MEDIAPIPE CNN BASED PRETRAINED PATIENT BODY POINT IDENTIFICATION FOR UNCOMFORTABLE SLEEP

[1]B. Yogitha Singh, P.G Scholar, Department of E.C.E, NRI Institute of Technology, Pothavarappadu, Agiripalli (M), A.P, India Email: b.yogitha123@gmail.com
[2] K. Prathyusha, Associate professor, Department of E.C.E, NRI Institute of Technology, Pothavarappadu, Agiripalli (M), A.P, India Email id:prathyushakuncha@gmail.com
[3]Ashraf Ali Shaik, Associate professor, Department of E.C.E, NRI Institute of Technology, Pothavarappadu, Agiripalli (M), A.P, India Email id:shaik.ashraf1@gmail.com
[4]YalamanchiliArpitha, Associate professor, Department of E.C.E, NRI Institute of Technology, Pothavarappadu, Agiripalli (M), A.P, India Email id:yalamanchiliarpitha@gmail.com

Abstract

There exists an enormous amount of data within the internet where the information is travelled from one source to another destination simultaneously, it is almost beyond thought that this data can be analysed and retrieved in the search capability and processed for a specific use. While there exists several commercial systems that have been designed to build and retrieve the data from images and videos using the initial meta-data provided, they much lack in the description of the system. Content-based retrieval (CBIR) systems have the capability to retrieve the existing images based on the visual content but lacks in few parameters. Here we propose a method of retrieving the data from images and videos by estimating the pose of a human. We propose the structure using different algorithms such as human pose estimation algorithm which can estimate the pose of human being from a proposed image further, we estimate the pose by representing in the form of pose lets which could be estimated from body detectors and later we represent the lower dimensional pose-sensitive manifold by embedding an image and finally conclude the work by demonstrating on a real time video retrieval system which can similarly match with the 2-D pose of a human pose of a query. Using Random Forest of K-D Trees to locate nearby inquiries allows for a quick responsiveness.Designers demonstrate that, apart from the query modalities, a low dimensional representation is sufficient for posture retrieval. Second, we demonstrate that by pooling the findings of many human posture estimate methods, accuracy may be greatly enhanced. Using a variety of pose inquiries, the system's efficacy is measured statistically.

DOI Number: 10.14704/nq.2022.20.8.NQ44214 **NeuroQuantology 2022; 20(8): 1928-1939**

1 Introduction:

The ubiquity of video recording devices has a huge growth in storing the image and video data. This requires processing the data by decoding it and indexing by inputting the data of images and videosconcurrently which is termed generally has content-based information retrieval (CBIR). Here the automation is done particularly on the people who have enormous application in identifying them which can be further developed for various business-relatedimplementations such in supervision, archival salvage, etc. Human presence is one of the important concepts that try to convey some information about a particular person. Here we estimate and demonstrate the importance of them from searching the pose from animages and videos and demonstrate the need to human pose estimation.

There exist various methods [20] that can be used to search images from a database. All instances in the database that are comparable to a designated item in an image are retrieved in response to a query using [19]. The prediction of the image's object labels [24, 55] is another approach that may be utilised to create annotations. The image and video retrieval systems include three major building blocks that include the user interface which can take the input and display the retrieved results, Image/Video Descriptor Generator to imply the easiness of user interaction, the search and indexing algorithms to find the suitable matching for queries in the database system based on the entities. The primary objective involves in eliminating the unwanted information and utilizing only the known information and encodes them in terms of effectively and efficiently. Position and orientation of an item relative to a coordinate system is its "pose," while the spatial arrangement of a human body is called a "pose." Due to the intricacy of the human body, the issue of human pose retrieval presents significant difficulties. Human posture identification has been around for more than 40 years in the field of computer vision [16, 17]. Some issues that arise while trying to solve the human body pose problem includes the following: Color and texture variety: human beings have the ability to dress in a wide range of ensembles that include a wide range of both. Since this is the case, the HPE algorithm must not care how the input looks. Humans may appear anywhere in the picture and in any size. They can even change direction. Humans may appear in many performances in some contexts, such as sports. Thus, the HPE method must be robust to changes in position, orientation, and size. The human body isn't completely rigid; its various components have some range of motion, allowing for a certain amount of variation. The limbs, for instance, are free to move even if the rest of the body remains immobile. Therefore, accurate articulation modelling is essential for HPE. Our goal here is to create the first working model of a really comprehensive job-searching platform. When it comes to user input, the text query is by far the most common approach, thus that's where we placed most of our focus in our proposal. A human stance is far more difficult to describe, thus this is of little utility in that context. In contrast, query-by-image is widely used despite being time-consuming and limiting. This process requires the user to discover a picture through the desired stance, and then submit the image to the system in order to generate comparable positions.Given that this technology is the first of its kind, this seems counterintuitive, since there is currently no reliable method for acquiring photographs with the required stance. A very natural and easy query interface is a primary focus of our effort. Finally, there's the description generator, the lifeblood of every successful search engine. Word descriptors have proved particularly effective in picture and video search engines. Gradient data has always been recorded. The most significant drawback of these identifiers is that they do not allow for fine-grained spatial reasoning, which is essential for human position but is not yet achievable. Human Pose Estimation (HPE) methods provide an alternative by estimating where various body components should be. These algorithms, sadly, are notoriously error-prone, rendering them useless. The second objective of this study is to develop new, compact position descriptors that are based on HPE techniques and to enhance the dependability of existing ones.

2 Literature Survey:

For the study of human pose according to pose of the person body is well articulated in [11] which illustrate the model parts and its relations between them. For learning the human pose and belief propagation [7] of the structure could be well illustrated in structured output SVM which intrudes the human pose estimation algorithms [3]. For further understanding of pictographic picture structures Markov random field is utilized which can model spatial relations between the defined entities [10]. For building a Structured output SVMs its parameters and insightful works that being performed out in [6]. For the features scope of HOG [9] can be widely used.

