survey and comparision of human pose estimation

[1]narmatha.M, [2]S.N shivappriya

[1]PG STUDENT ,DEPARTMENT OF ECE KUMARAGURU COLLEGE OF TECHNOLOGY
[2]ASSISTANT PROFESSOR , DEPARTMENT OF ECE ,KUMARAGURU COLLEGE OF TECHNOLOGY

Abstract

In this project human pose estimation based on deep neural network is proposed. The main stream of work in this field has been motivated mainly by the first challenge, the need to search in the large space of all possible articulated poses. This work uses a multi task learning called manifold and Eigen decomposition algorithm which results in high precision pose estimates. Human pose estimation in video is usually conducted by matching 2D image features and retrieving relevant 3D human poses. In the process of retriving, the mapping between images and poses is critical. By using global joint localization and local joint detection, pose estimation can be recovered with better accuracy. The approach has the advantage of reasoning about poses in a holistic fashion so that has a better performance can be achieved compared to the traditional methods. Pose estimation is extracted and classified by different poses such as running, walking, playing ,jogging etc.. Human pose estimation finds its application in image recognition, theft and terrorist activities and video games.

Index Terms—*manifold and eigen decomposition ,global joint localization, pose estimation ,holistic fashion.*

I. INTRODUCTION

Deep Neural Network is a network that has an input layer, an output layer sand at least one hidden layer and the additional layers increasability of the neural network to discriminate between classes with better results. The main advantage of using this sophisticated network is dealing with unlabeled or unstructured data Learning can be supervise classification) and unsupervised (pattern analysis). Deep learning is part of a larger family of machine learning methods that is based on representation of learning data, as opposed to task-specific algorithms. Deep learning architectures such as deep neural netwok, deepbelief network and recurrent network have been applied to various fields that already including computer vision, speech recognition, and game programs, where the results produced are comparable and in some cases superior to human knowledge.

The Convolution neural network is a class of deep feed forward artificial neural network most commonly applied to analyze the various visual imaginary. CNN uses a variety of multilayer perceptron mainly designed to require preprocessing compared to image classification algorithms.In deep learning, each level learns how to transform their input data into a slightly more abstract and composite representation. In an image recognition application, the raw input are considered to be a matrix of pixels; the first representational layer may portray about the pixels and encode edges; then second layer may composed and encoded arrangements of various edges; the third layer may encode a nose and eyes; and the fourth layer may recognize whether the image contains a face. Artificial Neural Networks (**ANNs**) or connection systems are computing systems inspired by the biological neural networks that constitute animal brains. Such systems learn and say how to do tasks by considering examples, generally without task-specific programming.

An ANN is based on a collection of connected units called artificial neurons (same as biological neurons in a biological brain). Each connection between neurons can transmit a signal from one neuron to another neuron. The receiving neuron can process the signal and also transform signal downstream neurons connected to it. Neurons may have stated that, generally it is represented by real numbers, typically between 0 and 1. Neurons and synapses may also have a weight that varies as that of learning procedure continues, which can increase or decrease the strength of the signal by sending it towards downstream. Typically, neurons are organized in the shape of layers. Different layers may perform various and different kinds of transformations on their inputs. Signals travel from the first to the last layer, that is after traversing the layers multiple times. A deep neural network (DNN) is an artificial neural network (ANN) with multiple layers that finds its application between the input and output layers. The DNN finds the correct mathematical manipulation to turn the input into the output, whether it is a linear relationship or a non-linear relationship. Convolutional layers apply a convolution operation to the input, and then passing it to the result to the next layer. The convolution transform the response of an individual neuron to visual stimuli. Each convolutional neuron processes data only for its receptive purposes. A very high number of neurons would be necessary, due to the very large input sizes that are associated with images,where each pixel is a relevant variable. Neural networks have been used on a variety of tasks, including computer vision, speech recognition etc..

Automatic speech recognition: Speech recognition is the inter disciplinary sub field of computational linguistics that develops methodologies and technologies that enables the recognition and translation of spoken language into text by computers. It is also known as automatic speech recognition (ASR),and also known as computer speech recognition or speech to text (STT). It incorporates knowledge and research in different fields such as the linguistics, computer science and electrical engineering fields.

Image recognition: Computer vision is an interdisciplinary field that deals with the computer are to gain a high-level understanding from digital images or videos. From the perspective of engineering, it refers to automate tasks that can be done by the human visual system .Computer vision tasks also include the methods for acquiring, processing, analyzing and understanding of digital images, and the extraction of high-dimensional data from the real time data in order to produce numerical or symbolic information.

In this project, pose estimation of humans are estimated using deep learning. Pose estimation is a difficult problem and an active subject of research because the human body has [Wanli Ouyang et al,2014]244 degrees of freedom with 230 joints. Although not all movements between joint are evident, the human body is composed of 10 large parts with 20 degrees of freedom. Algorithms must account for the large variability introduced by differences in appearance due to clothing, body shape, size, and hair style. Additionally, the results may be with error due to partial occlusions that is from self-articulation, such as a person's hand covering their face, or from external objects. Finally, most algorithms are there to estimate poses from monocular (two-dimensional) images, taken from a normal camera. Other issues include varying lighting and camera configurations. Pose estimation finds its application in assisted living, Character animation, intelligent driver assisting system Video games, video surveillance system and cloth parsing. To estimaste human poses using deep learning, aim of this work is to propose a framework that combines computer vision based on deep neural network to recognize human body poses from images taken by a camera. Recovering human pose is very critical in computer vision by precisely revealing the body joints from 2D images. A dataset is a collection of discrete items of related data that may be accessed individually or in combination or managed as a whole entity. Here dataset are collected as an individual images and also as a group. Real time images and also images related to pose estimation are taken.Nearly 350 images with different poses are taken for training and tested and classified. Images that are taken to estimate the pose have a holistic view.Also local body parts and joint localization can be estimated accurately.

