International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Research Paper
Available online at: www.jrrset.com
UGC Approved Journal No: 45483
Received on: 23.03.2018 Published on: 30.03.2018 Pages : 34-42

ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
Volume 6, Issue 3, March 2018.
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

GAIT ANALYSIS AND MACHINE LEARNING FOR UNIQUE BIOMETRIC IDENTIFICATION OF INDIVIDUALS (GMLUBI)

[1]Greshma KP, [2]V Vineet , [3]Rishav Gupta, and [4]P. Mohamed Fathimal
[1,2,3] UG Scholars, [4] Assistant Professor
Department of Computer Science and Engineering
SRM Institute of Science and Technology
Corresponding Author: greshmapai@gmail.com

Abstract

Human identification has become a flourishing area where different techniques are used for identification and verification, each of them having separate and exclusive claims and challenges. Gait analysis refers to the study of animal locomotive behaviour which includes body mechanics, and the activity of muscles. It has applications in various fields such as medical diagnostics and osteopathic utilizations. A unique application of gait analysis is biometrics and identification. This paper describes a representation of gait analysis by identifying people based on their gait patterns. The steps include background modeling, feature extraction, and processing on neural network. The background model is built by getting input from a high-resolution camera which detects the movement. This input in the form of raw RGB video frames which is then passed to a neural network which measures gait biomechanics. This helps in feature extraction of the biometric information such as pace, step angle, height from torso to feet and hand displacement. The created data set will be processed, followed by identifying humans by their unique gait pattern using SVM classifier and neural network. This network consists of two subnet works linked in cascade- Joint identification and tracking (for pose estimation) and Gait recognition and estimation (for gait recognition). Using machine learning in combination with traditional, computer-based gait analysis, one can perform biometric identification while learning the user’s locomotive style.

Keywords: gait analysis, feature extraction, neural network.

1. Introduction

The use of electronic devices has significantly increased in the past decade and has made our life fast, convenient and easier. Since most of these devices are connected to the internet and their numbers are increasing day by day it is important to protect our information. Many authentication methods are used to enhance the security. One such method is biometric authentication, most of which need the continuous intervention of the user and the system. Common methods among them being iris, face, fingerprint and voice. The best use of using biometric authentication is that the user does not need to remember the password. Gait authentication or recognition has an additional benefit beyond typical biometric authentication mechanisms. It is entirely unobtrusive, meaning that there is no active user interaction with the authentication system. Gait recognition extracts and analyzes gait patterns from users for authentication purposes [8].

Human gait not only means the "a manner of walking" but also a distinguishing feature of a person which includes weight of the individual,footwear, limb length, and posture collaborated with characteristic motion. Additionaly, the definition of gait also include certain characteristcs of the appearance of the person, such as the clothing, the period and phase of a walking cycle, the aspect ratio of the torso, the amount of arm swing etc[1]. It has applications in various fields such as medical diagnostics and chiropractic and osteopathic utilizations. It is used in medical diagnostics especially in sports biomechanics to help athletes run efficiently. Apart from this, there are other applications like biometric forensics for gender identification. It is mostly used for detection of the people while walking and human recognition in videos in many applications, such as video surveillance, traffic monitoring and human motion capture . The gait pattern of a person includes gait velocity, stride length, and stance and swing phase times changes with age.ie., gait details of a person at a younger age will be different during old age and these changes can be determined to reduce the frequency of falls of elderly by identifying diagnostic measures that are reliable predictors of falls and to prevent it.

One of the advantages of using this as an identification method is that gait can be detected at a low resolution when compared to iris or face information which needs high resolution for recognition and can be done at a distance. The challenges intricate in this study include changes in the clothing of the subject, variations in the camera viewing angle with respect to the walking subjects, and changes in gait as a result of mood or speed change, or as a result of carrying objects. Although the mentioned challenges exist, it permits some changes in clothing and other minute changes which do not change the gait substantially.

In this paper, we present a new method by using gait as a biometric application by recognizing humans by their gait. The gait motion of the human body is extracted from the image sequences using a high-resolution camera and is processed using a neural network and thereby calculating gait sequences for analyzing the gait motion by their gait. As such, the new system aims to derive measures of biometrical significance, in part demonstrated by recognition capability.