Convolutional Neural Networks [2] works for a better case when it comes to visual tasks due to its internal detailed structure. Current techniques have been shown to be built on CNN architectures [1] that can simulate the combination of functions in images.Visually the model can be seen having the series of layers where the upcoming layer will be the function of the previous existing layer. The speciality of the CNN is that it first addresses the higher order layers present first later decides upon the lower-level layers. The facial expressions in the picture can be demonstrated using a spring model [12] which can be viewed has a graph starting from the forehead its left edge and ending till right edge and other parts of the body. The connections between them can be further demonstrated using graphs and line segment structures.

For assessment of pose which is necessary component can be given using human pose retrieval for determining the upper body detection using pose lets[13], ETHZ UBD [14]. Additionally, the estimation algorithms to estimate the human body components that are visible in the picture from the correspondence between the person and his human parts. Pose of the person estimation algorithms [5] further the review is stated in [22] where the pose is defined as input from the image or a video.

3 Existing structure:

The human body detection can be estimated using pose lets [13] which were earlier developed for person detection schemes and a classifier to model subset of body parts which require larger amounts of data and may not be feasible. Using the CNNs the structure can be well improved [14] and can detect the particular object using parts that constitute it using training data provided. The upper body evaluations used for pre-processingin HPE flow chart[5, 26, 80]anopenlyaccessibledetection is used[32] that can combine [35] and [98] for part-base model and face detector using running window estimation.

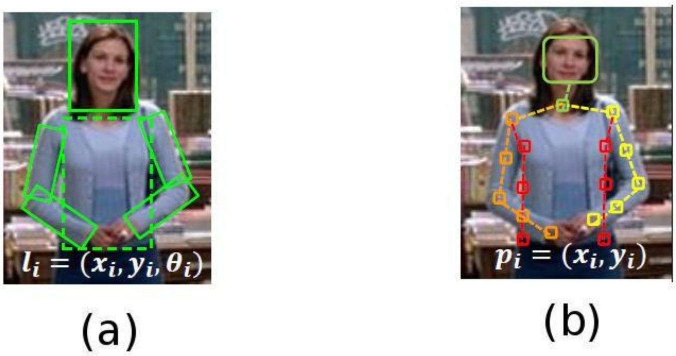

Fig 1 relation of the parts

Figure 1: The two different approaches to defining pieces and their relationships are shown graphically above. To illustrate (a), complete anatomical body parts are used as components in graphic constructions. The head, both limbs, and torso are all shown in the model. Each component's position and orientation constitute its state space. Joints and a few other locations on the body are represented as components in the graphic architecture of (b). Each component's state space would comprise its position and orientation. Both diagrams use rectangles to represent the various components, and the edges of adjacent rectangles linked by lines serve as edges between the two diagrammatic structures.

4 Algorithm

If there are several items in a picture, a single shot detector like YOLO can do so with only a single frame by using the multibox technique.

- START: Submit a Live-Streaming Video Sequence.

- STEP-1: Frame 1 bone joint estimation (t=0).
- STEP-2: Matching bone points and compare using Media pipe CNN architecture $(t = t + \Delta t)$.
- STEP-3: Inliers density estimate for inter-point density calculations Z(p).
- STEP-4: Computation or variation using sigma parameter estimation $\sigma\,(Z)$ i.e., standard deviation.
- STEP-5: Adaptive band based matching filtering i.e., S U Snew.
- STEP-6: Update body data points (p).
- STEP-7: Calculate distance between p & p i.e.t&$t -$. P or p −=p(x,y).

$$\sqrt{(x2 - x1)2} + \sqrt{(y2 - y1)2}$$

- STEP-8: If distance >Threshold&Δt > Time threshold (updates alert system) Else (continues tracking the patient).