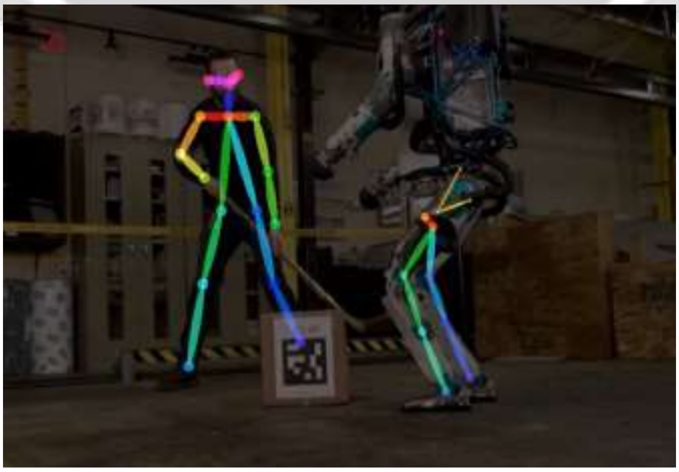


FIG 1.example of datasets for human pose estimation

II. LITERATURE SURVEY

Deep learning provides determination of various poses accurately and finds its application in various fields. Human pose estimation known as the problem of localization of human joints has managed to gather an attention in the computer vision. The main stream of this work provides lot of articulated poses as holistic view of human pose estimation. This approach uses DNN based regression method to estimate the human poses and this approach has the advantage of predicting poses in a holistic fashion and it is simple but powerful formulation. However, due to its fixed input size of 220 ^ 220 its capacity is limited. so,it cannot be easily increase the input size because it may increase the large number of parameters. DNN-based regression has the advantage of capturing context and reasoning about pose in a holistic manner[1]. Human pose recovery in videos is usually done by matching the 2D image features and retrieving the relevant 3D human poses.In the retrieving process mapping between the images and poses is critical. Traditional methods mapped as local joint detection or global joint localization, that limits the performance of a network.To overcome this disadvantage of global joint localization, continuous learning are made to estimate the holistic view of an image. Multiple manifold learning is used to calculate the various parameters.Multiple manifold is integrated and generalized eigen decomposition is mainly used to achieve parameter optimization. The tree structure was utilized in order to make the recovery procedure more efficient. The tree structure has the advantage of achieving an efficient computation by using the dynamic programming. Multitask learning approach for Human Pose Recovery (HPR), is used here[2]. First, (Multi task learning Auto encoder model)MTAM it extracts multiple features from both global and local body parts. A multi graph learning based approach is proposed to integrate those features. Second, MTAM provides an one shared auto encoder model in order to obtain hidden representations for both global and local parts. Third ,it incorporate the tasks of joint localization detection into MTAM.A new architecture [3] for human pose estimation that uses a multi layer convolutional network architecture and a modified learning technique which learns low-level features. The most challenging is to extract human pose from monocular RGB images with no specification or prior assumption. This specific variation of deep learning acheives a great performance when compared to the traditional architectures. The first end-to-end learning approach for full body human pose estimation are proposed and also the Deformable Part Models (DPMs) on a modern challenging dataset is also done here, so that an analysis of improves joint localization for approximately 5% of the test set case.To estimate human poses in videos[4] since multiple frames are also available.

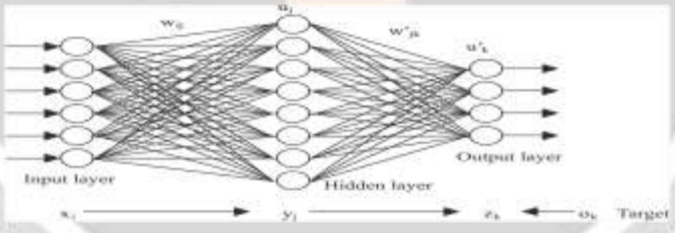


Fig 2 Neural Network

when compare to the still images, recognition of human poses in videos provide an incomparable position of human interms of their joints. A deeper network architecture are used for regressing the heat maps; (ii) spatial fusion layers that learn an spatial model; (iii) optical flow is used to align the heat map predictions; and (iv) a final parametric pooling layer which learns to combine the aligned Heat maps into a pooled into confidence map. Both classification and the segmentation can be done quite appropriate in videos compared to the images. Human pose recovery[5] from the videos can be retrieving by comparing with the images,which has been proposed as a novel pose recovery method using non-linear mapping with multi-layered deep neural network. It is based on feature extraction with multimodal and back-propagation using deep learning. Auto encoder are used to compact the representation of the input to get the accurate result. The Multimodal Deep Auto encoder (MDA) has a three-stage architecture. The first and third stages employ two auto encoders for learning the inner representations of 2D images and 3D poses, respectively. The second stage incorporates a two-layer neural network to transform the 2D representation into the 3D representation. However, for Human3.6M, performance is obviously affected by the noise level. To estimate and track articulated human poses in sequences [6]from a single view, real-time range sensor. MarkovChain Monte Carlo (MCMC) framework are used to find an optimal pose that is based on the comparisons between the depth images. Bottom detectors are used to speed up the convergence particularly hand and forearm locations. Tracking performance is quantitatively evaluated using specific trained models. Data sets include four different subjects with different limb widths and lengths.