2. Related work

L.Lee [1] has proposed the accuracy of recognition using gait video sequences collected over different days and times and under varying lighting environments and this result is used for gender classification based on gait appearance.[2] and [3] the application of the gait analysis is explained in elderly. The gait patterns changes with age and these changes can be determined to reduce the frequency of falls by identifying diagnostic measures that are reliable predictors of falls and to prevent it and to analyse data for disease identification and classify them in various categories such as Neurological and Neuromuscular Diseases or Juvenile Idiopathic Arthritis using gait analysis data and identify this at an early age. A. Sokolovaa and A.Konushina [4] proposed a deep learning approach using the optical flow as a source of motion information and combine neural feature extraction with the additional embedding of descriptors for representation improvement. In [7] Win Kong and et al artificial neural network (ANN) based classification of human gait. This gait comprising of six states are classified based on the similarity of the lower limbs' figure and the state of gait is beneficial to real-time human tracking and occlusion handling. In [5], an algorithm was developed that is capable of detecting human presence and moving objects widely used in human tracking and monitoring video surveillance using SAD (Sum of Absolute Difference) and background subtraction and background scene modeling followed by foreground object detection. In [6], Zoran Zivkovic has proposed a pixel level approach for background subtraction using Gaussian Mixture Model(GMM).

3. Neural Network Architecture

The input for the neural network is raw RGB video of pedestrian and the output is a one-dimensional vector which is also known as gait descriptor. The gait descriptor of two persons should be linearly separable. The whole system can be divided into two different sub-networks connected in series. Subnetwork-1 performs the joint identification and tracking while subnetwork-2 performs gait identification and classification.

Subnetwork-1 performs joint identification and extracts spatial features from input video frames. The features are then stored as 2D pose vector which is then fed into Subnetwork-2. The responsibility of the second subnetwork is to convert the 2D pose into a 1-dimensional feature which can be linearly separated from other 1D pose descriptors. These features are aggregated using Average Temporal Pooling which is then used to classify test subjects by their gait(using SVM classification). Subnetwork 2 performs the gait identification. The input for this network is the output of the first subnetwork which is a one-dimensional pose descriptor. The one-dimensional pose vector should be in the center of the frame for an efficient output.