5 Flow chart

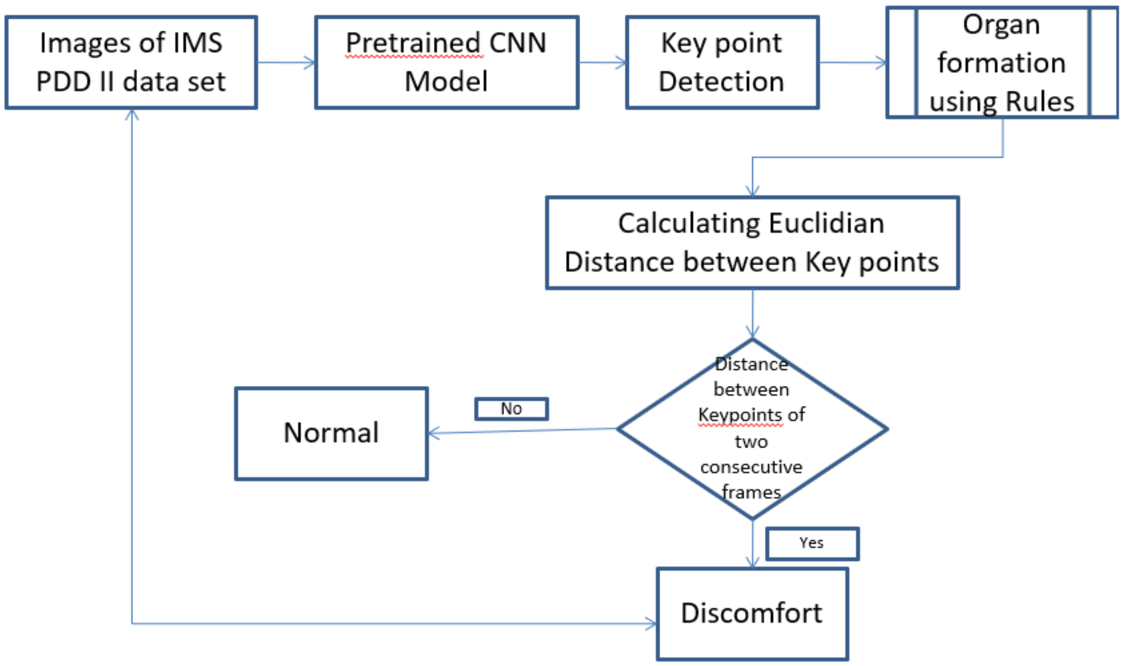

Fig2: flow chart

5 Proposed Architecture:

These datasets, which include the recently released Buffy2 Stickmen [50] and Movie Stickmen [50], the ETH PASCAL dataset [27], the H3D dataset [13], the FLIC dataset [79], the MPII Human pose dataset [4] and the Poses in the wild dataset [19], depict a person in a broad variety of positions, attires, lighting conditions, and scene backgrounds. The images in Buffy2 Stickmen were taken from episodes 2 and 3 of the fifth season of the television series Buffy the Vampire Slayer. It is intended that this collection will serve to both increase the quantity of postures available in the Buffy Stickmen dataset [39] and to augment the stances that are already there.

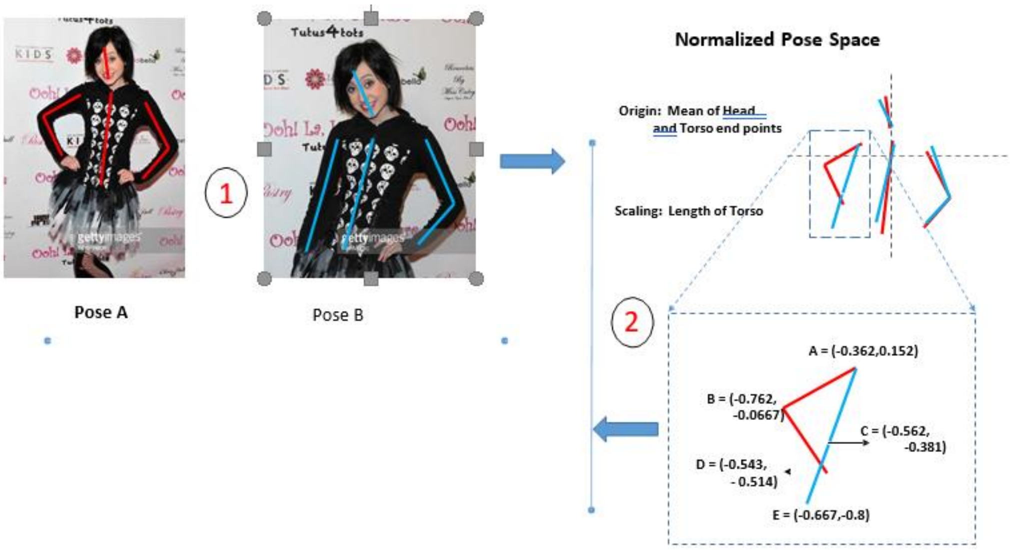

Fig 3 Pose similarity measures

Figure 3: Quantitative comparisons of poses Here, we use a pair of postures to calculate the principal component pair wise distance and angle difference. If we have two postures (1), we can map them into a standard pose space (2). This normalized posture space has its origin at the halfway point between the body's longitudinal axis and the axis of the head. The torso length serves as the scaling factor. Taking the averages of the various limb and body lengths is also performed in certain contexts. In the end, we have the PCP and angle difference measurements (3). The accompanying picture shows the precise formulae. Only the left lower arm's worth is shown here for simplicity's sake. The following measures for assessment are elaborated upon: Precision on average and the area under the receiver operating characteristic curve (ROC) are two important metrics.

Fig4 Ground-truth samples (stickman overlay) from Movie Stickmen (first row) and Buffy-2 (second row) (second row).

In order to assess the accuracy of human posture detections, we frame the issue as one of success and failure categorization. In order to capture the scenarios in which an HPE method is most vulnerable to error, we propose a set of characteristics. We classify characteristics as either I utilised directly by the HPE method (such as the HPE score, edgeallocation, and optimal arrangement of components L) or (ii) not implemented directly through the HPE algorithm (such as the location of the expanded detection window, its overlap with other detections, etc.). Next, we'll explain the differences between the two kinds of capabilities.