Developed a robust pose tracking system that combines bottom-up candidates and are efficient using a data driven MCMC framework on range images it can process multiple parallel Markov chains and rasterize more complex limb models. Conventional human detection[7] is mostly done using the images taken by visible-light cameras and it portrays about the detection process. Histograms Of Oriented Gradients (HOG), or extract points in the image, such as Scale-Invariant Feature Transform (SIFT), are used here. Proposed a model based approach, which detects humans using 2D head contour models and a 3-D head surface model. This algorithm can effectively detect the persons in all poses and their appearances from the depth, and it provides an accurate estimation of the whole body contour of the person. This method can easily adjust to new datasets, no training is needed for this datasets. Second, the algorithm uses a two layer detection process with 2D matching in the first layer which largely reduces computational cost. Third, assume person's pose for accurate detection. The limitation is that this algorithm has a high dependency based on the accurate head detection. 3D human poses that are to be extracted from 2D joint locations [8] for the analysis of people in images and videos are provided here.Ill poses is a major problem and to overcome the invalid poses first, to collect a motion capture data set that explores a wide range of human poses and from this a learning on pose-dependent model of joint limits is achieved. A new dataset of human motions that includes an extensive variety of stretching poses performed by trained athletes and gymnasts are recollected. This method significantly outperforms the current state of the art methods, both quantitatively and qualitatively. This prior and the optimization can be applied to many problems that takes place in human pose estimation. For jointly inferring human body pose and human attributes[9] in a graph with attributes, attributes and-or grammar (A-AOG) model are used. This method explicitly represents the decomposition and articulation of body parts, and account for the correlations between poses and attributes. The A-AOG model is an amalgamation of three traditional grammar formulations:(i) Phrase structure and grammar representing the hierarchical decomposition of the human body from whole to parts; (ii) Dependency of grammar modeling based on the geometric articulation by a kinematic graph of the body pose and (iii) Attribute grammar accounting for the compatibility relations between different body poses in the hierarchy. The mean average precision is 2.6% (8-layer) and 2.7% (16-layer) lower. The advantage of this approach is the ability to perform simultaneous attribute reasoning and part detection, unlike previous attribute models that use large numbers of region-based attribute classifiers without explicitly localizing parts. Recovering 3D full-body human pose [10] is a challenging problem with respect to many applications. It has been successfully addressed by the motion capture systems, with the multiple cameras. Deep learning approaches have shown remarkable abilities to learn 2D appearance features. Joint location uncertainties can be conveniently figured out during inference. The 3D poses are modeled by a sparse representation and the 3D parameter that are estimated are realized through an Expectation-Maximization algorithm, where it is shown that the 2D joint location uncertainties can be conveniently figured out during inference. The uncertainty is modeled by a Gaussian centered at the annotated joint location of the body poses Experiments demonstrates that 3D geometric priors and temporal coherence does not only help 3D reconstruction but also improve 2D partial localization. The EM algorithm usually converges in 20 iterations with a CPU that uses time less than 100s for a sequence of 300 frames. Alternative part detectors, pose representations, and temporal models are not integrated but can also be integrated and detected.The visual appearance and mixture type and deformation are three important information sources for the human pose estimation[11].This method Proposed to build a multi-source deep model in order to extract a non-linear, representation, from these different aspects of information sources. This is a post-processing of pose estimation results and can flexibly integrate with the existing methods by taking their information sources as input. By extracting the non-linear representation from the multiple information sources that have been taken, the deeper model outperforms state-of-the-art by 8.6 percent on three public benchmark datasets. A multi-source deep model is then applied to an individual of all body locations in order to determine whether the body locations are correct or not. This approach is limited because information sources with different statistical properties are mixed in the first hidden layer. A better solution is to have their high-level representations constructed before they are mixed. Deep Convolutional Neural Networks [12] have been applied to the task of human pose estimation, and have shown its potential of learning better feature representations and capturing contextual relationships. However, it is difficult to incorporate domain prior knowledge such as geometric relationships among different body parts into DCNN. Approach [12] significantly improves the performance compared with state-of-the-art approaches, especially on benchmarks with challenging articulations. As a consequence, during the training stage, these approaches may produce many imperfect results, Errors on these regions will be back propagated to penalize the features correspond to head detection, which is inappropriate. This proves the effectiveness of joint training DCNNs 18 and a deformable mixture of parts. Part detector and message passing are jointly learned, but separately learned . Second, to build a loopy model based on tree-structured model by adding an edge between knees, and get a 0.5 % improvement. Action recognition [13] and human pose estimation are closely related, but both problems are generally handled as different tasks in the literature. In this work, proposed a multi task framework for jointly 2D and 3D pose estimation from still images and human action recognition from various video sequences. One of the most important advantages in our proposed method is the ability to integrate high level pose information with low level visual features in a multitask framework. Moreover, the full optimization also improves by 3.3%, which says the importance of a fully differentiable approach. And finally, by averaging results from multiple video clips we gain 1.1% more. MoCap). However, due to the use of infrared projectors, these depth sensors are limited to indoor environments. Moreover, they have a low range precision and they are not robust to occlusions, frequently resulting in noisy images . 3D pose estimation, 2D action recognition, 3D action recognition with a single model very efficiently compared to dedicated approaches.

