International Journal of Future Generation Communication and Networking
Vol. 13, No. 3, (2020), pp. 2618–2641

A Deep Convolutional Neural Network (DCNN) and Squirrel Search Algorithm (SSA) based Classifier Framework by Extracting Human Body Skeleton points based on Silhouette Images for Human Action Recognition in Image Processing

Ratnala Venkata Siva Harish[1], Dr. P. Rajesh Kumar[2]

Research Scholar[1], Professor[2]
Department of Electronics and Communications Engineering
AU COLLEGE OF ENGINEERING (AUTONOMOUS)
VISAKHAPATNAM-530 003-ANDHRAPRADESH.

Abstract

Due to the development of cost-effective depth sensors and rapid poses estimation algorithms, skeleton-based action recognition is widespread. Historical approaches based on pose descriptors often fail with large-scale data sets as engineered features are limited. In this paper we intend to reinforce the geometric connections between joints to identify behavior. Three basic geometries are incorporated: joints, edges and surfaces. For action detection, the DCNN-based network uses a novel perspective transition layer and time dropout layers to learn robust images. Consequently, we propose using the Squirrel Search Algorithm (SSA) algorithm in order to make the Deep Convolutional Neural Network (DCNN) classification more efficient. In this report, we extract human body skeleton based on silhouette images by using a distance gradient, to classify crucial points from silhouette images that play a significant role in the recognition of human activity. Experiments of 3d large-scale measuring identification datasets show that joints, edges and surfaces for specific behavior are efficient and complementary. Our solutions greatly exceed present state-of-the-art strategies for identifying functions.

Keywords: *Deep Convolutional Neural Network (DCNN), Squirrel Search Algorithm (SSA), skeleton-based action recognition and silhouette images.*

1. INTRODUCTION

Physical processes are of phenomenal scientific and practical importance [1] and are focused on computer vision. It offers many important theoretical values, including a video surveillance system in real time, sport events interpretation and human computer interaction, and a wide range of possible applications [2]. The identification of human behavior has been a hot spot for study in recent years. Trivializing of human operations in video sequences [3] was a very difficult job due to issues such as the size, the point of view, illumination and the look of people. Analysis of human works from video sequences or photographs is a difficult task [4].A broad spectral range-recognition framework is needed for several implementations involving video monitoring systems, human-computer interaction, and robotics for characterizing human behaviour [5]. In human-to-human interacting and interpersonal interactions, it also plays a crucial role. It is hard to acquire as it includes information about a person's appearance, emotional intelligence and psychological status [6]. One of the major focuses for the study of machine learning and computer vision is the living organism ability to perceive others' actions [7].

A variety of methodologies of human identification including image segmentation, extraction of feature techniques, classification detection techniques and so on are applied in literature [8]. Segmentation methods to identify individuals include pattern mixing, foreground recognition and context subtraction, but with many humans in the scene [9] such methods have ended in failure. Moreover, several techniques for extracting features such as a gradient oriented histogram (HOG), HWF, haar-like features, movement features, edge features, ACF, ISM [10], etc. are used for human activity recognition. These methods of extraction do not work well if living organisms are not easily identifiable or if their postures are very distinct. We also noted the considerable improvement in classification results for human activities in the selection of relevant features [11]. The classification of human

ISSN: 2233-7857 IJFGCN

activities comprises two main phases: the collection and/or extraction of insightful features and implementation of an algorithm of classification [12].

A required feature set will significantly decrease the classification algorithm workload in such a program and even with a small biased feature set an efficient classification algorithm will operate effectively [13].Computational complexity of a grading algorithm is indeed a restricting problem [14]. Consequently, it might not have been the most sustainable by enhancing the classification algorithm as it is the appropriate way of handling the issue as it might cause difficulties to classify the data as rapidly as it is gained for some sign handling problems [15]. But by selecting any appropriate way of designing the classifier, system efficiency can be enhanced by enhancing classification accuracy [16].