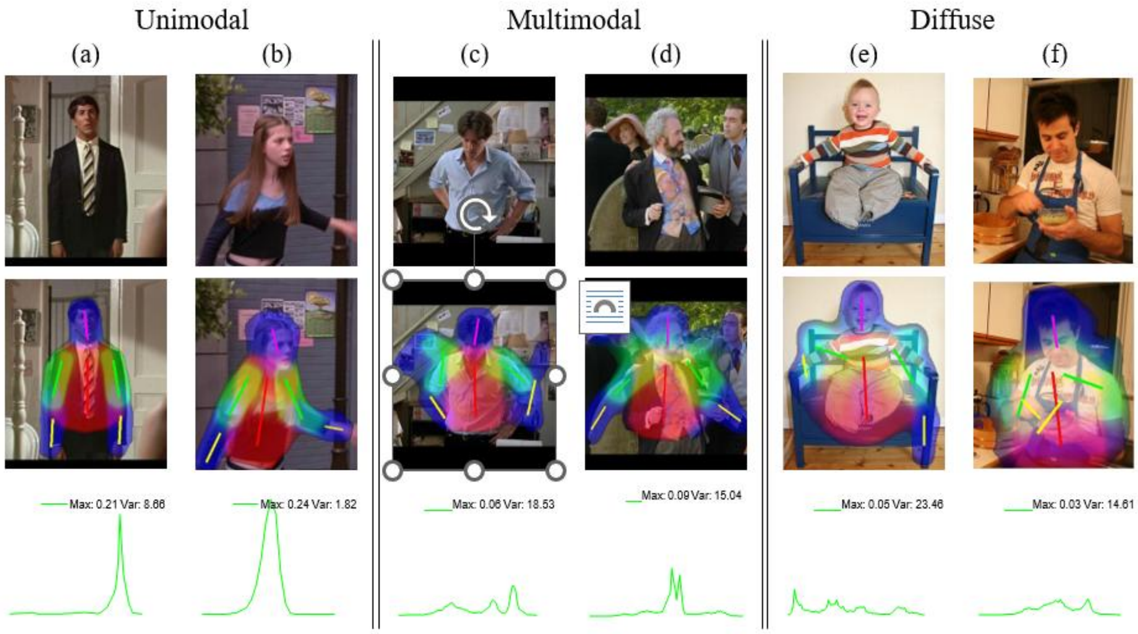

Fig5 Functions derived from HPE output

HPE-derived characteristics are shown in Figure 4. Here are some illustrations of single-, multi-, and wide-spread posture analysis. Semi-transparent masks based on the optimal configuration (sticks) and the posterior marginal distribution across the location of body parts are superimposed on each picture (first row) (displayed in second row). Examining this distribution allows us to put a numerical value on characteristics like "Max" and "Var" (third row).Unimodal distributions are the most common kind, and as shown in (a) and (b), the mode nearly invariably corresponds to the proper component placement. While one mode of a multimodal distribution will often correlate to the right portion position, as in (c), this is by no means guaranteed (d). Lastly, for distributions with a lot of scatter (e and f), the numerous barely distinguishable modes do not provide any information regarding the precision of component positions. When compared to bimodal and multimodal distributions, unimodal distributions are more indicative of proper assembly. The 'Max' and 'Var' options are what establish the distribution type. From a unimodal peak to a more diffuse multimodal distribution, both the highest value and the variance spread out.

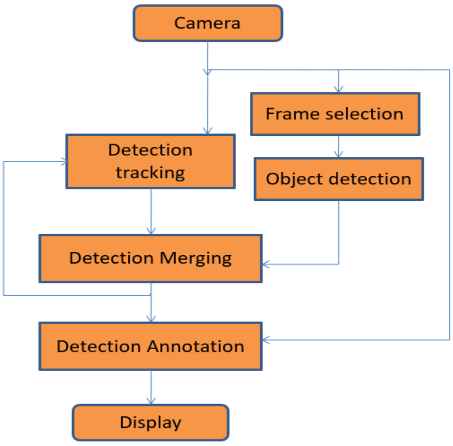

Fig 6 Proposed architecture

6 Results and Discussions

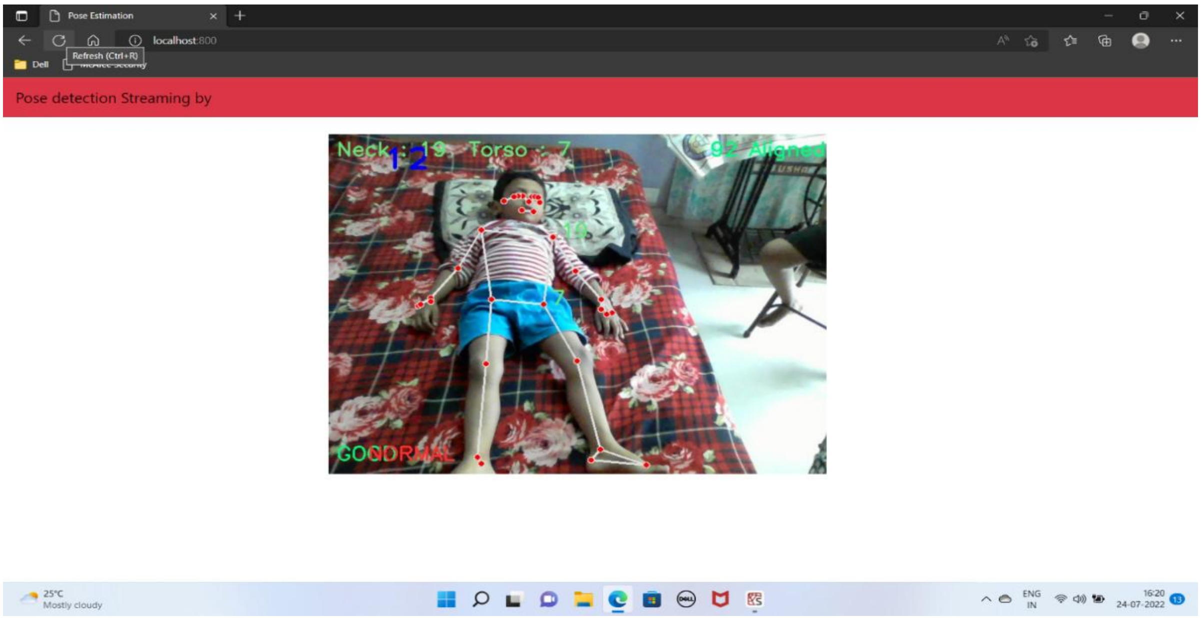

Fig7 : Patient 1: Detecting the patient condition as good because there is no movement in body parts

