IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-5, May 2019 E-ISSN: 2347-2693

Human Pose Detection From a Digital Image With Machine Learning Technique

Munindra kakati[1*], Parismita Sarma[2]

[1] Department of Information technology, Gauhati University, Guwahati, India
[2] Department of Information Technology, Gauhati University, Guwahati, India

[*]*Corresponding Author: kakatimunindra@gmail.com, Tel.: 917002339857*

DOI: https://doi.org/10.26438/ijcse/v7i5.13891393 | Available online at: www.ijcseonline.org

Accepted: 25/May/2019, Published: 31/May/2019

Abstract— Continuous development of Information Technology in the field of computer vision and digital image processing has brought new hope to many unfold problems of our daily life. In digital image area problems are solved by representing image as set of pixels with different amplitudes. These pixels are sent as input to the computer system and the system process these values using different algorithms. In this paper, we are working on a model, which is capable of automatically recognizing sitting and standing poses of human from digital photographs. We are also able to tag the respective images with correct pose. In this paper, we have shown standing and sitting poses which are executed using Matlab machine learning technique.

Keywords—Computer vision, Image processing, Human pose, HOG

I. INTRODUCTION

Human generally can give many poses with the help of their hand, finger, leg, foot etc. Human pose is considered to be one of the important research field in computer vision and a lot of research has been done on it. This topic has higher demand in the field of human computer interaction as well as activity recognition. The main focus of this problem is to automatic detection of different body parts and localization of the vulnerable joints from a digital image. Human pose detection is thought to be the first stage of human behavior understanding. Some systems has to strictly rely on the precise position of the human body parts, as it is of interest in their behavior analysis in an automotive environment. In the field of action recognition it is used to detect unusual behavior in different environments like crowds, hospitals, sports event etc. Action Recognition can also be useful in translating human action into animation form.

Problem Statement: Our aim of this project work is to detect two most important human pose standing and sitting using machine learning technique which can imitate human visual structure. For this important human body parts and joints are estimated and their virtual configuration to each other in terms of angle created between them are evaluated.

Image Processing: In this work we are using different digital image processing techniques. Some of the mathematical operations such as addition, subtraction, multiplication and division are represented by some image morphological operations. Among them dilation, erosion, open, close, boundary extraction and fill up are mostly used. Input for these processing may be a series of images or a video clip as series of video frames. The output of image processing may be matrix of numbers related to image parameter or an image itself.

Computer Vision: Computer Vision, in general can be said as the high-level understanding of the numeric representation of the binary 2-d images i.e. digital images. This is achieved through the venture of automating and incorporating wide range of procedures and representations utilized for vision perception. The implementation method is through the identification and passage of information of the processed image using mathematical operations. The output of which can be a set of image characteristics or a single image. The past decade has seen a continuous rise in its applications which has enabled computers with attached camera to intelligently see and recognize images thus bringing justification to the formulated phrase "Computer Vision".

Machine Learning in Computer Vision: Machine learning is embraced in an extensive variety of areas where it demonstrates its predominance over customary lead based calculations[1]. The main purpose is to allow the computer to learn robotically without any other assistance, also it learns to adjust different actions accordingly. Machine Learning is generally used in two ways to train the models. They are:

© 2019, IJCSE All Rights Reserved

1. Supervised Learning: This approach trains model to predict future events based on past known facts, information or data. The learning algorithm after sufficient training compares output with correct intended output to find errors and modify the model accordingly.

2. Unsupervised Learning: It is used to train models based on unclassified or unlabeled information. Its main task is to find patterns or constructions, which may be hidden or natural from the unlabeled data. Unsupervised classification model do not compare points to be classified with training data[2].

Artificial Neural Network: Neural networks are artificial simulations of biological brains. The biological brain consists of a huge number of interconnected elements called neurons. The input layer is the image features in specific format and the output layer is the pose or behavior classification. Based on the power of interconnection to different neurons in the hidden layer, the input signal is forward to the next level layer. The control of the connection is called a weight. The value of neuron in each layer will be based on the weight of the connection and the values of the neurons of the preceding layer. One of the important capacity of the brain is its ability to categorize or differentiate. To construct our artificial visual system we also need the ability to categorize. Our artificial neural network should learn the pose from a set of features value.

II. RELATED WORK

We have done a number of literature review throughout our work. A number of journal and conference papers as shown below are found important for our work.

