Human Body Posture Recognition based on Inclination

Neelima Rajawat
nrajawat2022@fau.edu
Florida Atlantic University,
Boca Raton, Florida, USA, 33486

Abstract

Nowadays, almost everywhere object recognition and detection are required and for that there are so many algorithms and methods used. Object detection implementation such as obstacle detection if something is coming in front of the vehicle can be its implantation. In case of human body recognition, the most basic implantation is detecting human bodies and counting the number of humans in the shown image or video. This can be done using pre-saved image or video and even it can be done using a live camera for image/video capture and then detection of the number of humans. If we follow this and try to go into more details, we can detect human bodies and body parts separately. This implementation can be helpful for medical purposes, where minor human body details are required. There are many implementations for this purpose and one of them is using deep learning methods. With the help of deep learning algorithms, image/video processing can be done, and the next step is the detection and recognition of human bodies.

A method for human body posture recognition based on inclination has been proposed. In this method using HAAR Cascade, and HOG Descriptor, human posture is detected. Basically, focusing on angle of inclination, which helps in identifying the correct posture for standing, sitting and lying down. Here, MediaPipe, developed by Google, is the most helpful framework for working with the model.

Keywords — Computer vision, HAAR, HOG, Mediapipe, Inclination, SVM

1. INTRODUCTION

Though Human Body Recognition is one of the harder concepts of artificial intelligence to implement, it is also very helpful in numerous fields such as in sports, healthcare, autonomous vehicles, e-commerce, Security, and surveillance, and many other applications. This recognition of the human body can be implemented using computer vision algorithms, and it is part of artificial intelligence. Subdivision of computer vision can be understood in Artificial Intelligence with the help of figure 1.

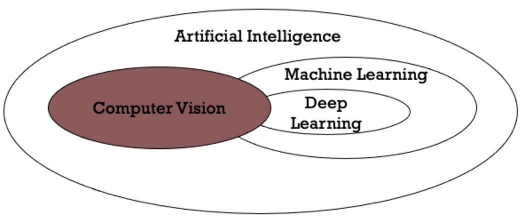


Figure. 1

Basic goal of AI research is to make a model which can understand human actions, activities, and intentions appropriately. Suppose a person is working on his exercise at home but the patient is not sure about the accurate angle of inclination while sitting in a posture or standing in some specific posture. In all these situations, it is necessary to know the correct angle of inclination for not making mistakes. Because if patients sit or stand in a wrong position and inclination for a long time then it can start causing problems in our bones, joints and muscles. So, sitting and standing in an appropriate position, and inclination angle is very important. That's why, this research is focused on the angle of inclination of Human Body Posture.

1.1 Computer Vision

Computer vision is a subfield of artificial intelligence which makes computer systems come up with meaning information collected from digital images, videos, and other visual inputs. Based on the gained information, these systems make decisions and give suggestions. We can conclude if AI makes computers think then computer vision makes them see, detect, and recognize the given input.

1.2 Deep Learning

Deep Learning is another subcomponent of machine learning which works on artificial neural networks with representation learning. Deep learning is essentially a neural network with multiple layers and these neural networks try to simulate all similar features like a human brain. It basically learns from a large amount of data.

1.2.1 Neural Networks in Deep Learning

Basically, Neural Networks are the subset of machine learning, which are used for different Deep Learning applications. Neural Network algorithms function the same as the human brain. It is basically a network of artificial neurons, or we can call nodes, which are connected to each other. This can be used for recognizing patterns by processing the data. Neural networks can be used in different applications such as speech recognition, image processing, human body recognition and autonomous robots.

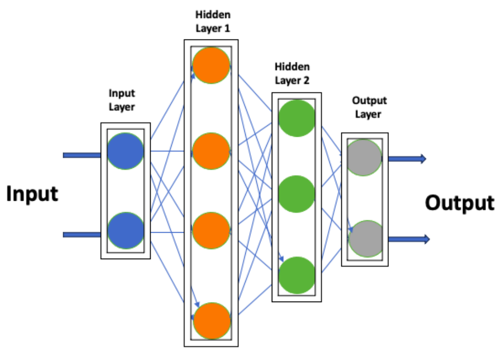


Figure. 2

1.2.2 Types of Neural Networks

There are different types of neural networks, some of them are discussed as follows:

1. **Convolution Neural Network (CNN)**

   CNNs are used for image and video processing with the help of convolution layers for pattern detections.

2. **Feedforward Neural Network (FNN)**

   This is most basic form of neural networks, in which information moves from one direction to other that is from input layer to output layer without any cycles or loops.

3. **Recurrent Neural Network (RNN)**

   This network is used for sequential data. This is used to feature the loops for processing information from previous steps.

4. **Long Short-Term Memory (LSTM)**

   LSTM is one of the specific types of RNN. This is designed for capturing the long-term dependencies in sequential data and for resolving gradient problems.

5. **Radial Basis Function Network (RBFN)**

   For recognizing different types of patterns for classification, RBFN utilizes radial basis function.

1.3 Convolution Neural Network (CNNs)

Most of the computer vision algorithms work using convolution neural networks, as it shows a drastic performance improvement as compared to traditional image and video processing algorithms. CNNs are multilayered architecture which gradually minimize data and its calculation for the most relevant set. Later this set is used for comparison with known data.

1.3.1 Working of CNN

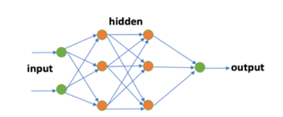
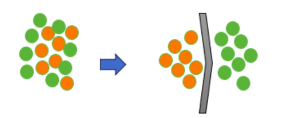
An image is processed in CNN, where colors (green, blue, red) are represented as different matrix values. These matrix values are calculated and converted into 3D tensors for color images. As these are collected feature map stacks, which are tied to the specific part of an image. These tensors are built by taking the image through a series of convolution and pooling layers. This process needs to be repeated based on the number of convolutions layer, but final features are extracted by the convolutional process. These are sent to a fully connected layer, which is used for generating predictions.