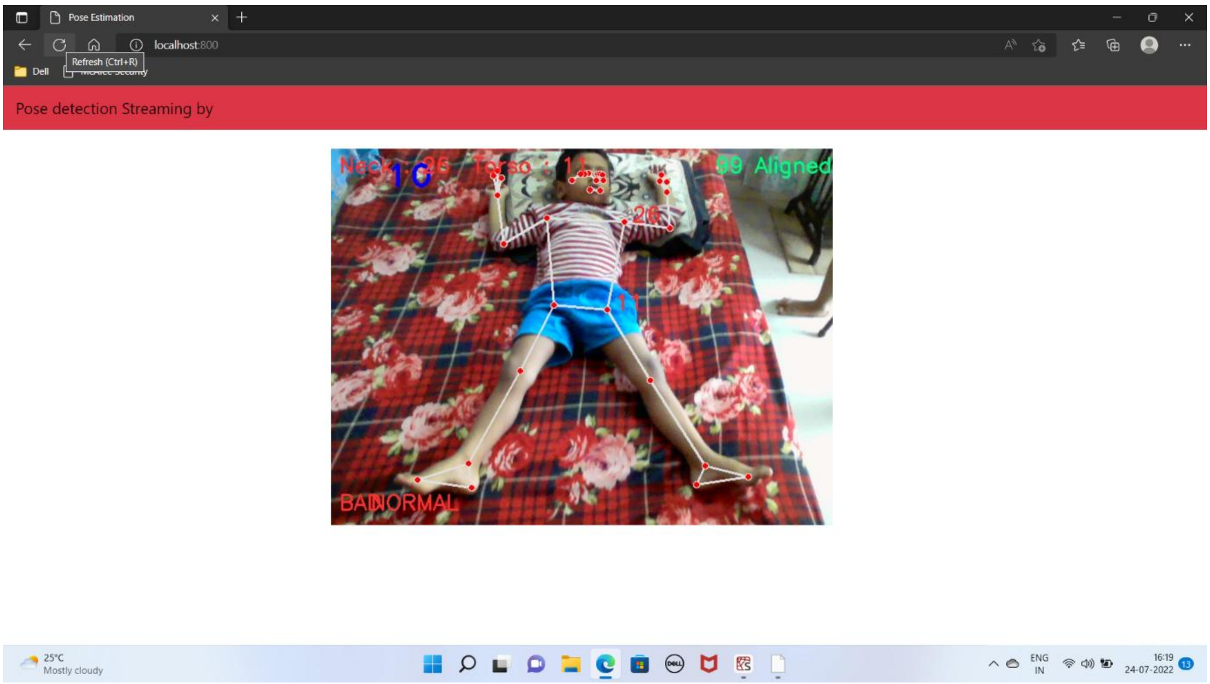

Fig8: Patient 1: Detecting the patient condition as bad because there is movement in body parts

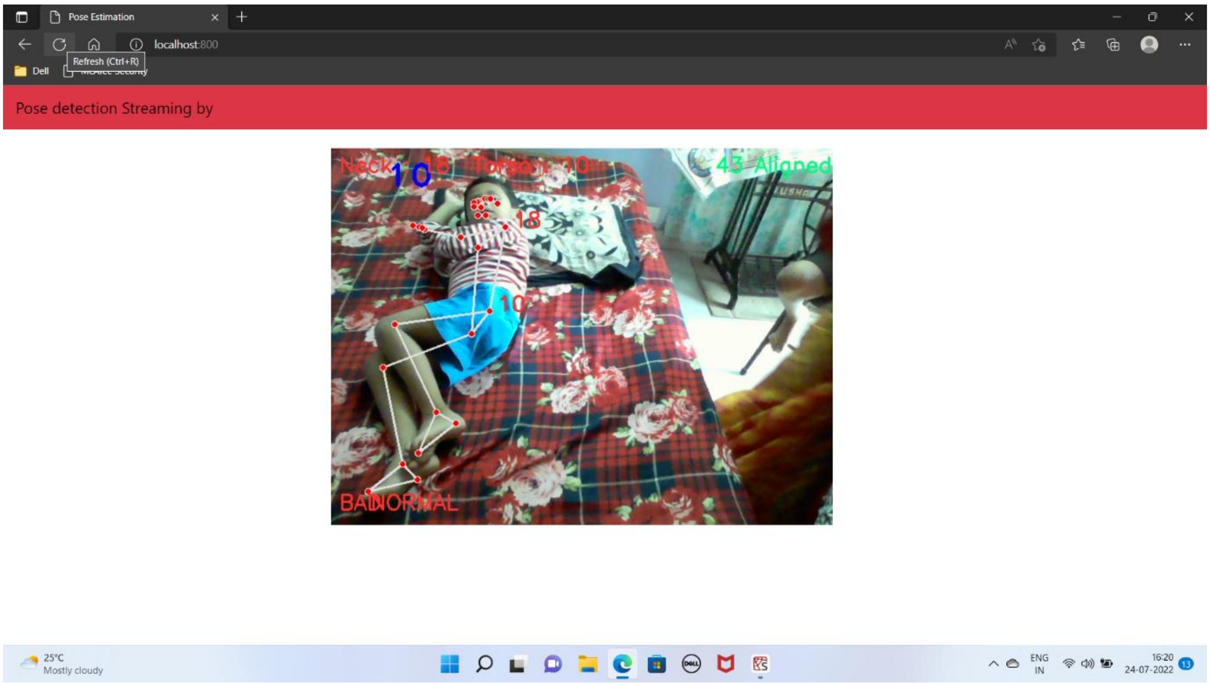

Fig9 : Patient 1: Detecting the patient condition as bad due to rapid movements in his hands and legs