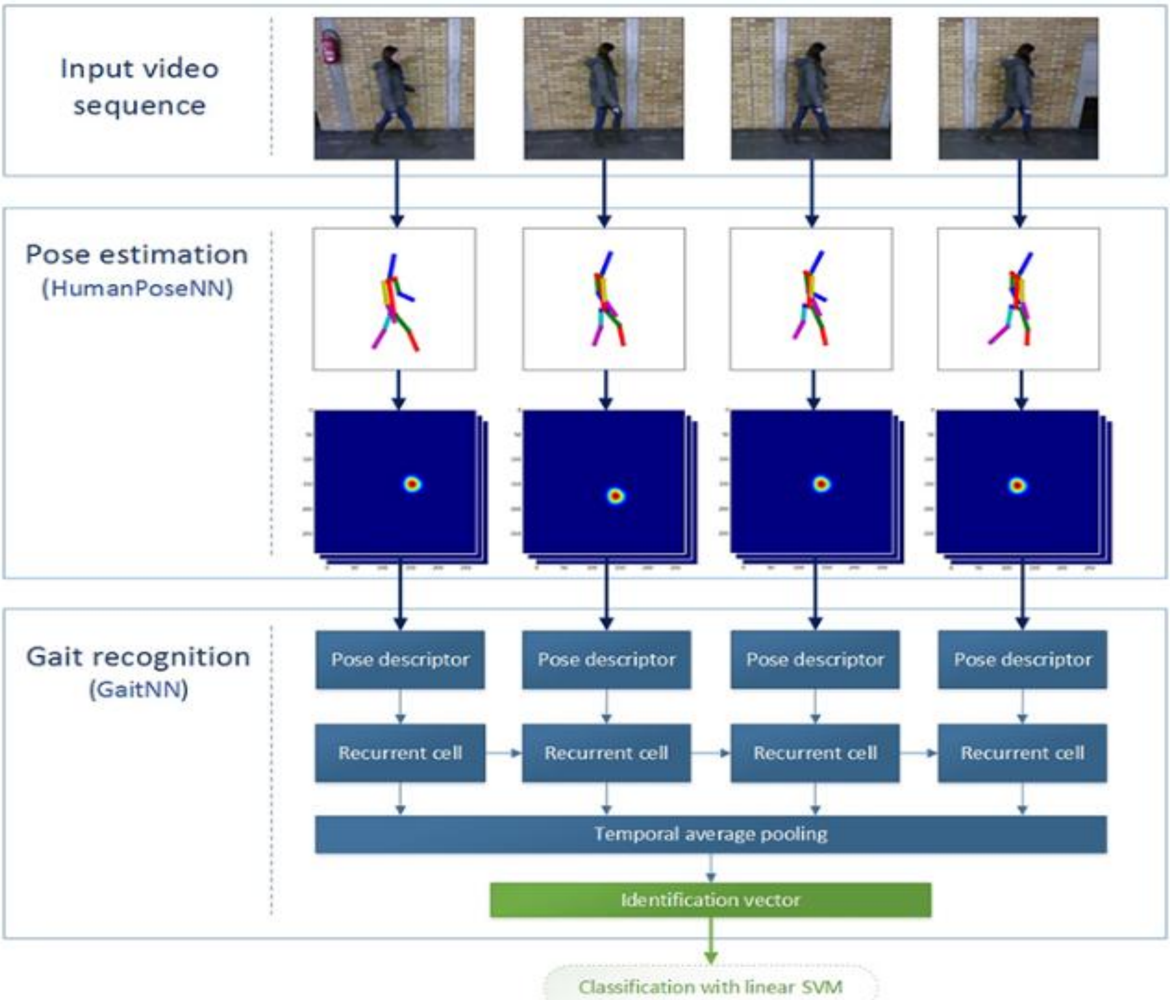

Fig 1Architecture diagram of the neural network

4. Methodology

The methodology consists of the following steps.
1. Data acquisition
2. Joint extraction and tracking
3. Output and classification