1.4 Support Vector Machines (SVMs)

SVMs are a sub-class of supervised learning algorithms. These algorithms are used for different tasks such as classification and regression. SVMs used to work with traditional machine learning but now the combination of SMVs is used with deep learning techniques for specific tasks. With deep learning, SVMs are used to feature extractions and classification. Deep Neural Network is used as a feature extractor and in that SVM is used for final classification based on these features.

SVMs are even used in computer vision for object detection by using a combination of gradient computations techniques in computer vision using linear SVM technique.

Major Differences

CNNs	SVMs
CNNs are deep learning models, which are used for grid data types such as images	SVMs are traditional supervised learning algorithms, which are used for classification and regression
It learns hierarchical features from the given raw datasets	It basically requires manual selection of feature
It basically requires a large amount of labeled data	It can even perform well on smaller datasets
CNNs can be less interpretable due to their complex structure	SVMs can make clear decision boundaries by aiding interpretability
It has complex architecture with convolution and pooling layers	SVMs has simpler architecture, and these models are effective for less complex problems
CNNs are well suited for grid and image data tasks	SVMs are versatile and these models are applicable to various classification tasks
	

1.5 Human body recognition

Human body recognition is a way of interpreting the actions, activities, and postures of human bodies with the help of different artificial intelligence algorithms, especially by using computer vision methods which is one of the components of artificial intelligence. For Understanding human body recognition, a method is required to detect human bodies, their faces, facial expressions, actions, and activities. There are two different parts of human body recognition, as shown in Figure.3, which can be understood with the following details:

1. Human Activity Recognition
2. Human Body Posture Recognition

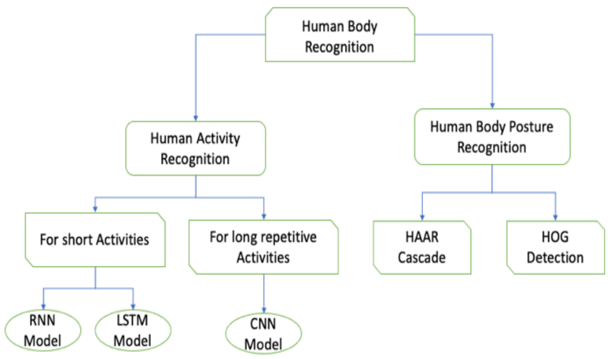


Figure. 3

1.5.1 Human Activity Recognition

Human Activity Recognition (HAR) is one of the methods from human body recognition which interprets human activities or motions with the help of computer vision and machine learning algorithms. With the help of computer vision techniques, human action recognition systems can be developed with the capabilities of automatically recognizing actions and movements of human bodies based on the collected data. These different human actions can be interpreted as activities, gestures, behaviors, and emotions.

Human activity Recognition can be implemented for short activities or for long activities. In case of short activity recognition, deep learning models such as recurrent neural network models, and long short-term memory models can be used as these models follow the natural time order relationships for actual input data. For long repetitive activities, the convolution neural network model is best to use as these CNN models are more efficient to learn features deeply which contain recursive patterns.

1.5.2 Human Body Posture Recognition

Human body posture recognition is a method for interpreting and analyzing human body posture based on their positions, orientation and inclination shown in input data of images or videos. Computer vision is used to execute image and video processing of human body postures based on estimations with the help of specific key points shown in the input image or video. Human Body Posture Recognition is helpful in numerous fields such as healthcare, sports analysis, security, surveillance, and human computer interaction.

To understand the human body posture recognition process, we need to go through the below steps of the process one by one in detail:

Step1: Image Acquisition
For the input, try to get image or video of human body postures.

Step2: Human Detection
Use methods for detecting and locating human bodies.
Ex: HAAR Cascade, HOG Descriptor

Step3: Key Point Detection
Detect critical points such as pose estimation or joint detections.

Step4: Pose Estimation
Estimate the body posture by connecting all the key points.

Step5: Posture Recognition
Analyze the estimated pose to recognize the specific posture or gesture.

Here, we will be focusing on two human body posture recognition methods, which are discussed further:

[1]. HAAR Cascades

For human body posture recognition, HAAR cascade is a method for detecting human body postures. HAAR methods are used for posture detections, and it works by using a simple rectangle filter that has HAAR-like features for scanning an image at different scales. HAAR cascades have been efficient for real-time object detection. However, they might have issues such as variations in scale, pose, and lighting. Below Figure. 4 shows the complete process of HAAR Cascade for human body recognition.

Working of HAAR Cascade

1. **Input Image**
Input images for human body posture recognition using HAAR cascade method has been given.

2. **Image Normalization**
Image normalization is the process of transforming the pixel values of input images into standard or normalized format. In HAAR cascade, it is used to assure that input images that have consistent and comparable characteristics can be converted to grayscale, resize or crop the images.

3. **Create Image Window**
Creating an image window is the process of sliding a window of fixed size across an input image at various positions and scales. This is the most basic step of human body recognition in HAAR cascade.