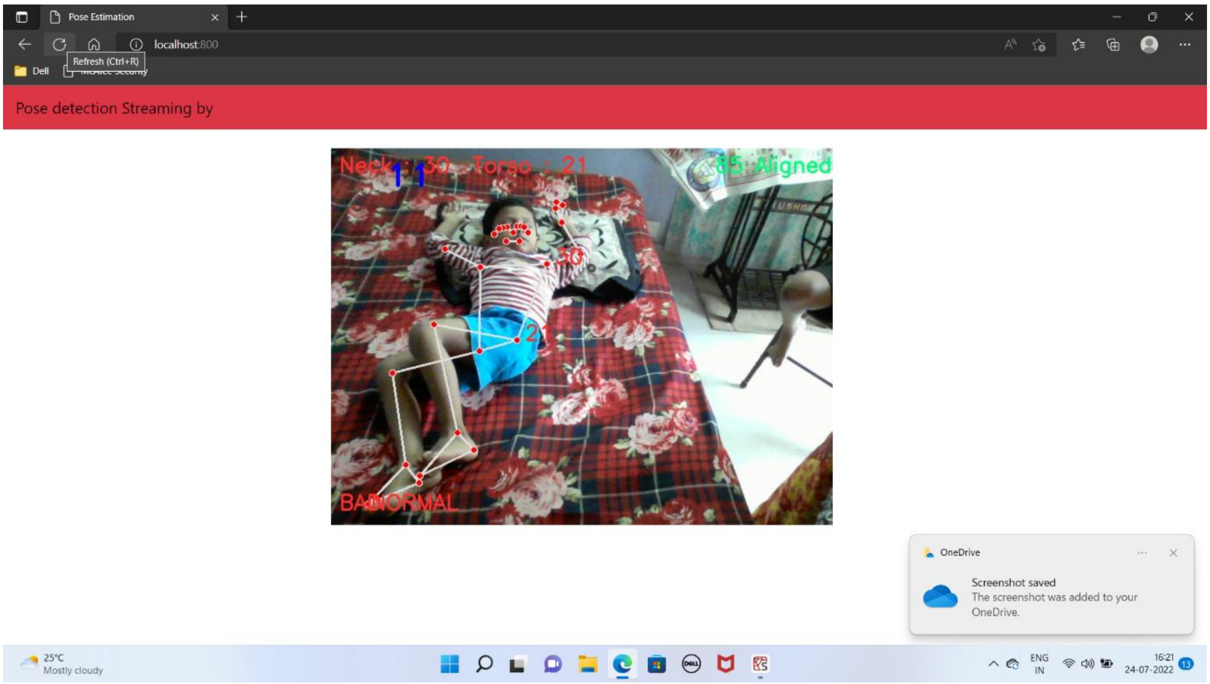

Fig10 : Patient 1: Detecting the patient condition as bad due to abrupt movement in his legs

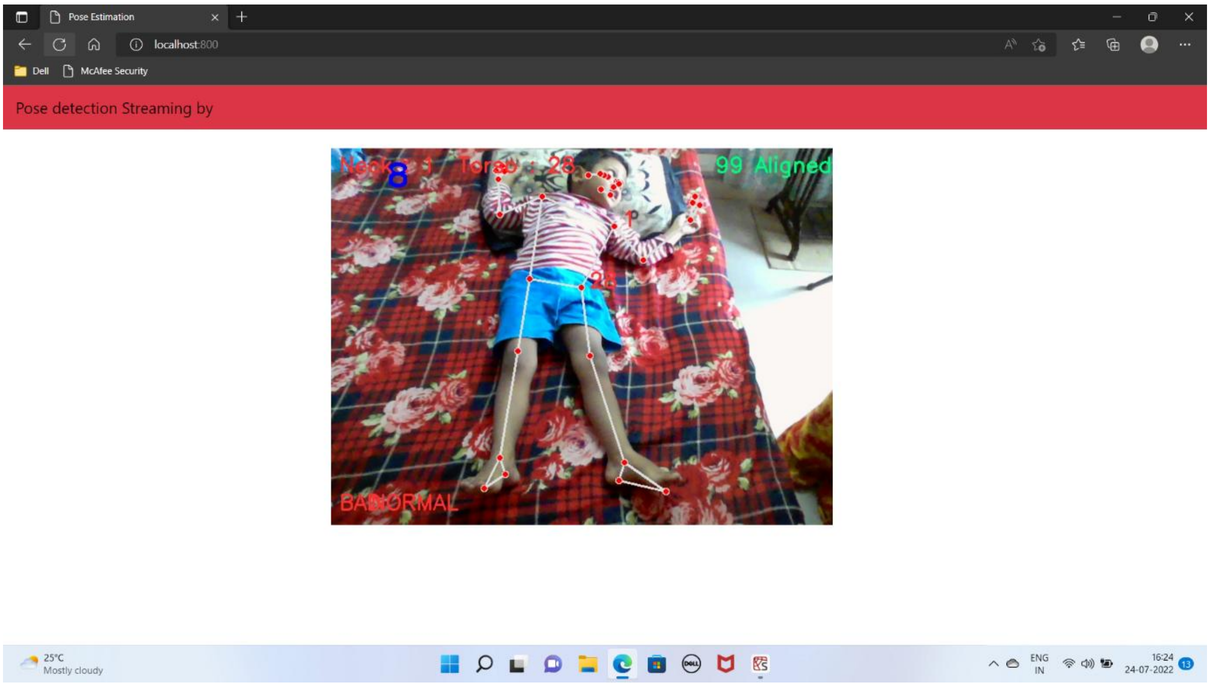

Fig11: Patient 1: Detecting the patient condition as bad due to abrupt movement in his hands