III. DEEP NEURAL NETWORK

The pose estimation as a joint regression problem and it can be successfully cast it in DNN settings. Deep Neural Network (DNN), have shown far better performance on visual classification tasks compared to traditional methods and more recently on object localization. The location of each body joint is regressed to using as an input as full image and a 7-layered generic convolutional DNN. There have been two advantages of this formulation.First, the DNN is capable of capturing the each body joint each joint regressor uses the full image as a signal. Second, the approach is simpler to formulate than methods based on graphical models and then no need to explicitly design feature representations and detectors for parts; no need to explicitly design a model topology that saves time and interactions between joints. Detecting human poses from images or videos is a challenging problem facing nowadays that have variations in different types of pose, clothing, lighting conditions and complexity of the backgrounds. Human pose estimation are based on these steps: Detecting human on input images, Extracting human silhouette from and human body pose based on silhouette using neural networks. It is not only important to make the detection of pedestrian but it is also crucial to understand in which poses the pedestrian is such as standing, walking, running, etc.. The recognition part was achieved by hierarchically separating all features that were extracted from the spatial body language ration. The recognition part was achieved by hierarchically separating all features that were extracted from the spatial body language ration. Machine learning is one of the most powerful ways of performing analysis in computer vision with methods like SVM and RF receiving some of the highest success rates. These methods receive high success rates in many cases, but their overall performance often falls short of the very high accuracy required for fully automated systems. Neural networks and deep learning try to solve this problem by learning features themselves. This approach increases accuracy given large enough dataset for training. A successful tracking system can find applications in motion capture, human computer interaction, and activity recognition.

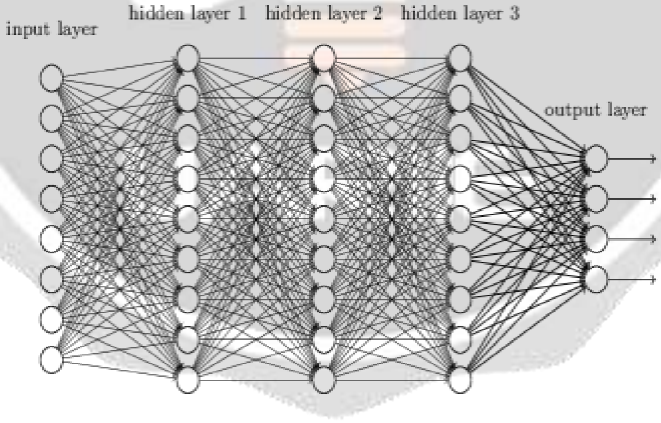


Fig 4.Deep Neural Network

CONCEPT OF POSE ESTIMATION

The proposed method has a manifold regularizer on the framework and simultaneously optimizes the parameters of linear projection and task shared parameter. In this way, the effects of different tasks can be adaptively modulated. Multi-Task Auto encoder Model (MTAM) is designed for HPR. First, MTAM extracts multiple features from both global and local body parts. A multigraph learning based schema is proposed to integrate these features. Second, MTAM employs one shared auto encoder to obtain hidden representations for both global and local parts. The auto encoder's encoding process can initialize the shared parameter γ for the following multi-task learning procedure. Third, we incorporate the tasks of joint localization and detection into MTAM. A prominent characteristic of this method is that the multi-task learning is able to enhance the parameter estimation. Hence, it is more efficient than existing methods in HPR. Because MTAM is supervised, multi-source and multitask, the inner representations and the appropriate relationship between the tasks of joint localization and joint detection be simultaneously explored through optimization.This figure is taken from "HUMAN BODY POSES RECOGNITION USING NEURAL NETWORKS WITH CLASS BASED DATA AUGMENTATION"by Karl-Kristjan Luberg.

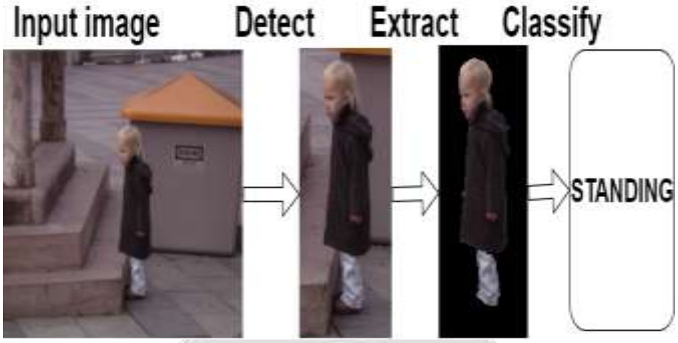


Fig 4.Methodology step by step

Detection:

Detection of image is the first step in pose estimation. Only after detecting the images we can classify by extracting their features. At a structural level, a human has a head, arms, torso and legs. Computer vision can be used to exploit these basic traits to detect humans from a random image. Despite the fact that idea of applying HOG descriptor for object recognition is nearly a decade old, it is still used a lot and shows good results. Therefore, for our preliminary step of detecting pedestrians, we used OpenCV library, which has pre trained descriptors to be applied for human detection. The process starts by using images of interest that are all resized to 480x640 pixels if necessary. sometimes multiple and overlapping bounding boxes are detected for each human. To effectively use the found bounding boxes in the steps of problem resolving, its necessary to extract a clear bounding box for each human on the image.This figure is taken from "Human body pose recognition using neyral network with class based data augmentation"by Karl-Kristjan Luberg.