In 2013, Chaitra B H, Anupama H S and Cauvery N K[3] mentioned in their paper that they have developed a recognition system for human actions using a novel self organizing map based retrieval system.

In 2018, Yi Liu, Ying Xu and Shao-bin Li[4] published a review paper and focuses on the state of art progress of two dimensional human pose detection methods depend on deep learning. In this paper they discussed about single Convolutional neural network method, Multi stage Convolutional neural network method, Multi branch method, Recurrent Neural Network method and Generative Adversarial Networks method.

In 2017, Vasileios Belagiannis and Andrew Zisserman[5] proposed a Convolutional neural network model for estimation of two dimensional human body poses in an digital image. This model discussed about heatmap representation for its body keypoint and is capable to learn and represent both the body part appearances and the statement of the body part configuration.

In 2016, Xiao Chu, Wanli Ouyang, Hongsheng Li and Xiaogang Wang[6] proposed a structured feature learning model to check the correlations among body joints at the feature level in the field of human pose detection.

In 2018, Wenlin Zhuang, Siyu Xia and Yangang Wang[7] proposed a novel approach. The approach uses a nonparametric representation, which they refer as Direction Maps, to study the direction information of human body parts. The whole architecture is basically divided into two parts. The first part is designed to jointly learn parts location and direction of each body part relative to the center of body. The second part integrates the location and direction information to obtain more accurate posture.

In 2012, Alexandros Iosifidis and Anastasios Tefas[8] proposed a novel view invariant action recognition technique based on neural network representation and recognition is proposed. This proposed method based on learning human body posture prototypes using self organizing maps.

In 2010 Poppe published another survey [9], and focused on Human Action Recognition. In [10], Weinland et al. describe approaches to Action Representation, Segmentation and Recognition.

In 2014, Toshev and Szegedy[11] applied Deep Convolutional Neural Networks for pose recovery. In this work, pose estimation is formulated as a joint regression problem, in which the coordinates of the body joints are directly regressed through a cascade of convolutions, followed by fully connected networks, which outputs a vector with the coordinates of the target joints. This approach is typically considered the earliest work in Human Pose Estimation using novel Deep Learning techniques.

In 2016, Lifshitz Ita, E. Fetaya, and S. Ullman[12] proposed a novel approach where each location in the image votes for the position of each keypoint using a convolutional neural network. The voting scheme collect the information from the image, rather than rely on a sparse set of keypoint locations.

In 2018, N. Geetha and E. S. Samundeeswari[13] published a review paper and discussed about different types of recognition techniques of human actions represented in a digital images. They discussed about existing methods of human activity recognition.

© 2019, IJCSE All Rights Reserved

III. METHODOLOGY

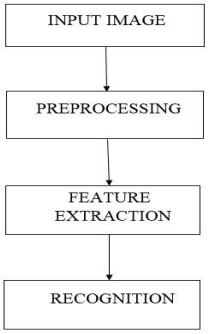

Figure *1. Block Diagram of Human Pose Detection System*

In our proposed system the computer or the system should be able to detect two most important human pose sitting and standing using machine learning techniques. Human pose detection by machine is helpful in different inspection application like activity or behavior analysis. In Human machine interaction control systems like athlete performance analysis take help of pose identification techniques. Figure 1 shows a basic block diagram for pose detection method. According to this block diagram, the input digital image is preprocessed for more enhancement. Next important and necessary features of the interested regions are extracted. In the last phase recognition can be done from the extracted image. Figure 2 shows the flow chart of our proposed system. The figure clearly shows the training and testing phases with different modules. At last accuracy will be calculated which we expect to be satisfactory.

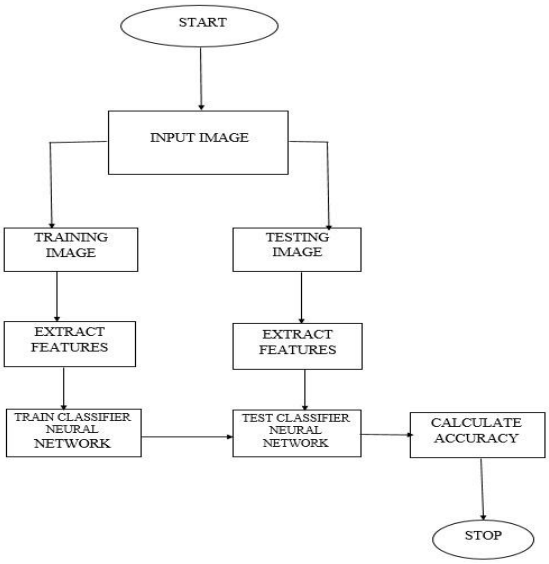

Figure *2. Flowchart for the proposed system*

Image Database: The image database is created for the recognition of human pose and is classified into two classes for distinct training and testing purposes. During training purpose, 1700 images are used, containing two human pose image sitting and standing. Figure 3 shows a preview of database of training and testing image.