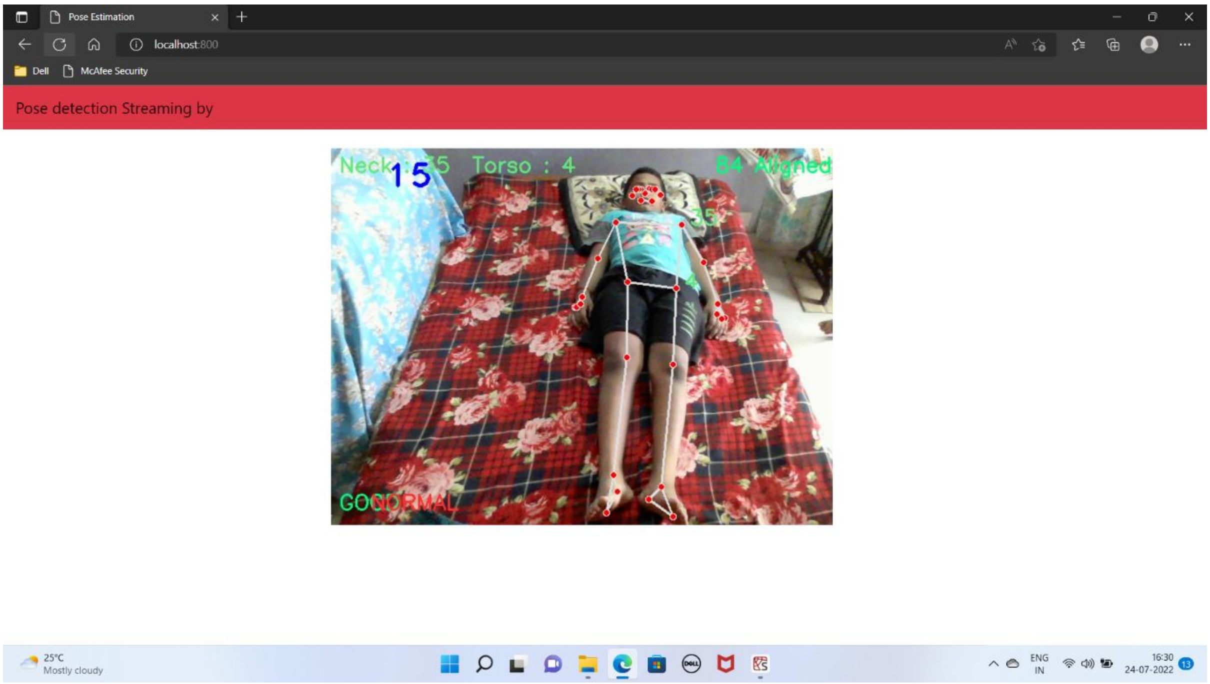

Fig12 : Patient 2:Detecting the patient condition as good because there is no movement in body parts

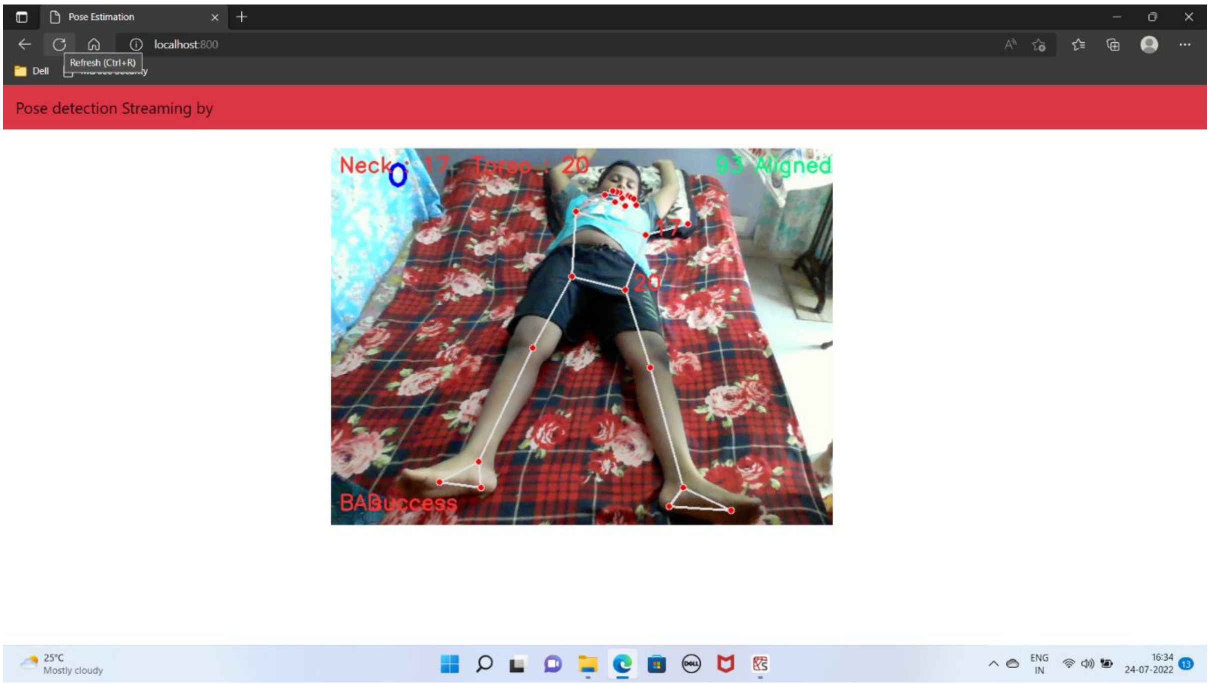

Figure13: Patient 2: Detecting the patient condition as bad because there is movement in body parts

