases are required, face detection, face classifier and face recognition.

Face detection.

There are many face detection algorithms available, what is desired is that the humanoid has high accuracy in face detection, but with low resource consumption, in this work we used algorithms of the well-known OpenCV library [11], which is an open source software, for its simplicity we tested the Haarcascade algorithm, it is worth mentioning that this model requires images with good conditions in the environment, ie, you need to have good lighting, stable background and people have to be as close to the camera; A pre-trained model with deep learning was also used, this has a high detection accuracy and consumes few resources, with this algorithm it is not necessary to have such a controlled environment, i.e., good results are obtained even if there is poor lighting, clutter in the background and the faces of the people are not shown very much in front of the camera.

Face detection is used to create a personalized database of people to recognize, for this it was decided to obtain at least 500 photos and store them in a folder separately per person. The photos can be registered one by one, which is very time consuming, or a video can be taken and from there the respective photos can be obtained directly, for this project we chose the second option, besides the fact that a new person can be updated or added to the database when he/she is in front of the humanoid. The process of capturing the database is shown in the block diagram in Figure 5.

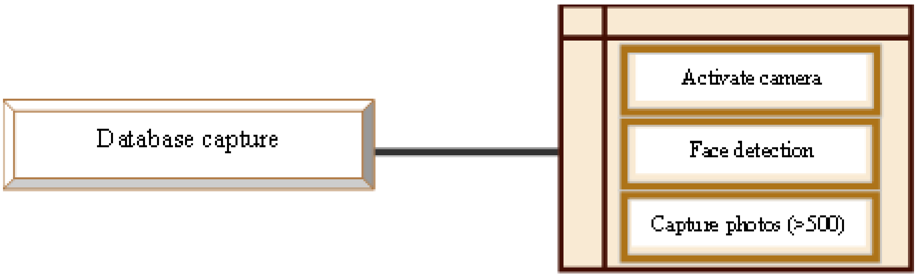

Figure 5. Block diagram of the capture of the customized database of faces to be recognized.

Face sorter.

Face recognition has to be performed with a face classifier, OpenCV has Python-based algorithms to perform this, in this work the following models were tested: EigenFaceRecognizer, FisherFaceRecognizer and LBPHFaceRecognizer, the last one is the one that best developed this function, the model obtained is stored in a file with .xml extension. The process of face classification or trainer of the captured data is shown in the block diagram in Figure 6.

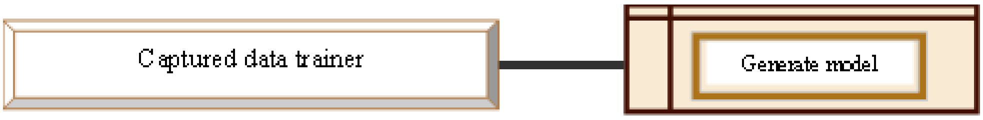

Figure 6. Block diagram for the classification of faces to be recognized.

Face recognition.

The face recognition needs first of all the face detection, for this phase due to the fact that the environment is poorly controlled, the pre-trained deep learning model is used, then it is required to load the classifier to compare the detected face with the ones in the customized database, finally the calculated prediction is given; if a match is not found the face is determined as unknown. The process of recognizing faces is shown in the block diagram in Figure 7.

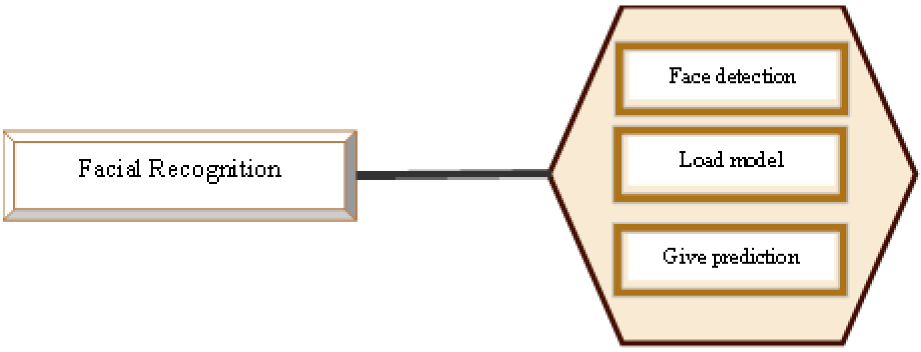

Figure 7. Block diagram of real-time face recognition for MIKE humanoid vision.