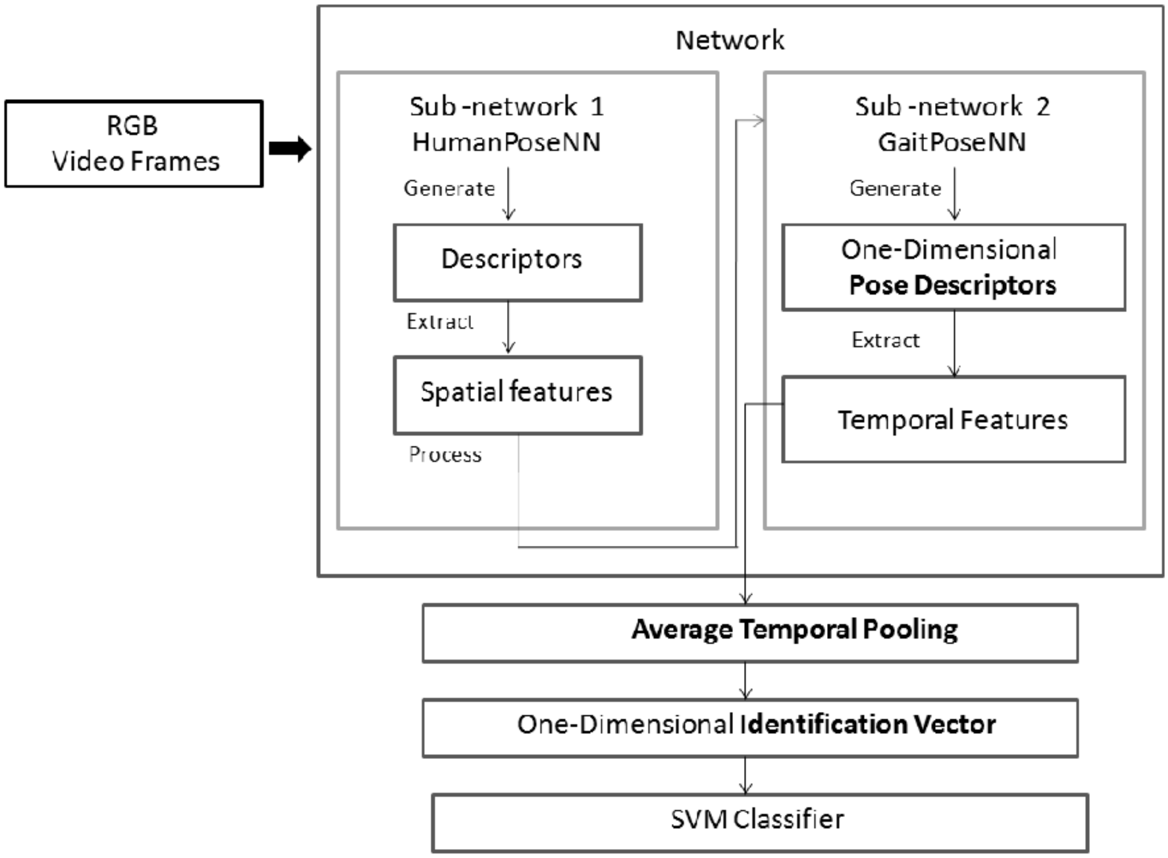

Fig 2.workflow of the video processing

4.1 Data acquisition

An image is obtained from the camera which is used as input video sequences where a person's walk is recorded. Unlike the other methods of identification, this one is difficult since the human walking styles of people will be similar and the difference is hard to notice[19]. Once the data set is obtained it is then passed to the neural network for processing. Two features are extracted from and individual. Spatial features is based on space (motion based on arms and legs) whereas temporal features is based on time(speed of motion). Identification vector is a combinaton of both spatial and temporal features.

4.2 Joint extraction and tracking

Human 2D pose estimation can be done using two approaches: Top-down approach and bottom-up approach. The top-down approach detects the person and executes single person pose estimation. The major disadvantage of this approach is if it fails to detect a human in the frame. Then the recovery of the human pose is very difficult also the runtime of the system will increase with the increase in the number of human in the frame.

The bottom-up approach, on the other hand, provides a remedy to the above mentioned approach and also has the ability to reduce the runtime complexity . In practice, previous bottom method does not retain the gain in efficiency as the final parse requires costly global inference. But the proposed approach that joins and labels the part detected candidate and associates them to individual people. But solving an integer linear programming problem over a fully connected graph is an NP-hard problem and average execution time is per hour.

The input to the network will be the raw RGB video frames of a pedestrian. This produces 1D vector-gait descriptor, which exposes an identification vector (fig 3) that would be thoroughly separable, reason which the gait of two-person is always linearly separable. As a result, the gait descriptor produces the one-dimensional vector as an output. The two types of network used are:

Subnetwork 1 – Joint identification and tracking

Subnetwork 2 - Gait recognition and estimation.