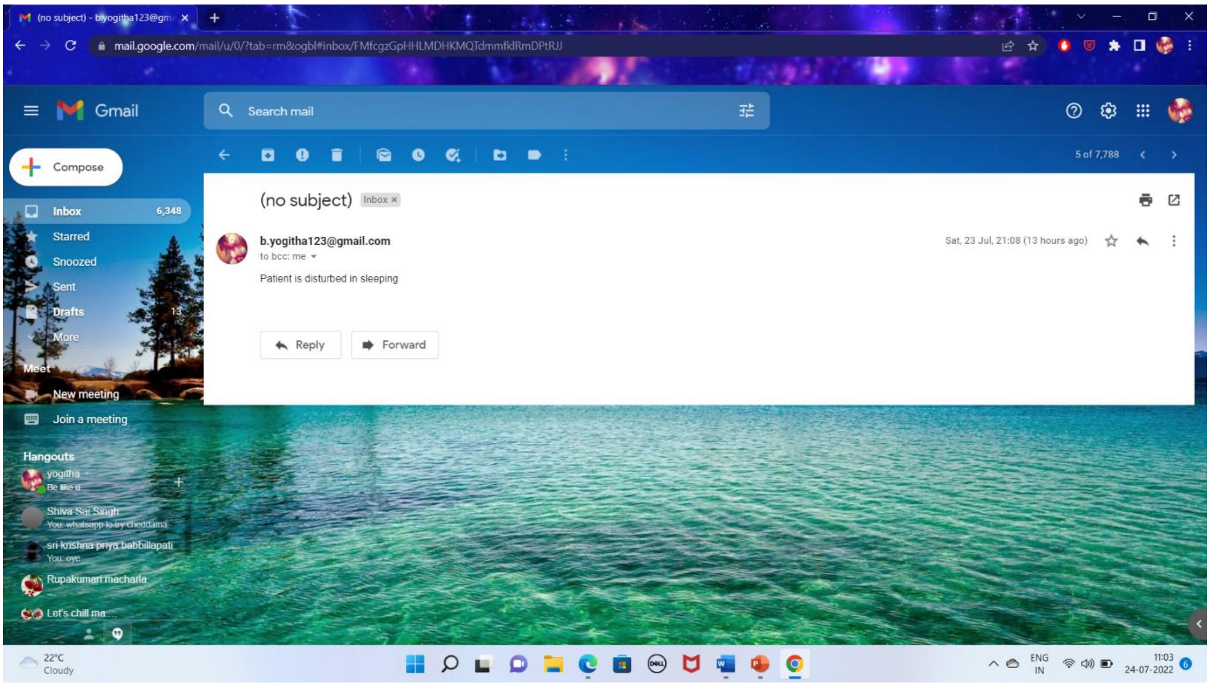

Figure 14:Alerting message sent to their relatives when we detect discomfort in the patient in the form of mail

Conclusion and future scope

Human posture conveys information about the person's actions and gestures, which is mostly relevant to forensics and chemistry. Enhancing content-based information retrieval (CBIR) systems via the use of images acquired with human posture is a top priority. In this paper, we create a real-time job search system and show that the results of many HPE algorithms together beat those of the individual methods, providing empirical evidence that picture retrieval utilising human posture is significantly more achievable than previously thought. Further next, we used the new test image to run all of these posture sets & maximum collect them to get 2nd position evaluator, which we used to characterise the position recoverymethod. We further analysed the projection function that maps an input picture to a position-sensitive splitter in a lower dimensional space, and provide empirical evidence that the resulting position descriptor compares well to the best available. Automatic Algorithm Evaluator, Deep Embedding, and Action Retrieval Systems will be useful in the future for developing this.

References

[1] G.Aggarwal,S.Biswas,P.Flynn,andK.Bowyer.Predictingperformanceoffacerecognitionsystems:Animagecharacterizationapproach.InProc.CVPRworkshops,pages52–59,2011.

[2] B.Alexe,T.Deselaers,andV.Ferrari.Measuringtheobjectnessofimagewindows.IEEEPAMI,34(11):2189–2202,Nov2012.

[3] S.Andrews,I.Tsochantaridis,andT.Hofmann. Supportvectormachinesformultiple-instancelearning.

[4] InAdvancesinNeuralInformationProcessingSystems(NIPS),2002.

[5] M.Andriluka,L.Pishchulin,P.Gehler,andB.Schiele.2dhumanposeestimation: Newbenchmarkandstateoftheartanalysis.InProc.CVPR,June2014.

[6] M.Andriluka,S.Roth,andB.Schiele.Pictorialstructuresrevisited:Peopledetectionandarticulatedposeestimation.InProc.CVPR,2009.

[7] O.Arandjelovic andA. Zisserman.Automatic face recognitionfor filmcharacter retrieval in feature-lengthfilms.InProc.CVPR,2005.

[8] F. Bastien,P.Lamblin,R.Pascanu,J.Bergstra,I.J. Goodfellow,A. Bergeron,N. Bouchard,and Y.Bengio.Theano:newfeaturesandspeedimprovements.DeepLearningandUnsupervisedFeatureLearningNIPS2012Workshop,2012.

[9] S.BelongieandJ.Malik.Matchingwithshapecontexts.InCBAIVL00,pages20–26,2000.

[10] J.Bergstra,O.Breuleux,F.Bastien,P.Lamblin,R.Pascanu,G.Desjardins,J.Turian,D.Warde-Farley,andY.Bengio.Theano:aCPUandGPUmathexpressioncompiler.InProceedingsofthePythonforScientificComputingConference(SciPy),june2010.OralPresentation.

[11] A.Blake,P.Kohli,andC.Rother.MarkovRandomFieldsforVisionandImageProcessing.MITPress,2011.

[12] M.Blank,L.Gorelick,E.Shechtman,M.Irani,andR.Basri. Actionsasspace-timeshapes. InProc.ICCV,2005.

[13] M.BoshraandB.Bhanu.Predictingperformanceofobjectrecognition.IEEEPAMI,22(9):956–969,2000.

[14] L.BourdevandJ.Malik. Poselets:Bodypartdetectorstrainedusing3dhumanposeannotations. Proc.ICCV,2009.

[15] L.D.Bourdev,F.Yang,andR.Fergus.Deepposeletsforhumandetection.CoRR,abs/1407.0717,2014.[15] J.Bromley,I.Guyon,Y.LeCun,E.Säckinger,andR.Shah. Signatureverificationusingasiamesetime delayneuralnetwork.InNIPS,pages737–744,1993.

[16] C.J.C.Burges.Atutorialonsupportvectormachinesforpatternrecognition.pages121–167,1998.[17] K.Chatfield,K.Simonyan,A.Vedaldi,andA.Zisserman.Returnofthedevilinthedetails:Delvingdeep intoconvolutionalnets.arXivpreprintarXiv:1405.3531,2014.

[17] X.ChenandA.Yuille.Parsingoccludedpeoplebyflexiblecompositions.InComputerVisionandPatternRecognition(CVPR),2015.

[18] A.Cherian,J.Mairal,K.Alahari,andC.Schmid.Mixingbody-

partsequencesforhumanposeestimation.

[19] InProc.IEEEConferenceonComputerVisionandPatternRecognition,2014.

[20] S.Chopra,R.Hadsell,andY.LeCun. Learningasimilaritymetricdiscriminatively,withapplicationtofaceverification.InProc.CVPR,pages539–546,2005.

[21] N.DalalandB.Triggs.HistogramofOrientedGradientsforHumanDetection.InProc.CVPR,volume2,pages886–893,2005.

[22] J.Deng,W.Dong,R.Socher,L.-J.Li,K.Li,andL.Fei-Fei.ImageNet:ALarge-ScaleHierarchicalImageDatabase.InProc.CVPR,2009.