The turning and tilting of the humanoid's head is done automatically in 2 ways:
1. When it detects and recognizes faces, the real-time tracking algorithm of people is activated.
2. When it detects objects and obstacles, the algorithm is activated so that the robot can walk freely.

Assembly of the prototype

On the right side of Figure 8, the mechanical design of the head of the MIKE humanoid is shown, in the center you can see the complete assembly of the 3D printed prototype with carbon fiber filament, on the right side you have the head with a mask to protect the internal elements.

Figure 8. Complete prototype of the MIKE humanoid head.

III. Result

Four experimental tests were carried out for 20 individuals, 4 are women and 16 are men, 500 photos with different poses and different characteristics were obtained from each individual, and one test for individuals not found in the database. The tests performed yielded the following results:

Test 1. Night hours with little or no illumination. For this prototype version with the Haarcascade algorithm, the recognition could not be performed for any case, because the camera module does not have infrared or built-in illumination.

For tests 2 to 5, the pre-trained deep learning model was used and the following results were obtained.

Test 2. Night time with good illumination at a distance of 1mt., with the data of the Table. 1.

Table 1: Characteristics of selected data

	covered head	lenses	masks	acknowledgment
1	NO	NO	NO	YES
2	NO	NO	NO	YES
3	NO	NO	NO	YES
4	YES	NO	NO	YES
5	NO	NO	NO	YES
6	YES	NO	NO	YES
7	NO	YES	NO	YES
8	YES	NO	NO	YES
9	NO	NO	NO	YES
10	YES	NO	NO	YES
11	NO	NO	NO	YES
12	NO	NO	NO	YES
13	NO	NO	NO	YES
14	NO	YES	NO	YES
15	YES	NO	NO	YES
16	NO	NO	NO	YES
17	NO	YES	NO	YES
18	NO	NO	YES	YES
19	YES	NO	NO	YES
20	NO	YES	NO	YES

Test 3. In daylight with good illumination at a distance of 1mt, 1 ½ mt. to 2 mts, with the data of the Table 2.

Table 2: Characteristics of selected data

	covered head	lenses	masks	acknowledgment
1	NO	NO	NO	YES
2	NO	NO	NO	YES
3	NO	NO	NO	YES
4	YES	NO	NO	YES
5	NO	NO	NO	YES
6	YES	NO	NO	YES
7	NO	YES	NO	YES
8	YES	NO	NO	YES
9	NO	NO	NO	YES
10	YES	NO	NO	YES
11	NO	NO	NO	YES

12	NO	NO	NO	YES
13	NO	NO	NO	YES
14	NO	YES	NO	YES
15	YES	NO	NO	YES
16	NO	NO	NO	YES
17	NO	YES	NO	YES
18	NO	NO	YES	YES
19	YES	NO	NO	YES
20	NO	YES	NO	YES

Test 4. In daylight with good illumination at a distance of 1 ½ to 2 meters, with multiple individuals with the data in the Table 3.

Table 3: Characteristics of selected data

	covered head	lenses	masks	acknowledgment
1	NO	NO	NO	YES
2	NO	NO	NO	YES
3	NO	NO	NO	YES
4	YES	NO	NO	YES
5	NO	NO	NO	YES
6	YES	NO	NO	YES
7	NO	YES	NO	YES
8	YES	NO	NO	YES
9	NO	NO	NO	YES
10	YES	NO	NO	YES
11	NO	NO	NO	YES
12	NO	NO	NO	YES
13	NO	NO	NO	YES
14	NO	YES	NO	YES
15	YES	NO	NO	YES
16	NO	NO	NO	YES
17	NO	YES	NO	YES
18	NO	NO	YES	YES
19	YES	NO	NO	YES
20	NO	YES	NO	YES

Test 5. In daylight with good illumination at a distance of 1 ½ to 2 meters, with one or multiple individuals not found in the custom database model: yields false positives in 5% of cases, i.e. try to find matching faces.