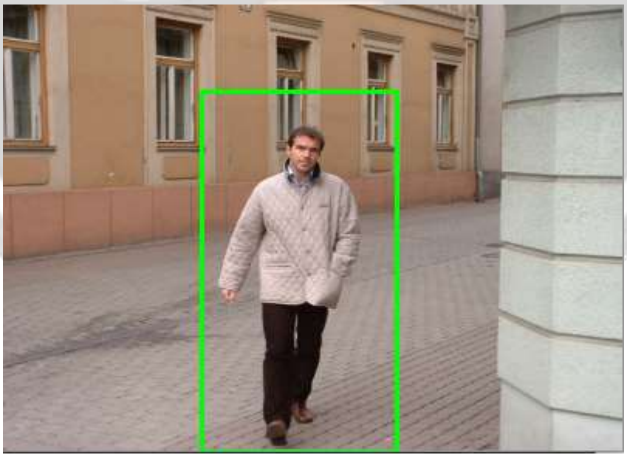


Fig 5.Example of walking human detected and highlighted with green rectangle.

Extraction:

The human silhouette extraction from videos can be a straightforward process with the assumption of that the environment is always under control. We can rely on the moving object when comparing the frames and detect and extract the human body. However, this task becomes a little bit challenging when working with single images. After detection phase, once we are sure that the human is in the segmented image, the image can be divided into two parts – background and foreground. The foreground is the bounding box from the previous step where human has been detected. The background is the part of the image outside of the bounding box. The background is the part of the image othat is outside of the bounding box. GrabCut utilizes this information and the algorithm is introduced as a solution for foreground extraction. Then a bounding rectangle is drawn. The foreground object (human) must be completely fit in the rectangle. This algorithm segments foreground iteratively to get the best result. This algorithm segments pixels into foreground and then into background. Background pixels are coloured with black and the end result is an image where the foreground object is clearlyy brought out. This figure is taken from “Human body pose recognition using neyral network with class based data augmentation”by Karl-Kristjan Luberg.

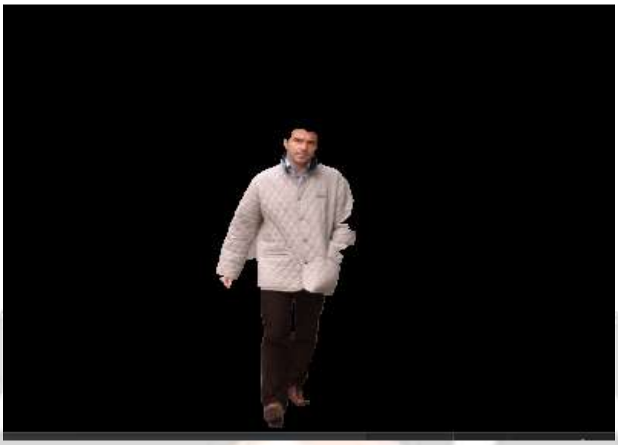


Fig 7.Example of walking human silhouette extracted.

Classification:

The previous steps – that is detecting a human from an image and extracting silhouette are applied to hundreds of images. In this thesis, different poses are classified standing and walking. The dataset for training classifiers consists of 226 images of silhouettes, which are augmented to form a dataset of 2260 images of silhouettes in total.

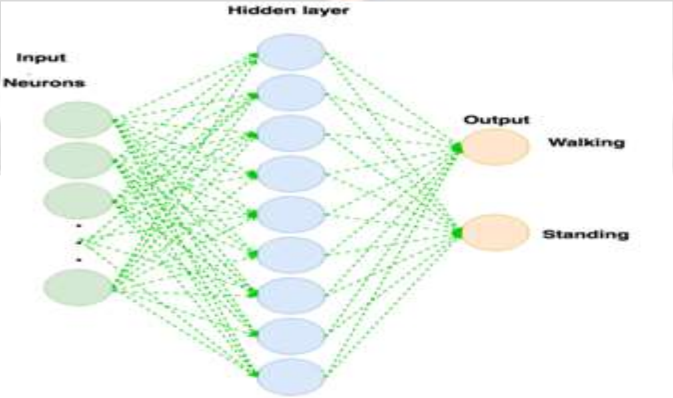


Fig 9.Neural network structure visualized.

The neural network outputs classification value, which is translated into text and written on the input image of a human silhouette. This figure is taken from “Human body pose recognition using neyral network with class based data augmentation”by Karl-Kristjan Luberg.

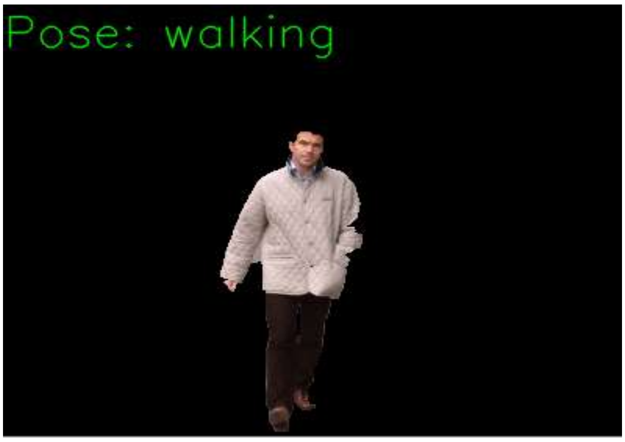


Fig 10.Example of walking human body pose correctly being classified by the neural network.