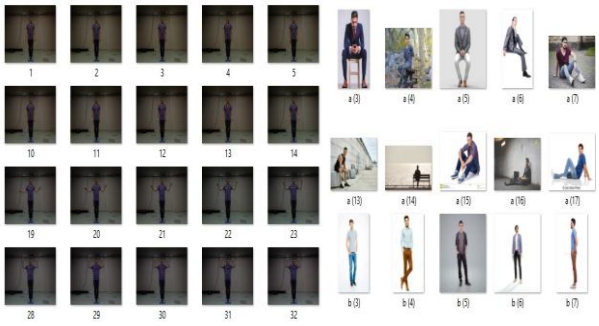

Figure *3. Preview of database of image*

Preprocessing: The aim of pre-processing of digital image is to improve the image data which may be suppressed unwilling or for removal of salt and pepper as well as Gaussian noises. Pre-processing performs two tasks, they are calculation of Region of Interest (ROI) and conversion of colored image to gray scale image[14]. Enhancement of digital image is important as this process make the image clear, bright and distinct, but does not add extra information to it. Some preprocessing techniques that are used in this process are mentioned below.

Grayscaling: A grayscale image is one which has only shades of gray colors scaling from 0 to 255. 0 means total black color and 1 means absolute white color. Using gray scale in digital image processing amount of information to be processed is reduced. Which in turn reduces processing cost. Gray scale poses equal intensity for all the components of Red, Green and Blue in the common RGB scale and hence single intensity value for every pixel is enough for processing. Figure 4 shows one of our example image which is converted from RGB to black and white.

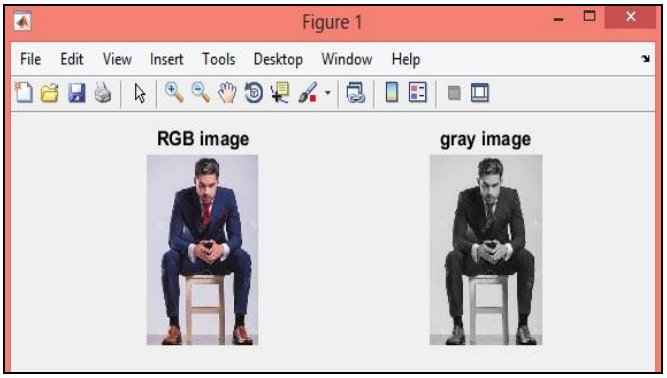

Figure *4. RGB to gray Conversion*

© 2019, IJCSE All Rights Reserved

Filtering: Filtering is very necessary in digital image enhancement phase. Basically there are three types of filters high pass, band pass and low pass. High pass filter allows only high frequency to pass, for example line, border etc. low pass filter allows only low frequency regions like background or any other region with slight variation in gray values. Image smothing, blurring or contrasting are done with the help of these filters.

Image Scaling: image scaling is related with computer graphics as well as in digital image processing. The term image scaling means change the existing coordinates of all the objects in the image to a new one. After scaling a digital image, a new image must be generated which are composed of higher or lower number of pixel in comparison with original image.

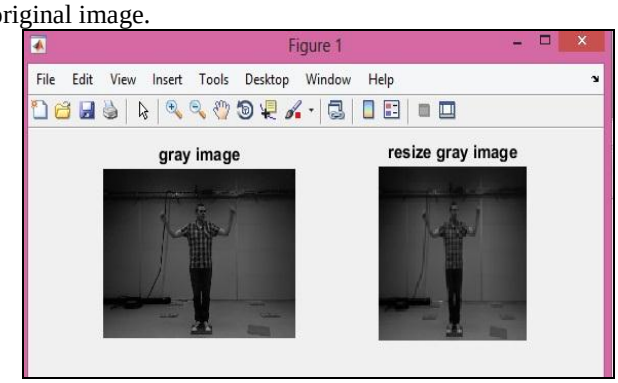

Figure *5. Resized image*

Feature Extraction: In the field of digital image processing feature plays a very important role. Various image preprocessing methods like gray scaling, thresholding, resizing, filtering etc. are applied on the image before feature is extracted. The features are extracted using specific feature extraction methods which can be later used for classification and recognition of images.