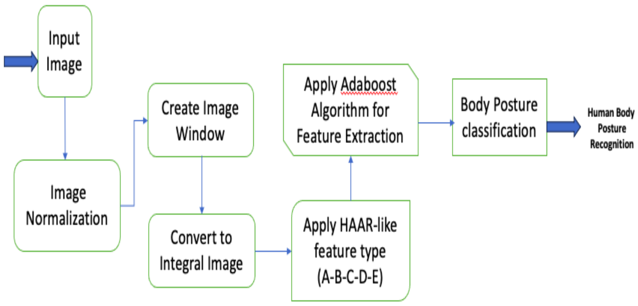


Figure. 4 HAAR Cascade

1. **Convert to Integral Image**
Integral image is used for optimization in HAAR-like feature based human body posture detection methods. This gives calculation of rectangular features at different positions within the image. It makes the posture detection process computationally more efficient. For a grayscale input image, the integral image is constructed by computing the cumulative sum of pixel values in the original image.

S(x,y)=∑i=0x∑j=0yI(i,j)S(x,y)=∑i=0x∑j=0yI(i,j)

Integral image is used for rapid calculation of pixel value sum within the rectangular region, which is created by four values from the integral image.

Sum=S(D)+S(A)−S(B)−S(C)

In Python with OpenCV, we can compute the integral image using the **cv2.integral** function.

2. **Apply AdaBoost Algorithm for feature extraction.**
Basically, AdaBoost is used for boosting by combining multiple weak classifiers to create a strong classifier. In the HAAR cascade method for human body posture recognition, AdaBoost is deployed during training to select and weight the most effective distinguishing features from different human body posture dataset.

3. **Apply HAAR like feature type (A-B-C-D-E)**
HAAR like features are the rectangular filters representing local patterns in images. Here, Type A, B, C, D, E capture variations in horizontal, vertical, diagonal, and line intensity. In human body posture recognition, these features are applied at different scales using the previously discussed sliding window approach. This classifier's decision is based on the combination of features and learned parameters.

4. **Body Posture Classification**
Once the HAAR cascade classifier is identifying regions in an image that contains human bodies then human body postures can be identified. So, different human body postures can be separated using this method.

[2]. Histogram of Oriented Gradients (HOG) Descriptor

In human body posture recognition, HOG descriptor is used for body posture recognition. HOG works by capturing the gradients and their orientation for an image. HOG Descriptor works for the local image appearance and shape of the image can be described based on the distribution of gradients intensity and even it can be measured based on edge directions. In this process, the image is divided into small, connected regions, which are known as cells and for calculating every cell's pixel, histogram of gradient directions is compiled. Basically, a descriptor is the concatenation of these histograms. Figure. 5 is used for describing the complete process of the HOG descriptor.

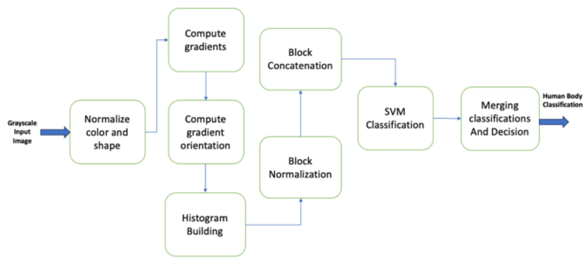


Figure. 5 HOG Descriptor

Working of HOG Descriptor

HOG descriptors are effective for capturing shape and edge information in images, making them suitable for human body posture recognition. They are commonly used in combination with SVMs for accurate and efficient object detection in various computer vision applications.

1. **Input Image**
Input images for human body posture recognition using the HOG Descriptor method has been given.

2. **Normalize color and shape**
Normalizing color and shape of an image, using Histogram of oriented gradients (HOG) descriptors, is a way of enhancing the robustness of the feature extraction process. Normalizing input images provides us with variations in lighting conditions and contrast of different images.

3. **Compute Gradients**
At this step, it is required to compute the gradient of pixel intensities in both the x and y directions. This provides information about the local intensity changes in the image.

4. **Compute Gradient Orientation**
Compute the magnitude and orientation of the gradients for each pixel. The magnitude can be calculated by $\sqrt{(grad_x)\text{^}2 + (grad_y)\text{^}2}$. While the orientation can be calculated by arctan2($grad_y, grad_x$).

5. **Histogram Building**
Based on the gradient, it will try to build the histogram of the shown images. So far, building histogram is one of the important parts of this process.

6. **Block Normalization**
The cells are grouped into larger blocks, and the histogram values are normalized within each block. This normalization helps make the descriptor more robust to changes in lighting conditions.

7. **Block Concatenation**
Finally, the HOG descriptor is created by concatenating the normalized histograms from the different cells and blocks as well.

8. **SVM Classification**
Basically, HOG Descriptor is used as a machine learning classifier which is a support vector machine for the purpose of training. These classifications are done on the different inclination classes, which are set based on the calculated angles in between different postures of human bodies.

9. **Merging classifications and decision**
At the end, based on the classification, these posture decisions have been merged and the final decision of image posture has been made as to which class that posture should be chosen, where it fits.

Major Differences

HOG Descriptor	HAAR Cascade
It is based on gradient information in local regions, which captures gradient information and orientation	it is based on simple rectangular filters, which basically utilizes intensity differences in adjacent regions
It involves gradient magnitude and orientation computations	It requires computations of sums and differences of pixel intensities
It is efficient, but sometime slower than HAAR like features in some cases	It has fast computation due to integral image representation
It explicitly captures gradient orientation in different image region	It does not explicitly capture orientation information
It may require more memory as it stores gradient histograms	It typically requires less memory
Application: Object Detection, Human body detection, general image classification	**Applications:** Face Detection, Object detection, human body recognition

1.6 Applications

Human body posture recognition method has numerous applications in different fields, some the applications are discussed below as follows:

1. **Healthcare & Physiotherapy**

   Human Body Posture Recognition applications are helpful in the healthcare and physiotherapy process. In the field of medicine and healthcare, for monitoring patient's movements, applications of human body posture recognition can be helpful. Even in different physiotherapy exercises, while doing at home, it can monitor and guide for the appropriate exercise and postures.