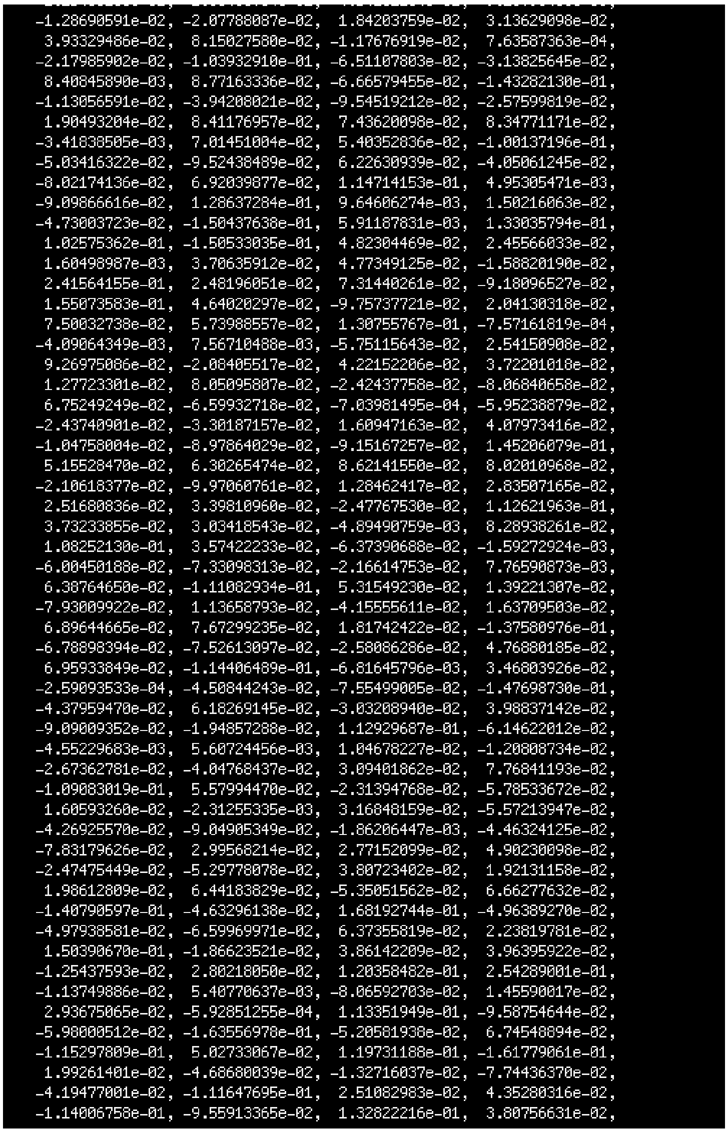

Fig 3: identification vector

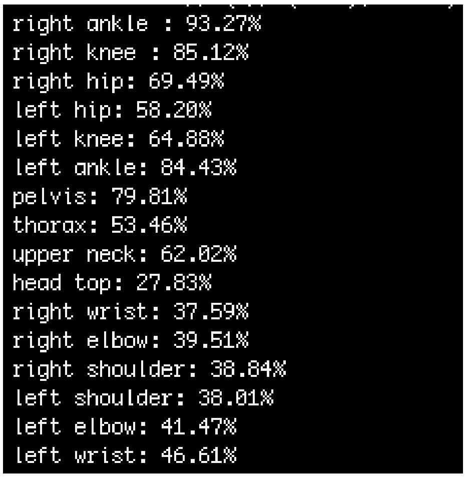

Fig 4: joint estimation

4.2.1 Subnetwork-1

An open pose is the first multi-person system to detect the human body, hand, and facial key points on a single frame. The functions of the open pose is to detect real-time multi-person key points. This includes estimation of key points of the body, hand, and face.
The input for the neural network is raw RGB video with size wxh. It is fed into subnetwork 1 where the human pose is estimated using open pose algorithm which further gives a 2-dimensional figure of the human pose. The system is divided into two branches, the top branch projects the confidence maps s, bottom branch projects the affinity fields.

4.2.2 Subnetwork-2

Spatial features of the pedestrian such as pose of the person will be generated from subnetwork 1 by the descriptor. Further, pre-processing of the spatial features of subnetwork 1 into one-dimensional pose descriptors with the help of the Residual convolutional network. From the one-dimensional pose descriptors uses of the multilayer recurrent cells (LSTM OR GRU), the temporal features are extracted and recorded. Average of all the temporal features is taken using the average temporal pooling and each of this will be termed as 1-D identification which will be continuously separable with each and every different person and finally classifies using SVM classifier (fig 5). SVM classifier is supervised model used for classification and regression.

4.3 Output and classification

Thus the obtained one-dimensional identification vector is then classified using SVM (Support Vector Machine) classifiers. It is used for analyzing the data used for classification which helps in identifying an individual. The workflowof the system is shown in fig 2.

4.3.1 SVM (Support Vector Machine)

Support vector machine or SVM is based on the concept of decision planes that defines decision boundaries. A decision plane separates a set of objects having a different class membership. In the fig 5 given below objects belong to GREEN or RED and the line separating the green object and red object is called the decision boundaries.