IV. RESULT

Human poses are estimated and compared using machine learning. Different types of poses are detected and estimated here. Support Vector Machine, Articial Neural Network algorithm and k nearest neighbouring are used. The **svm**, is derived entirely on the basis on several simple mathematical techniques. Morphological based segmentation, hsv feature extraction and matching score based classification are the steps used. lower performance and highest system complexity are the drawbacks of svm. Knn has the working of taking the neighbor sample ansd calculating the average rather than taking the distance samples.ANN has the advantage of using target value equal to input value.so it overcomes the disadvantage of svm and knn. Also,in feature extraction autoencoder are used so it reduces the dimensionality and gives better accuracy.some of the pose estimation results are given here.

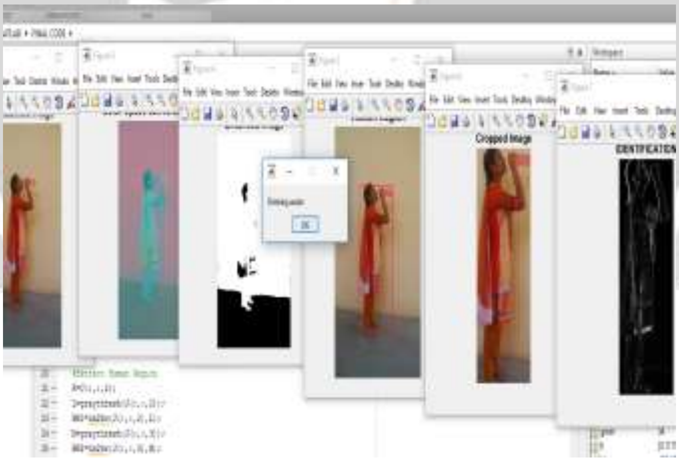


Fig 12human pose are estimated and classified as drinking water.

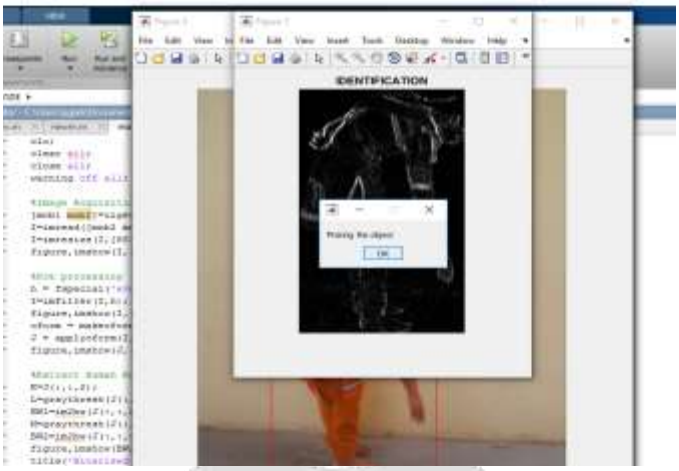


Fig 13. Human pose is classified as picking the object

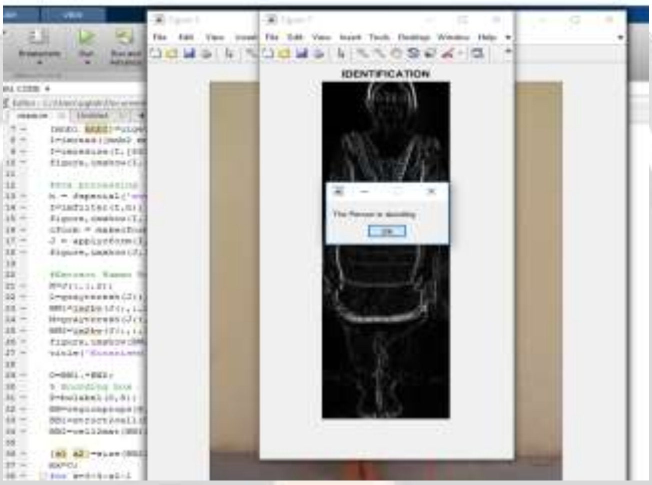


Fig 14.Human pose is classified as standing.

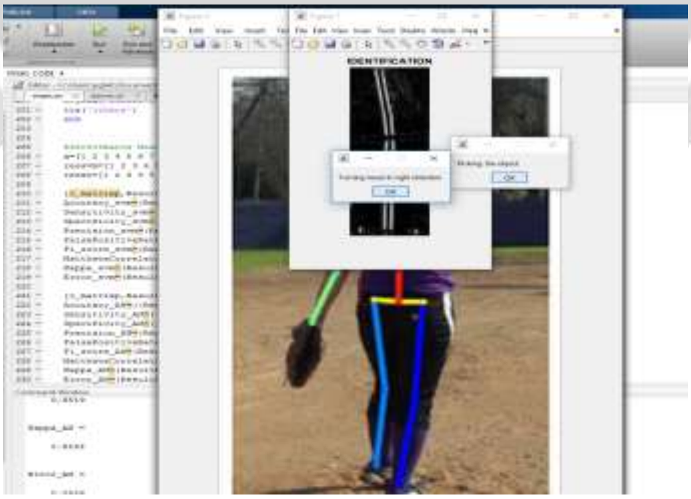


Fig 15.Human pose is estimated and classifies as turning head in right direction and picking the object.