2. **Sports & Fitness Analysis**

   In different sports, movements can be analyzed using the human body posture recognition method, by using these techniques, it is helpful to understand and closely monitor the poses while playing different sports. It can monitor and provide feedback on the body postures during fitness activities to ensure proper form and reduce the risk of injuries.

3. **Security and Surveillance**

   In the field of security and surveillance systems, Human body recognition is one of the valuable applications, which is using different computer vision techniques for analyzing the activities, positions, and movements of human body postures in real time. Basically, this method can be applied for enhancing the security by detecting and reposing while seeing some suspicious activities or poses which look threatful. It can help in intrusion detection and crowd monitoring.

4. **Human Computer Interaction**

   In Human Computer Interaction, human body posture recognition can be implemented for two applications such as in Gesture Recognition for enabling intuitive interactions with computer devices based on body gestures and postures. Augmented Reality (AR) and Virtual Reality (VR) by tracking Human body posture recognition helps in enhancing AR-VR experience.

5. **Educational Training**

   By creating a virtual learning environment, try to create integrative educational content where body posture is important in simulations and learning. Human body posture recognition can be used to train the students in professional fields such as in, military or law enforcement training.

6. **Gaming Industry**

   In Gaming industry, for motion capture by capturing and replicating with real world body movements in virtual environments, gaming experience can be enhanced. Even human body posture recognition is used to create a more immersive and responsive gaming experience.

2. Literature Review

2.1 Deep Learning enabled medical computer vision

This research focuses on the transformative impact of different deep learning techniques in medicine using computer vision applications. This includes the in-depth examination of different architectures using deep learning models such as convolutional neural networks (CNNs), recurrent neural networks (RNNs). Author discussed the utilization of huge medical datasets for training deep learning models, where it focuses on quality and diversity. This survey highlights use of different deep learning and computer vision techniques in medical imaging, pathologies, cardiology, dermatology, ophthalmology by learning in depth on human activities with ambient intelligence.

2.2 Research on Human Body Detection and Tracking Algorithm Based on Kinect

This research is focused on advancement in human body detection and tracking methods using Kinect Technology. The key points work on the exploration of Kinect's capabilities in capturing depth information and its integration with sophisticated algorithms for robust human body detection and tracing. Using a 3D motion sensing camera, this research first analyzes the human body, then using the SVM classifier, it tries to work on a tracking algorithm and its implementation with the HOG feature detection method.

2.3 Human activity recognition in artificial intelligence framework: a narrative review

This survey provides the overview of state-of-the-art techniques and advancements in human activity recognition in AI. HAR worked as collecting activity signals, pre-processing, model training, and activity interface. This paper discussed the HAR process, different devices, and application AI framework. Overall, it represents current implementation using HAR, and talks about challenges and future directions.

2.4 BEHAVE: Dataset and Method for Tracking Human Object Interactions

This research discussed modeling interactions between human and object in different applications i.e. gaming, virtual reality, and human-robot collaboration. To highlight the challenges of this, authors introduce the BEHAVE dataset, that is full body human object interaction dataset. This dataset contains multi-view RGB D frames, 3D SMPL fits. This dataset comprises 15k frames over 5 different locations with 8 subjects interacting with 20 common objects. Main approach is to predict the human and object interaction into a statistical model.

2.5 Face mask recognition system using CNN Model

This paper talked about need of face mask detection during pandemic. So, this has been done using basic machine learning tools such as Tensor Flow, Keras, Scikit-Learn, and OpenCV. Proposed method can identify faces in shown images and videos and finds the presence of masks. It gets an excellent accuracy in case of surveillances and when investigates optimal values for convolution Neural Network (CNN)model to ensure accurate mask detection without overfitting.

3.Methodology

Using a computer vision algorithm with a combination of deep learning approach and support vector machine, a two-way method for human body posture recognition has been implemented. Figure 6 shows our proposed solution model for human body posture recognition.

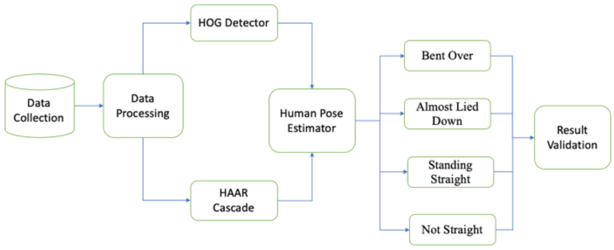


Figure. 6 Proposed Model

3.1 ALGORITHM

Let's understand the implementation process by the algorithms.

- **Step1: Data Collection**
  First step is to collect the dataset in the format of images or videos of humans/ Human bodies.

- **Step2: Data Processing**
  Here, we are working on two things:
  - Resizing dataset for model input
  - Normalize the pixel values in common ranges for all the images.

- **Step3: Select Model for Recognition**
  Here, we have given two options for human body recognition, which are as
  1. HOG Detector
  2. HAAR Cascade

- **Step4: Choosing Input Method**
  In this step, we have two options:
  Option -1: Live Camera
  Option -2: Input Dataset

- **Step5: Human Pose Estimation**
  In this step, we train the pose estimate model for the input dataset. This model is used to predict the human body posture based on the inclination:

  Posture = ["Bent Over", "Almost Lied Down", "Standing Straight", "Not Straight"]

- **Step6: Post-Processing**
  In this step, we refine the pose estimation results by filtering out based on confusion matrix values false positive, false negative, true negative, and true negative. Focusing on noisy key points by looking closely on false positive values.

- **Step7: Visualization of Detected Bodies**
  Visualize the detected human body postures after recognition on the shown original image or video.

- **Step8: Validation of Results**
  In this step, we recognize the result's accuracy using appropriate matrices such as union of human body recognition, Test accuracy for pose estimation.