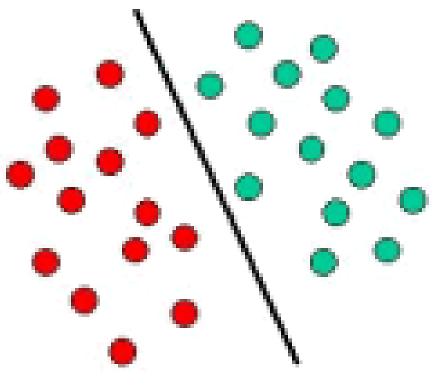

Fig 5: example for SVM

This example separates the object linearly i.e.it separates the object into the respective group. Sometimes this classification is complex and hence we need complex structure to classify the object in an optimal classification. Classification tasks based on drawing separating lines to distinguish between objects of different class memberships are known as hyperplane classifiers. Here the gait identification vector from the neural network will be the input to the SVM and will be plotted in the decision planes of every person and the SVM will classify the object of same characters in a group using the hyperplane classifiers. Hence the gait identification vector can be classified to identify the vector of the corresponding person.

Fig 6: Overall pipeline.

Our method takes the entire image as the input for atwo-branch CNN to jointly predict confidence maps for body part detection, shown in (b), and part affinity fields for partsassociation, shown in (c). The parsing step performs a set of bipartite matchings toassociate body parts candidates (d). We finally assemble them into full bodyposes for all people in the image (e).

5. Detection And Association

The method of working with the subnetwork is first it gets a frame (of size wXh) as an input and produce a 1D representation of anatomical key points on the human body. First, the network creates a 2D confidence map S of the body part location and then creates a 2D vector field L of affinity field. The set S =(s1,s2,s3……sj) has j confidence map one in each part. The set L=(L1, L2….Lc) has c vector field in every part .finally both confidence map and affinity field are concentrated together to obtain a 2D representation of the human pose.

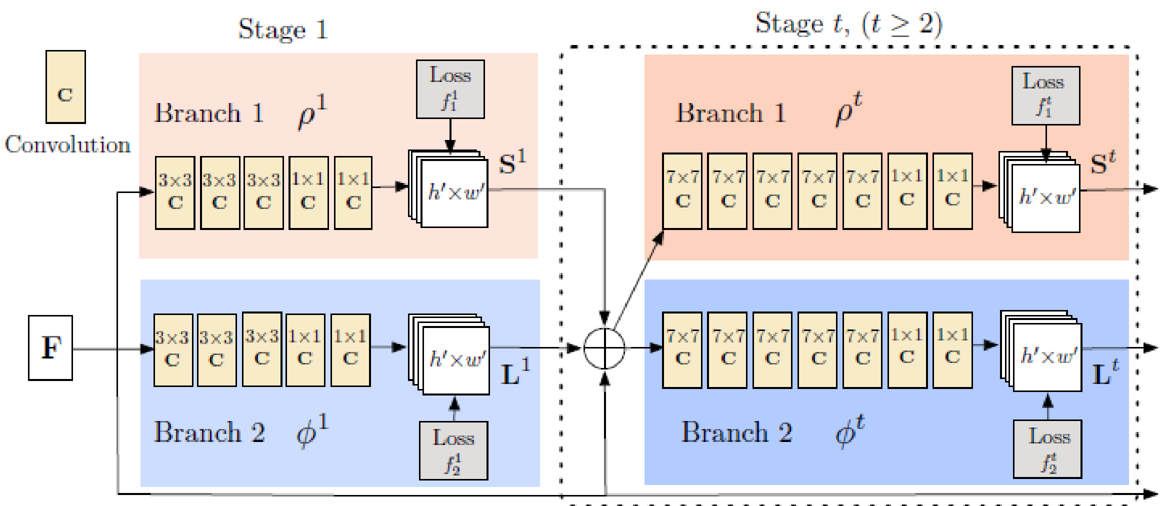

Fig 7: The architecture of the two-branch multi-stage NN.

Each and every stage in the first branch predict confidence maps St[1], and each stage in the second branch predicts PAFs Lt. After each stage, the predictions from the two branch.

The neural network architecture can simultaneously predict the detection of confidence map and part affinity field working in series. In other words, the network can be divided into two branches: branch 1, predicts the confidence map and branch 2 associates the part affinity field The architecture keeps refining the prediction in each and every successive stage t∈{1,....T}.The image is first examined by a network and tuned to produce a set of feature maps F which will act as input to the first stage of each branch. At the first stage, the network generates a set of confidence map maps S 1 = ρ 1 (F) and a set of part affinity fields L 1 = φ(1), here ρ 1 and φ 1 are the NNs for inference at stage 1.In each and every stage derivation of every branch is merged with the original image features F and the output is obtained.