IV. Conclusions

1. An original prototype of the humanoid's head was built. MIKE.
2. The binocular camera module employed fulfills the vision function of the humanoid.
3. A customized database of faces to be recognized was obtained.
4. The Raspberry pi 5 minicomputer used complies with the features for real-time face recognition, for tracking recognized faces from the customized database, as well as for object and obstacle detection.
5. Due to the fact that in this version of the humanoid no external illumination was incorporated and 5% of false positive cases were obtained, the developed algorithms of the AI system of the MIKE humanoid perform facial recognition with a percentage of 95% of the cases.

References

[1] IBM, ¿Qué Es La Visión Artificial?, Https://Www.Ibm.Com/Mx-Es/Topics/Computer-Vision, Retrieved On August 21, 2024.
Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki And K. Fujimura, "The Intelligent ASIMO: System Overview And Integration," IEEE/RSJ International Conference On Intelligent Robots And Systems, Lausanne, Switzerland, 2002, Pp. 2478-2483 Vol.3, Doi: 10.1109/IRDS.2002.1041641.

[3] K. Akachi Et Al., "Development Of Humanoid Robot HRP-3P," 5th IEEE-RAS International Conference On Humanoid Robots, 2005., Tsukuba, Japan, 2005, Pp. 50-55, Doi: 10.1109/ICHR.2005.1573544.

[4] T. Asfour, K. Welke, P. Azad, A. Ude And R. Dillmann, "The Karlsruhe Humanoid Head," Humanoids 2008 - 8th IEEE-RAS International Conference On Humanoid Robots, Daejeon, Korea (South), 2008, Pp. 447-453, Doi: 10.1109/ICHR.2008.4755993

[5] J. Hirth, N. Schmitz And K. Berns, "Emotional Architecture For The Humanoid Robot Head ROMAN," Proceedings 2007 IEEE International Conference On Robotics And Automation, Rome, Italy, 2007, Pp. 2150-2155, Doi: 10.1109/ROBOT.2007.363639.

[6] Lizhong Gu And Jianbo Su, "Gaze Control On Humanoid Robot Head," 2006 6th World Congress On Intelligent Control And Automation, Dalian, China, 2006, Pp. 9144-9148, Doi: 10.1109/WCICA.2006.1713769.

[7] C. Laschi, H. Miwa, A. Takanishi, E. Guglielmelli And P. Dario, "Visuo-Motor Coordination Of A Humanoid Robot Head With Human-Like Vision In Face Tracking," 2003 IEEE International Conference On Robotics And Automation (Cat. No.03CH37422), Taipei, Taiwan, 2003, Pp. 232-237 Vol.1, Doi: 10.1109/ROBOT.2003.1241601.

[8] R. Beira Et Al., "Design Of The Robot-Cub (Icub) Head," Proceedings 2006 IEEE International Conference On Robotics And Automation, 2006. ICRA 2006., Orlando, FL, USA, 2006, Pp. 94-100, Doi: 10.1109/ROBOT.2006.1641167.
[9] Tesla, IA Y Robótica | Tesla México, Https://Www.Tesla.Com/AI , Retrieved On August 21, 2024.
[10] E. Talpes Et Al., "The Microarchitecture Of DOJO, Tesla's Exa-Scale Computer," In IEEE Micro, Vol. 43, No. 3, Pp. 31-39, May-June 2023, Doi: 10.1109/MM.2023.3258906.
[11] Opencv. (2015). Open Source Computer Vision Library.