Most focused part is how to calculate the angle of inclination, so for showing that in details, prefer to use below table:

S.N.	Postures	Left Shoulder (s), Left Hip(h), Left Knee(k)
1	Bent Over	When s is Greater than h, & Greater than k
2	Almost Lied Down	When s is Greater than h, & less than k
3	Standing straight	When s is Less than h and less than k
4	Not Straight	When s is less than h, and greater than k

4.Implementation

For implementing the above method for human body posture recognition based on inclination angle, we have used python HAAR Cascade and HOG Descriptor methods. This is implemented using computer vision & Python. Using Python cv2 packages are imported for all computer vision libraries.

4.1 Input

For image/video input we have used two options to choose in the code:

- This code can perform live camera, human body posture recognition, so for that a system with a working camera is required, if live camera method is being used.
- Another method is just to pass an image/video dataset for further analysis. So almost 1200+ different posture images are used as input.

4.2 Method Discussion

After taking input, data has been processed and for further actions, added two options either you can go with the HAAR Cascade or HOG Descriptor. Based on the chosen model, human body posture is estimated. This posture is recognized based on the angle of inclination, which is implemented based on angle of knee, hip and shoulder.

4.3 Tools & Techs

A computer system with more than 8GB RAM and camera capability is used. For writing code, we have to use PyCharm tool. Here, Python version 3.11.2 has been used on a mac machine. In the PyCharm tool, after installing Jupyter notebook package, using the below command:

Python –m notebook

Used different computer vision packages i.e. **import cv2.**

MediaPipe

MediaPipe is a framework developed by google. It is used to make different machine learning models such as face detection, human body recognition, multi-hand tracking, object detection, and just for normal tracking. Basically, MediaPipe is a cross-platform, and open-source machine learning framework. So, for this implantation, imported this package as:

Import mediapipe as mp

4.4 Debug Output

After running the code, it will start with this point as shown in below figure 7 of output.

```
Select the recognition model:
1. HOG Detector
2. HAAR Cascades
Enter your choice (1/2):
 1
Choose an option:
Option-1: Live Camera
Option-2: Input Dataset
Enter your choice (1/2):
  2
Image: pose369.jpg, Posture: Standing Straight
Image: pose433.jpg, Posture: Standing Straight
Image: pose355.jpg, Posture: Standing Straight
Image: pose341.jpg, Posture: Standing Straight
Image: pose88.jpg, Posture: Bent Over
Image: pose427.jpg, Posture: Standing Straight
Image: pose63.jpg, Posture: Standing Straight
```

Figure. 7 Output

5.Results Analysis

By considering different posture image data, using two different models HAAR Cascade and HOG Descriptor, we can see the differences between recognized posture images.

For HOG Descriptor, it shows 97.97% Accuracy, as 114 images are almost lied down and it shows correct, but for one image bent down. These are the false prediction (FP, FN). While for 'bent over' postures, for 3 images it shows 'almost lied down'.

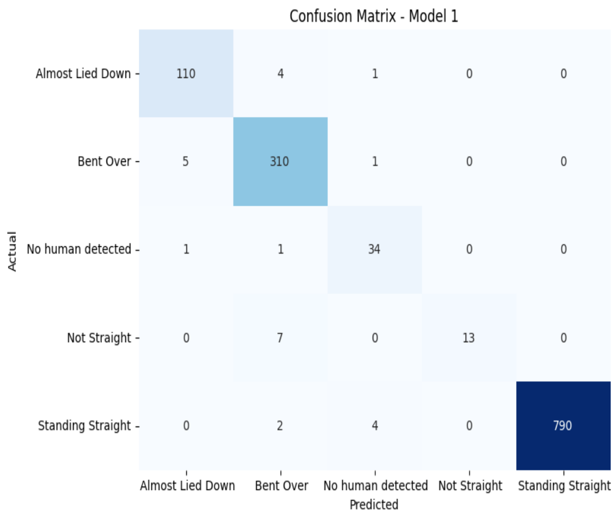


Figure. 8 Using HOG Descriptor

As a result, for this model, overall, it shows 97.97% accuracy.

```
Model: 1
Total Images: 1283
Correct Predictions: 1257
Accuracy: 97.97%
```

For HAAR Cascade, it shows 100% accuracy. So, it shows all the true positive values.

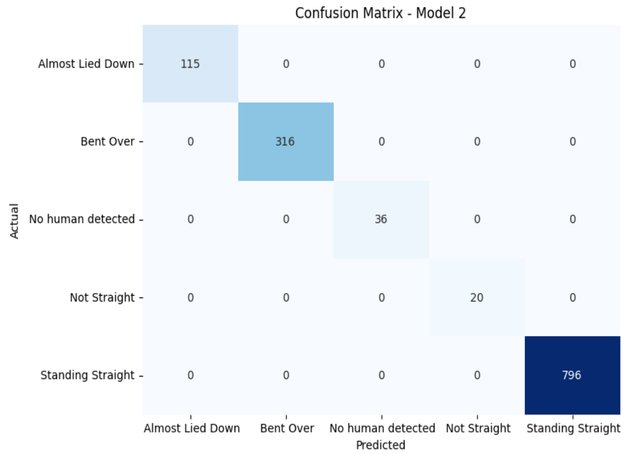


Figure. 9 Using HAAR Cascade

As it can be seen by using model 2, for 1283 postures it recognizes the correct posture. Therefore, it gives 100% accurate recognition results.

```
Model: 2
Total Images: 1283
Correct Predictions: 1283
Accuracy: 100.00%
```