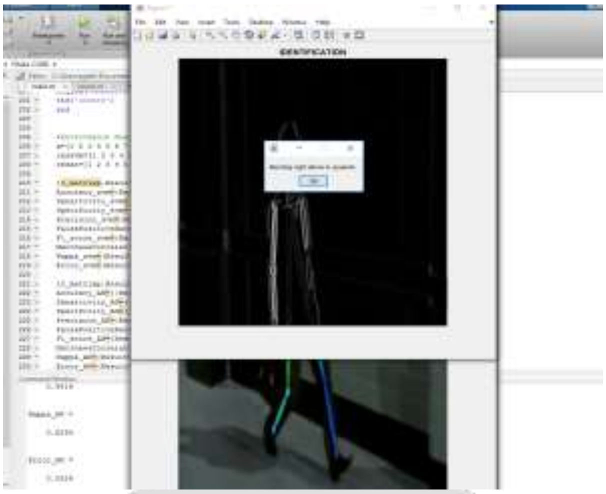


Fig 16.Human pose is estimated and classified as bending right elbow in upwards.

In detection ,morphological segmentation are used to view the skeleton view of an image. Then for extraction Hue Saturation Value are used because in this method dark colors are shaded as black and white colors are identified as white pigment. Finally , it is classified using different algorithms.

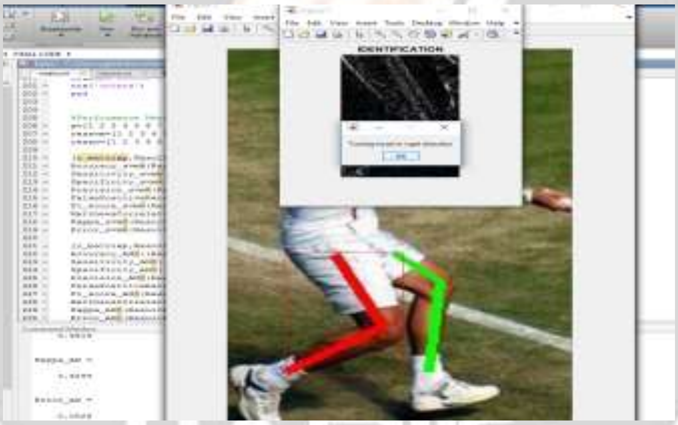
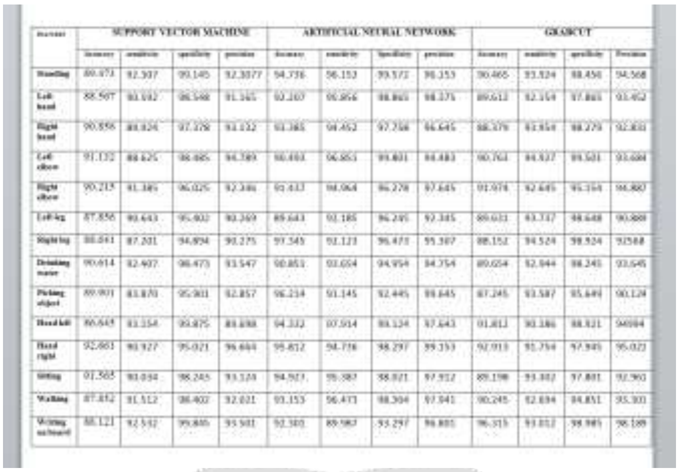


Fig 17. Human pose is classified as turning head in right direction.

The results for different algorithms are estimated and compared. In SVM, non separable classes are converted into separable classes but dimensionality increases is the disadvantage of this algorithm. Knn uses nearby samples so that minimam distance are cIn ANN, Artificial neural network is most commonly applied to analyze the visual imaginary. It uses variation of multilayer perceptrons mainly designed to require preprocessing compared to other image classification algorithms. Backpropogation is used in ANN by using the feedback error are reduced and accuracy is improved. Mathematical morphology is a tool for extracting image components that are useful in the representation and description of the region shape, such as boundaries and skeletons. The HSV color space is quite similar to the way In which humans perceive color. Improved performance with reduced complexity. Comparision for different algorithms are done in matlab and the plotted in table below.



[illegible]	SUPPORT VECTOR MACHINE				ARTIFICIAL NEURAL NETWORK				GRABCUT			
	Accuracy	sensitivity	specificity	precision	Accuracy	sensitivity	Specificity	precision	Accuracy	sensitivity	specificity	Precision
Standing	89.973	92.307	99.145	92.3077	94.736	96.153	99.572	98.153	96.465	93.524	88.456	94.568
Left hand	88.507	93.692	98.548	91.365	93.207	96.956	88.865	98.375	89.612	92.154	97.865	93.452
Right hand	90.856	88.926	97.378	93.132	93.365	94.452	97.756	96.645	88.376	93.954	98.279	92.833
Left elbow	91.132	88.625	98.685	94.789	98.493	96.851	99.801	84.483	90.763	84.927	89.501	93.684
Right elbow	90.215	91.386	96.025	92.386	93.433	98.064	96.278	97.645	91.978	92.645	95.164	94.800
Left leg	87.856	90.643	95.902	90.369	89.643	92.185	96.245	92.385	89.633	83.737	98.648	90.888
Right leg	88.041	87.201	94.894	90.275	97.345	92.123	96.473	95.307	88.152	94.524	98.924	92588
Drinking water	90.614	92.407	98.473	93.547	90.851	93.654	94.954	84.754	89.654	92.944	98.245	93.645
Picking object	89.901	83.870	95.901	92.857	96.234	91.145	92.445	88.645	87.245	93.587	85.649	90.124
Head left	86.645	93.354	99.875	88.698	94.332	97.914	88.134	97.643	91.812	90.186	88.921	94084
Hand right	92.803	90.927	95.921	96.664	95.812	94.736	98.297	99.353	92.913	91.754	97.945	95.022
Sitting	91.565	93.034	98.245	93.124	94.921	95.387	98.021	97.912	89.198	93.102	97.801	92.963
Walking	87.852	91.512	98.402	92.021	93.153	96.473	98.304	97.941	90.245	92.834	84.851	95.303
Writing on board	88.121	92.532	99.845	93.501	92.303	89.982	93.297	96.801	96.315	93.012	98.985	98.189