$S^{t}= \rho^{t}(F,S^{t-1},L^{t-1}),\ ),\forall t\geq 2,$ (1)

$L^{t}=\emptyset^{t}(F,S^{t-1},L^{t=1})\ ),\forall t\geq 2,$ (2)

To supervise the network to iteratively predict confidence maps of body parts in the first branch and affinity fields in the second branch, we apply two loss functions at the end of each stage, one at each branch respectively The loss function at both branch at stage t is derived from

$f_S^t = \sum_{j=1}^{J} \sum_p W(p) . \|S_j^t(p) - S_j^*(p)\|_2^2,$ (3)

$f_L^t = \sum_{C=1}^{C} \sum_p W(p). \|L_C^t(p) - L_c^*(p)\|_2^2,$ (4)

$S_j^t$is the groundtruth part confidence map L_C^t is the groundtruth part affinity field,

W is a binary mask with W(p) = 0

The intermediate supervision at every stage gives a vanishing gradient problem by restoring the gradient periodically.the overall goal is:

$$F= \sum_{t=1}^{T}(f_S^t + f_L^t) \quad (5)$$

5.1 Confidence Maps

A confidence map is a 2D representation of the body part in a frame at each pixel location. Preferably a single confidence map should occur for the single person in a frame. But a frame with many people there should be a peak for every body part j for every person k. we can generate confidence map for each person at a location by using:

$S_{j,k}^*(p)=\exp\left(-\frac{\|p-x_{j,k}\|}{\sigma^2}\right)$ (6)

where σ controls the width of the peak

The ground truth confidence map to be projected by the network is an assembly of the singular confidence maps via a max operator,

$S_j^*(p)= \max_k S_{j,k}^*(p).$ (7)

Instead of average we are considering the maximum so that the precision of the nearby peak remain distinct.

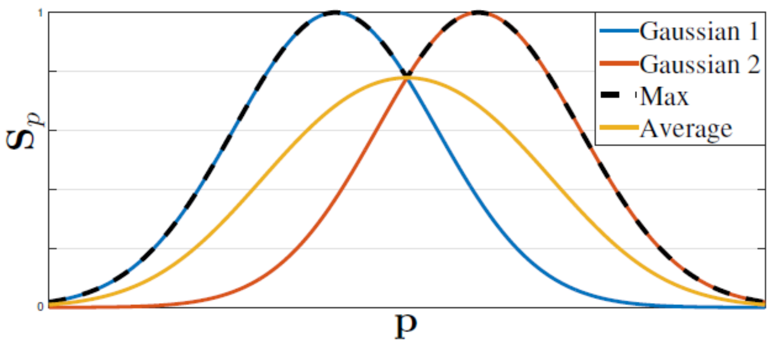

Fig 8 Diagram representing the Gaussian Curves.

During testing confidence map and produce body part candidate doing non maximum suppression.

5.2 Part Affinity Field for Association (PFA)

Once the body part is detected we need to convene them to create a full body pose. And to associate them we can detect an extra midpoint between each pair of the body parts.But there is some limitation to these association they are (1) it conceals position and not the orientation of each part.(2) the region of support is reduced to a single point.

To overcome this problem we have a feature called part affinity field which helps in storing position and orientation at the region of support of the body parts.the affinity field is 2D vector field for each body part for each pixel present in the body part. This vector contains the direction from one point to another point if the detected body part.