Performance Analysis

Based on the model comparison, performance is analyzed for both the models HOG Descriptor and HAAR Cascade for human body posture recognition. Performance analysis is done using different parameters, such as accuracy, recall, precision, and f1-score. Accuracy is the correctly classified instanced in total instances. Recall is positive instances correctly classified by the model. While Precision is the predicted positive instances which are actual positive. F1-score is used for making a balance between false positives and false negatives. We can understand it by seeing below figure. 10

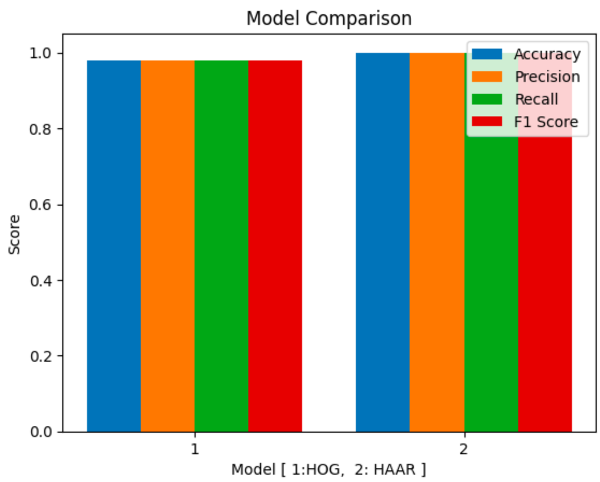


Figure. 10

Below grid plot shows the comparison between HOG Descriptor and HAAR Cascade, for human body posture recognition and this has been done based on inclination. In case of HOG posture recognition accuracy is 97.97% while in case of HAAR, it is 100%. Which can truly be seen in the below figure 11 performance evaluation graph.

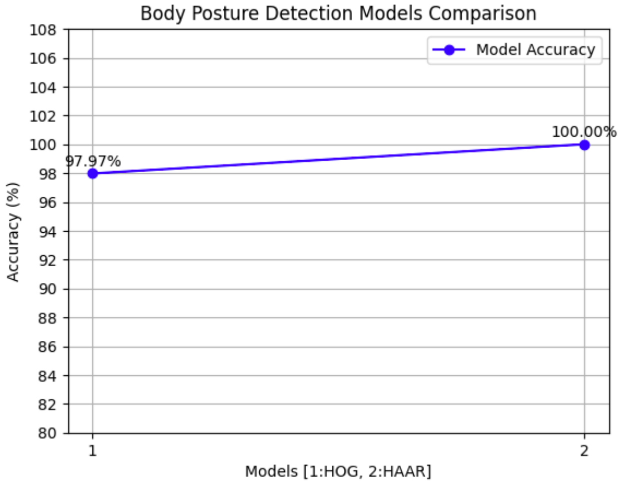


Figure. 11 Performance comparison

6.Conclusion

Based on the results analysis, we can say that the HOG method is better than HAAR as it gives more accurate results. By working on this project for human body recognition based on inclination, we are defining different human body postures. So, I worked on almost 1200+ different poses and tried to find their exact inclination based on the angle. This angle is calculated after recognizing exact human body posture.

By seeing the result, we can analyze that if we are using HOG descriptor then accuracy of the detected image posture is 99.53%, while in case of HAAR Cascade, we can be 100% sure about the body posture. So, if it is required to implement at a place where highly reliable results are only accepted then it is suggested to use HAAR Cascade instead of HOG Descriptor. Although HOG is almost good enough to implement at numerous places.

This implementation can help in proper standing or sitting or determining the exact human body posture. In many places, it can help to predict and predefine the postures and suggest a better way of specifically showing accurately. This can be a great application in different fields, such as sports, medicine, and healthcare.

7.Future Scope

In Future, we can even go in more depth for finding different postures based on other factors, which can help in medical, Healthline, wellness, sports, and so on. We can even work with different types of datasets. Other than angle, and posture, factors such as working on gestures, activities, sentiments, and it can even be something in more detail. We can work with different models, and different techniques. Even enhancement in the current model by going through a large dataset, or by choosing different aspects can be implemented.

REFERENCES

[1] Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., ... & Socher, R. (2021). Deep learning-enabled medical computer vision. *NPJ digital medicine*, *4*(1), 5.

[2] Wu, Q. (2021, December). Research on human body detection and tracking algorithm based on kinect. In *2021 International Conference on Electronic Information Technology and Smart Agriculture (ICEITSA)* (pp. 7-10). IEEE.

[3] Waseem, S., Abu-Bakar, S. A. R. S., Omar, Z., Ahmed, B. A., Baloch, S., & Hafeezallah, A. (2023). Multi-attention-based approach for deepfake face and expression swap detection and localization. *EURASIP Journal on Image and Video Processing*, *2023*(1), 14.

[4] Zhang, J., Yang, Y., Gao, D., Wang, R., & Wang, J. (2023, May). Deep Learning Based Cross-View Human Detection System. In *Journal of Physics: Conference Series* (Vol. 2504, No. 1, p. 012027). IOP Publishing.

[5] Zhu, H., Zheng, Z., & Nevatia, R. (2023). Gait recognition using 3-d human body shape inference. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 909-918).

[6] Daoudi, M., Vezzani, R., Borghi, G., Ferrari, C., Cornia, M., Becattini, F., & Pilzer, A. (2023). Computer Vision in Human Analysis: From Face and Body to Clothes. *Sensors*, *23*(12), 5378.

[7] Al-Faris, M., Chiverton, J., Ndzi, D., & Ahmed, A. I. (2020). A review on computer vision-based methods for human action recognition. *Journal of imaging*, *6*(6), 46.

[8] Cheng, M. M., Zhang, Z., Lin, W. Y., & Torr, P. (2014). BING: Binarized normed gradients for objectness estimation at 300fps. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3286-3293).