FIG 18.comparision table for different algorithms.

ACCURACY: The degree to which the result of a measurement, calculation, or specification conforms to the standard value.That is standard value measurement is often called as accuracy
PRECISION: Rfinement in a measurement, calculation, or specification, especially that is represented by the number of digits given.Repeated for different values to get the precise value.
SPECIFICITY: The extent to which experiment or training is specific for a particular condition, trait, etc.
SENSITIVITY: The quality that are being easily influenced, changed. by a physical activiy or effect that is easily changed to the soecific reaction.

V. CONCLUSION

Various papers were studies and referred to the overall literature survey about the human pose estimation. Theories and previous research have been the basic reference in order to define the different pose estimation of an individual. The existing methods uses different methodology such, as HOG, PSM, SVM to detect, extract and classify the human pose. But these method does not provide an accurate result because they does not include more no of images, which will have an impact in recognition. The proposed methodology uses ANN in which target value are given equal to the input. Threshold values are fixed and therefore complexity is reduced. Multiple task learning auto encoder model so that input are perfectly recognized using auto encoder and then manifold recognizer reduces dimensionality. These advanced technique provided better accuracy than traditional methods.

VI. REFERENCES

1.Alexander Toshev., Christian Szegedy**.,” Deep Pose: Human Pose Estimation via Deep Neural Network”,** IEEE XPLORE 2012.

2. 2. Jun Yu., Chaoqun Hong., Member, IEEE, Yong Rui, Fellow, IEEE,and Dacheng Tao, Fellow, IEEE **“Multi-Task Auto encoder Model for Recovering Human Poses”**,IEEE transactions on industrial electronics,2011

3. Arjun Jain., Jonathan Tompson .,Christoph Bregler ., **“Learning Human Pose Estimation Features with Convolutional Networks”,**Office of Naval Research ONR.

4. Tomas Pfister., James Charles., Andrew Zisserman **.,” Flowing ConvNets for Human Pose Estimation in Videos,”** published on IEEE conference,2015.

5. Chaoqun Hong, Jun Yu, Member, IEEE, Jian Wan, Dacheng Tao, Fellow, IEEE, and Meng Wang, Member, IEEE**” Multimodal Deep Autoencoder for Human Pose Recovery”**, IEEE transactions on image processing, VOL. 24, NO. 12, DECEMBER 2015.

6. Matheen Siddiqui., and G´erard Medioni ., **“Human Pose Estimation from a Single View Point, Real-Time Range Sensor”,** IEEE transactions on image processing,2010.

7. Lu Xia., Chia-Chih Chen and J. K. AggarwalHuman**.," Human Detection Using Depth Information by Kinect",** IEEE transactions on image processing,2014.

8. Ijaz Akhte.,r Michael J. Black., **"Pose-Conditioned Joint Angle Limits for3D Human Pose Reconstruction"**,IEEE explorer,2015.

9. Seyoung Park., Bruce Xiaohan Nie., and Song-Chun Zhu., **"Attribute And Or Grammar for Joint Parsing of Human Pose, Parts and Attributes",** IEEE explorer 2010.

10. . Seyoung Park., Bruce Xiaohan Nie., and Song-Chun Zhu., **"Attribute And-Or Grammar for Joint Parsing of Human Pose, Parts and Attributes",**IEEE explorer 2010

11. Xiaowei Zhou., Menglong Zhu., Georgios Pavlakos., Spyridon Leonardos.," MonoCap: "**Monocular Human Motion Capture using a CNN Coupled with a Geometric Prior",** to appear in IEEE transactions on pattern analysis and machine intelligence, 2018

12. Wanli Ouyang., Xiao Chu., Xiaogang Wang., **"Multi-source Deep Learning for Human Pose Estimation",**IEEE explorer 2014.

13. Sam Johnson.,Mark Everingham**.," Learning Effective Human Pose Estimation from Inaccurate Annotation",** In Proc. CVPR, 2009.

14. Wei Yang., Wanli Ouyang.,Hongsheng Li .,Xiaogang Wang**.," End-to-End Learning of Deformable Mixture of Parts and Deep Convolutional Neural Networks for Human Pose Estimation",**IEEE xplore 2016.

15. Diogo C. Luvizon1., David Picard., Hedi Tabi., **"2D/3D Pose Estimation and Action Recognition using Multitask Deep Learning",**IEEE xplore 2010.