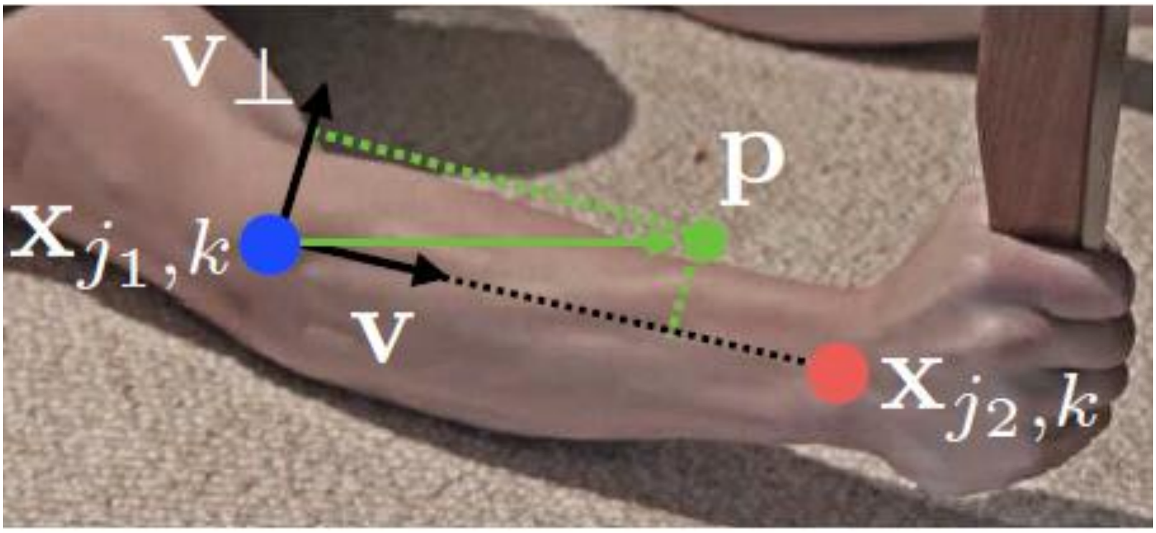

Fig 9 Figure showing representation of points

Let $x_{j1,k}$ and $x_{j2,k}$ be the inferred position of the parts j1 and j2 from limb c of a person k.if a point P lies in the body part .then the value at L ∗ c,k(p) is a unit vector that points from j1 to j2; to find Fl d, L ∗ c,k, at an image point p as

$$L^*_{c,k}(\mathrm{p})=\begin{cases} v & if\ p\ on\ limb\ c,k \\ 0 & otherwise \end{cases} \quad (8)$$

The average of the grounf truth affinity field and the affinity field of all the person in the frame is

$$L^*_c(\mathrm{p})=\frac{1}{n_c(p)}\sum_k L^*_{c,k}(\mathrm{p}), \quad (9)$$

Association of candidate part detection can be done by calculating the line integral on corresponding PAF.for two candidate part location s dj1 and dj2 ,the predicted part affinity field Lc can be measured by

$$E=\int_{u=0}^{u=1} L_c(\mathrm{p(u)}).\frac{d_{j2}-d_{j1}}{\|d_{j2}-d_{j1}\|_2}\,du, \quad (10)$$

6. Conclusion

In the proposed model, the collected data of every individual was processed to get the gait using techniques like open-pose algorithm which contained two neural network which in-turn gave a one dimensional vector. One dimensional vector called as identification vector, is further passed to SVM for classification and regression where the spatial and temporal feature for individual unique

7. Future Work

This system can be further implemented in surveillance in a crowd to identify the person. Methodology for the state and matter of subjects such as change in clothing or a backpack, object being carried by the person, way the subject views the camera in a different way every time, and changes in the pace according to the mood of the person(drunk, disgust, hypnotised etc can be further be analysed and implemented in the proposed system .

References

[1] L. Lee ,W. E. L. Grimson,MIT Artificial Intelligence Lab,Cambridge, MA 02139 “Gait Analysis for Recognition and Classification”

[2] Ranveer JOYSEEREE, Rami ABOU SABHA, Henning MUELLER.,”applying machine learning to gait analysis data for disease identification”.

[3] Gait in the elderly,F Prince, H Corriveau, R Hebert, DA Winter - Gait & posture, 1997 - Elsevier

[4] A. Sokolovaa, A.Konushina, Gait Recognition Based On Convolutional Neural

[5] S.sulaiman, A Hussain,n Tahir, SA Samad, M M Mustafa “human silhoutte extraction using background modeling and subtraction techniques”

[6] Zoran Zivkovic “Improved Adaptive Gaussian Mixture Model for Background Subtraction”

[7] Win Kong, Mohamad Hanif Saad, M A Hannan, Aini Hussain, "Human gait state classification using artificial neural network".

[8] Ik-Hyun Youn, Sangil Choi, Richelle LeMay, Douglas Bertelsen, Jong-Hoon You, New gaIt metrics for biometric authentication using a 3 axis acceletraction”