[9] Leskovec, J., & Sentjo, S. (1999). Detection of human bodies using computer analysis of a sequence of stereo images. *11th European Union Contest for Young Scientists*.

[10] Mori, G., & Malik, J. (2002). Estimating human body configurations using shape context matching. In *Computer Vision—ECCV 2002: 7th European Conference on Computer Vision Copenhagen, Denmark, May 28–31, 2002 Proceedings, Part III 7* (pp. 666-680). Springer Berlin Heidelberg.

[11] Dargazany, A. R. (2019). Human Body Parts Tracking: Applications to Activity Recognition. *arXiv preprint arXiv:1907.05281*.

[12] Ma, R., Zhang, Z., & Chen, E. (2021). Human motion gesture recognition based on computer vision. *Complexity*, *2021*, 1-11.

[13] Baker, B., Gupta, O., Naik, N., & Raskar, R. (2016). Designing neural network architectures using reinforcement learning. *arXiv preprint arXiv:1611.02167*.

[14] Bhatnagar, B. L., Xie, X., Petrov, I. A., Sminchisescu, C., Theobalt, C., & Pons-Moll, G. (2022). Behave: Dataset and method for tracking human object interactions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 15935-15946).

[15] Mamatkulovich, B. B. (2023). A DESIGN OF SMALL SCALE DEEP CNN MODEL FOR FACIAL EXPRESSION RECOGNITION USING THE LOW-RESOLUTION IMAGE DATASETS. *MODELS AND METHODS FOR INCREASING THE EFFICIENCY OF INNOVATIVE RESEARCH*, *2*(19), 284-288.

[16] Wu, M., Brandhorst, H., Marinescu, M. C., Lopez, J. M., Hlava, M., & Busch, J. (2023). Automated metadata annotation: What is and is not possible with machine learning. *Data Intelligence*, *5*(1), 122-138.

[17] Muneeb, M., Rustam, H., & Jalal, A. (2023, February). Automate appliances via gestures recognition for elderly living assistance. In *2023 4th International Conference on Advancements in Computational Sciences (ICACS)* (pp. 1-6). IEEE.

[18] Li, H., Zeng, N., Wu, P., & Clawson, K. (2022). Cov-Net: A computer-aided diagnosis method for recognizing COVID-19 from chest X-ray images via machine vision. *Expert Systems with Applications*, *207*, 118029.

[19] Xie, Y., Takikawa, T., Saito, S., Litany, O., Yan, S., Khan, N., ... & Sridhar, S. (2022, May). Neural fields in visual computing and beyond. In *Computer Graphics Forum* (Vol. 41, No. 2, pp. 641-676).

[20] Allugunti, V. R. (2022). Breast cancer detection based on thermographic images using machine learning and deep learning algorithms. *International Journal of Engineering in Computer Science*, *4*(1), 49-56.

[21] Kaur, G., Sinha, R., Tiwari, P. K., Yadav, S. K., Pandey, P., Raj, R., ... & Rakhra, M. (2022). Face mask recognition system using CNN model. *Neuroscience Informatics*, *2*(3), 100035.

[22] Mehendale, N. (2020). Facial emotion recognition using convolutional neural networks (FERC). *SN Applied Sciences*, *2*(3), 446.

[23] Singh, J., Aggarwal, R., Tiwari, S., & Joshi, V. (2022, October). Exam Proctoring Classification Using Eye Gaze Detection. In *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)* (pp. 371-376). IEEE.

[24] Tan, T. H., Wu, J. Y., Liu, S. H., & Gochoo, M. (2022). Human activity recognition using an ensemble learning algorithm with smartphone sensor data. *Electronics*, *11*(3), 322.

[25] Frühstück, A., Singh, K. K., Shechtman, E., Mitra, N. J., Wonka, P., & Lu, J. (2022). Insetgan for full-body image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7723-7732).

[26] Li, Y. (2022, January). Research and application of deep learning in image recognition. In *2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA)* (pp. 994-999). IEEE.

[27] Deng, S., Liang, Z., Sun, L., & Jia, K. (2022). Vista: Boosting 3d object detection via dual cross-view spatial attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8448-8457).

[28] Ahmad, K., Maabreh, M., Ghaly, M., Khan, K., Qadir, J., & Al-Fuqaha, A. (2022). Developing future human-centered smart cities: Critical analysis of smart city security, Data management, and Ethical challenges. *Computer Science Review*, *43*, 100452.

[29] Guo, C., Zou, S., Zuo, X., Wang, S., Ji, W., Li, X., & Cheng, L. (2022). Generating diverse and natural 3d human motions from text. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5152-5161).

[30] Huang, C. H. P., Yi, H., Höschle, M., Safroshkin, M., Alexiadis, T., Polikovsky, S., ... & Black, M. J. (2022). Capturing and inferring dense full-body human-scene contact. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13274-13285).

[31] Islam, M. M., Nooruddin, S., Karray, F., & Muhammad, G. (2022). Human activity recognition using tools of convolutional neural networks: A state of the art review, data sets, challenges, and future prospects. *Computers in Biology and Medicine*, 106060.

[32] Helmi, A. M., Al-qaness, M. A., Dahou, A., & Abd Elaziz, M. (2023). Human activity recognition using marine predators algorithm with deep learning. *Future Generation Computer Systems*, *142*, 340-350.

[33] Zhang, S., Li, Y., Zhang, S., Shahabi, F., Xia, S., Deng, Y., & Alshurafa, N. (2022). Deep learning in human activity recognition with wearable sensors: A review on advances. *Sensors*, *22*(4), 1476.

[34] Sahoo, J. P., Prakash, A. J., Pławiak, P., & Samantray, S. (2022). Real-time hand gesture recognition using fine-tuned convolutional neural network. *Sensors*, *22*(3), 706.