Histogram Oriented Gradient(HOG): Dalal and Triggs[15] developed a point descriptor which works by describing a point through information collected in a defined environment around it. HOG is based on image gradients and collects information in a rectangular environment. We have used this concept in our work. The HOG descriptor is calculated for all points in an image as it is frequently used in detection.

After the feature extraction process, the extracted value are classified into two pre defined target classes for training purpose which are defined by "01" for sitting pose and "10" for standing pose. Figure 6 shows the HOG features over one of our sample data. The second image of figure 6 shows this HOG feature values of the first one.

Figure *6. HOG features over the original image*

Training Neural network Classifier: For training purpose, around 1700 total sitting and standing images are collected. After extracting the input features, the training set is used to train using artificial neural network and the output is divided into two classes. They are sitting and standing. The implementation utilizes the Matlab Neural Network Toolbox. For each training session the input and targets vectors are randomly divided into three sets. 70% of the vectors are used to train the network. 15% of the vectors are used to validate how well the network generalized. The remaining 15% are used as a testing set.

Testing Neural Network Classifier: The trained Classifiers are then tested with a dataset of images which were not used for training the classifiers.

Tools/Environment/Experimental Platform: MATLAB is a 4th generation programming language. MATLAB allows different types of features which are implementation of algorithms, matrix manipulations, plotting of functions and data, creation of user interfaces etc. During the whole project windows 8 operating system is used. Figure 7 is a snapshot of NN training dialog box.

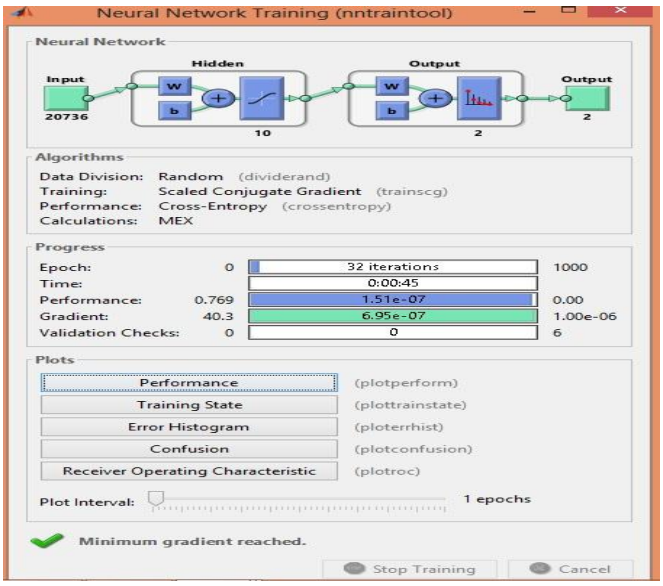

Figure *7. Neural network training dialog box*

© 2019, IJCSE All Rights Reserved

IV. RESULTS AND DISCUSSION

We have trained this system to detect the standing and sitting position of human being using ANN. To do so we have used 1700 images which are commonly found. We have tested the system with 50 images in preliminary phase. We have got an accuracy of almost 75%. Figure 8 shows a snapshot of two output testing images standing and sitting poses respectively. We are also able to tag the image with correct tag, which are shown in both the poses of figure 8 in yellow color.

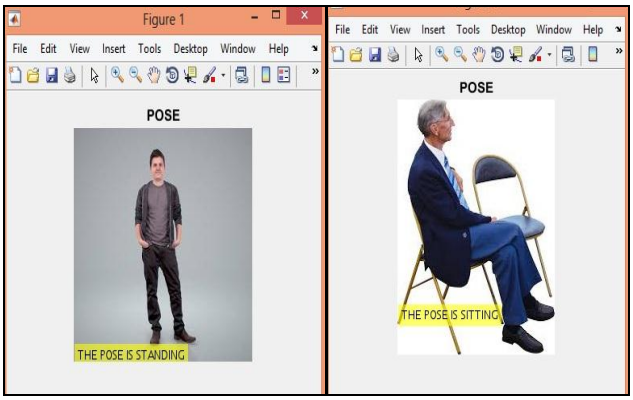

Figure *8. Output of the testing image*

V. CONCLUSION AND FUTURE SCOPE

We have completed detection of two positions sitting and standing with success. Our final aim is to detect at least four commonly found human poses standing, sitting, handshaking and waving. In future we will train and test the system with more data which will certainly increase efficiency of our model, more than now achieved 75%.