[35] Khan, I. U., Afzal, S., & Lee, J. W. (2022). Human activity recognition via hybrid deep learning based model. *Sensors*, *22*(1), 323.

[36] Basak, H., Kundu, R., Singh, P. K., Ijaz, M. F., Woźniak, M., & Sarkar, R. (2022). A union of deep learning and swarm-based optimization for 3D human action recognition. *Scientific Reports*, *12*(1), 5494.

[37] Jiao, C. (2022). Recognition of Human Body Feature Changes in Sports Health Based on Deep Learning. *Computational and Mathematical Methods in Medicine*, *2022*.

[38] Ullah, W., Hussain, T., Khan, Z. A., Haroon, U., & Baik, S. W. (2022). Intelligent dual stream CNN and echo state network for anomaly detection. *Knowledge-Based Systems*, *253*, 109456.

[39] Rastgoo, R., Kiani, K., & Escalera, S. (2022). Real-time isolated hand sign language recognition using deep networks and SVD. *Journal of Ambient Intelligence and Humanized Computing*, *13*(1), 591-611.

[40] Kong, Y., & Fu, Y. (2022). Human action recognition and prediction: A survey. *International Journal of Computer Vision*, *130*(5), 1366-1401.

[41] Sanchez-Caballero, A., de López-Diz, S., Fuentes-Jimenez, D., Losada-Gutiérrez, C., Marrón-Romera, M., Casillas-Perez, D., & Sarker, M. I. (2022). 3dfcnn: Real-time action recognition using 3d deep neural networks with raw depth information. *Multimedia Tools and Applications*, *81*(17), 24119-24143.

[42] Sahin, V. H., Oztel, I., & Yolcu Oztel, G. (2022). Human monkeypox classification from skin lesion images with deep pre-trained network using mobile application. *Journal of Medical Systems*, *46*(11), 79.

[43] Gupta, N., Gupta, S. K., Pathak, R. K., Jain, V., Rashidi, P., & Suri, J. S. (2022). Human activity recognition in artificial intelligence framework: A narrative review. *Artificial intelligence review*, *55*(6), 4755-4808.

[44] Le, N., Rathour, V. S., Yamazaki, K., Luu, K., & Savvides, M. (2022). Deep reinforcement learning in computer vision: a comprehensive survey. *Artificial Intelligence Review*, 1-87.

[45] Houssein, E. H., Hammad, A., & Ali, A. A. (2022). Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review. *Neural Computing and Applications*, *34*(15), 12527-12557.

[46] Baker, B., Gupta, O., Naik, N., & Raskar, R. (2016). Designing neural network architectures using reinforcement learning. *arXiv preprint arXiv:1611.02167*.

[47] Islam, A. R. (2022). Machine learning in computer vision. In *Applications of Machine Learning and Artificial Intelligence in Education* (pp. 48-72). IGI Global.

[48] Mehmood, S., Ghazal, T. M., Khan, M. A., Zubair, M., Naseem, M. T., Faiz, T., & Ahmad, M. (2022). Malignancy detection in lung and colon histopathology images using transfer learning with class selective image processing. *IEEE Access*, *10*, 25657-25668.

[49] Jiang, B., Hong, Y., Bao, H., & Zhang, J. (2022). Selfrecon: Self reconstruction your digital avatar from monocular video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5605-5615).

[50] Zhao, F., Yang, W., Zhang, J., Lin, P., Zhang, Y., Yu, J., & Xu, L. (2022). Humannerf: Efficiently generated human radiance field from sparse inputs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7743-7753).

[51] Yi, X., Zhou, Y., Habermann, M., Shimada, S., Golyanik, V., Theobalt, C., & Xu, F. (2022). Physical inertial poser (pip): Physics-aware real-time human motion tracking from sparse inertial sensors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13167-13178).

[52] Yan, X., Lin, L., Mitra, N. J., Lischinski, D., Cohen-Or, D., & Huang, H. (2022). Shapeformer: Transformer-based shape completion via sparse representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 6239-6249).

[53] Parvaiz, A., Khalid, M. A., Zafar, R., Ameer, H., Ali, M., & Fraz, M. M. (2023). Vision Transformers in medical computer vision—A contemplative retrospection. *Engineering Applications of Artificial Intelligence*, *122*, 106126.

[54] Suman, S., Mishra, S., Sahoo, K. S., & Nayyar, A. (2022). Vision navigator: A smart and intelligent obstacle recognition model for visually impaired users. *Mobile Information Systems*, *2022*.

[55] Dulhare, U. N., & Ali, M. H. (2023). Underwater human detection using faster R-CNN with data augmentation. *Materials Today: Proceedings*, *80*, 1940-1945.

[56] Tang, Y., Yang, D., Li, W., Roth, H. R., Landman, B., Xu, D., ... & Hatamizadeh, A. (2022). Self-supervised pre-training of swin transformers for 3d medical image analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 20730-20740).

[57] Sun, P., Cao, J., Jiang, Y., Yuan, Z., Bai, S., Kitani, K., & Luo, P. (2022). Dancetrack: Multi-object tracking in uniform appearance and diverse motion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 20993-21002).

[58] He, Y., Zhang, H., Arens, E., Merritt, A., Huizenga, C., Levinson, R., ... & Alvarez-Suarez, A. (2023). Smart detection of indoor occupant thermal state via infrared thermography, computer vision, and machine learning. *Building and Environment*, *228*, 109811.

[59] Pham, H. H., Khoudour, L., Crouzil, A., Zegers, P., & Velastin, S. A. (2022). Video-based human action recognition using deep learning: a review. *arXiv preprint arXiv:2208.03775*.