REFERENCES

[1] Afzal Ahmad, Mohammad Asif and Shaikh Rohan Ali, "*Review Paper on Shallow Learning and Deep Learning Methods for Network security*", International Journal of Scientific Research in Computer Science and Engineering, Vol.6, Issue.5, pp.45-54, October **2018**.

[2] N.S. Lele, "*Image Classification Using Convolutional Neural Network",* International Journal of Scientific Research in Computer Science and Engineering, Vol.6, Issue.3, pp.22-26, June **2018**.

[3] Chaitra B H, Anupama H S and Cauvery N K, "*Human Action Recognition using Image Processing and Artificial Neural Networks*", International Journal of Computer Applications (0975 – 8887)Volume 80 – No.9, October **2013**.

[4] Yi Liu, Ying Xu and Shao-bin Li, "*2-D Human Pose Estimation from Images based on Deep Learning : A Review*", 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference,**2018.**

[5] Vasileios Belagiannis and Andrew Zisserman, "*Recurrent Human Pose Estimation*" 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition, **2017**.

[6] Xiao Chu, Wanli Ouyang, Hongsheng Li and Xiaogang Wang, "*Structured Feature Learning for Pose Estimation*", IEEE Conference on Computer Vision and Pattern Recognition, **2016**.

[7] Wenlin Zhuang Siyu Xia ,Yangang Wang, "*Human pose estimation using Direction Maps*", 33rd Youth Academic Annual Conference of Chinese Association of Automation (YAC), **2018**.

[8] Alexandros Iosifidis and Anastasios Tefas, "*View-Invariant Action Recognition Based on Artificial Neural Networks*", IEEE Transactions on Neural Networks and Learning Systems, Vol. 23, No. 3, March **2012**.

[9] Ronald Poppe, "*A survey on vision-based human action recognition*", Image and Vision Computing, Volume 28 Issue 6, pp.976-990, June **2010**.

[10] Daniel Weinland, Remi Ronfard, and Edmond Boyer, "*A survey of vision-based methods for action representation, segmentation and recognition,*" Computer Vision and Image Understanding, 115(2):224-241, February **2011.**

[11] Toshev, Alexander and Christian Szegedy, "*DeepPose: Human Pose Estima-tion via Deep Neural Networks*", 2014 IEEE Conference on Computer Vision and Pattern Recognition, **2014**.

[12] Ita Lifshitz, Ethan Fetaya and Shimon Ullman, "*Human Pose Estimation Using Deep Consensus Voting*", European Conference on Computer Vision, March **2016**.

[13] N. Geetha and E. S. Samundeeswari, "*A Review on Human Activity Recognition System*", International Journal of Computer Sciences and Engineering, Vol-6, Issue-12, Dec **2018**.

[14] Karabi Barman, Parismita Sarma, "*Glaucoma Detection Using Fuzzy-C Means Clustering Algorithm and Thresholding*", International Journal of Computer Sciences and Engineering, Vol. 7, Issue. 3, pp. 859-864, **2019**.

[15] Navneet Dalal and Bill Triggs, "*Histograms of Oriented Gradients for Human Detection*", 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, **2005**.

Authors Profile

Munindra kakati is an M.Tech student in the Department of Information Technology, Gauhati University. He had done his B.Tech in computer science and engineering from Gauhati University Institute of Science and Technology. His area of interest is Image processing.

Parismita sarma is an assistant professor in the department of Information Technology, Gauhati University. She has PhD in the department of Information Technology from Gauhati University. She is a graduate BE in computer science and engineering from Jorhat Engineering College. Parismita has successfully completed MTech in Information Technology from Tezpur University, Assam. She has attended many conferences and seminars and has almost twenty publications both in conferences and international Journals.

© 2019, IJCSE All Rights Reserved