In this work we intend to improve the reliability of the identification of individuals by enhancing the system for classifying raw photographs from the human body skeleton collected on the basis of silhouettes (pose) Photographs such as walking, playing, dancing, etc. We agree in general that the arbitrary influence of the details on features cannot be entirely manipulated by individuals or a single identification technique. As the recognition problem grows in complexity, the vulnerability of single recognition strategies becomes apparent, particularly as multiple forms of behavior and/or similarities occur. Consequently, we suggest that the Deep Conventional Neural Network (DCNN) classification [24,25] should have a precise and efficient human action classification by using the Squirrel Search Algorithm (SSA) [26] algorithm. Thus, we retrieve human body skeleton centered on silhouette image by applying the distance gradient to identify crucial points in silhouette images that play a significant part in the detection of human activity. MATLAB follows the method proposed and the experimental findings will indicate the efficacy of the method proposed.

The remaining article is structured as defies: Section 2 describes various works pertaining to the use of human action recognition. Section 3 describes the proposed methodology on Skeleton based representations for action recognition. Section 4 illustrates the Deep Convolutional Neural Network (DCNN). Section 5 describes the proposed Squirrel Search Algorithm (SSA). Section 6 presents the simulation results. Section 7 defines the conclusion.

2. RELATED WORK: A BRIEF REVIEW

There are numerous research works based on the human action recognition by analyzing various types of features and classifications are carried out by the research scholars. Some of them are reviewed here in this section.

Chen et al. in [17] proposed a human action recognition system which works on the basis of a sensor fusion method, run in real time, concurrently using a depth camera and an inertial sensor. Two efficient, collaborative representative classifier robust and efficient depth image characteristics and inertial signals are provided. A fusion was then carried out at the decision-making stage. A multimodal, readily accessible identification of human behavior was used to test the evolved real-time framework by evaluating a wide variety of human activities.The performance evaluation for the developed real-time system was > 97%, which was at least 9% higher than for the individual application of each sensing mode. The tests, both offline and in real time, exemplify the reliability and real-time performance of the method.

A new spatial time vector (ActionS-ST-VLAD) execution stage to combine informative deep features across the entire video, based on video segmentation adaptive function segmentation and segment sampling adaptive feature (AVFS-ASFS) was proposed by Zhigang et al [18]. By using AVFS-ASFS, the ActionS-ST-VLAD security features have been selected and the correct deep features are separated automatically into segments that have features for a temporally compatible Behavior in ActionS.Then, a flow-guided warping technique was applied to detect and delete redundant characteristic charts, based on the keyframe attribute extracted in each section, and the insightful ones are clustered with their exploited weight. They also use the RGBF model to capture RGB images relating to the execution activity in motion-scale regions. Four public benchmarks-HMDB51, UCF101, Cinetics, and ActivityNet-were comprehensively experimented for assessment. The results showed that their

ISSN: 2233-7857 IJFGCN

approach has effectively pooled valuable profound features spatially and has resulted in state-of-the-art video action recognition efficiency.

In[19] Wang et.al tried to reinforce the geometrical relationships between junctions for practice. The three primitive geometries were initiated: joints, borders and surfaces. Therefore, the three inputs were accommodated by a common end-to - end RNN-based network. In the RNN-based network, a new layer and temporal drop-outs were used to recognize action to obtain robust images and they first defined frame wise for action detection and then use a new multi-scale sliding window algorithm. Experimental studies in large-scale 3D behavior recognition data sets indicated that joints, boundaries and surfaces for various actions have been successful and analogous. They surpass current state-of-the-art approaches for both action identification and action detecting tasks significantly.

Ping et al. in [20] suggested that latent states have composite features and that they incorporated a novel CLS model that described and considered 3D human behavior of skeleton sequences. A hierarchical graph was used to model a human action that describes the action sequence as sequential atomic action. A composite latent state consisting of a latent semantic attribute and a latent geometrical attribute was described by an atomic action. To know model parameters and the latent composite structures of human behavior a biased, EM-like algorithm was proposed. Considering a 3D skeleton sequence, iterative programming algorithms of the composed attribute were proposed to identify the action and deduce a latent temporal structure of the action. The approach was tested for the three complicated 3D action dataset: MSR 3D Action Dataset, 3D Event Dataset Multiview and UTKinect 3D Dataset. Extensive experiments on these data sets show the efficiency and benefit of the method proposed.

The method for human action detection from depth maps and posture data via Convolutional neural networks (CNN) has been presented by Aouaidjia et al in [21]. For action representation, two input descriptors were used. First was an illustration of a distance movement, which gathers successive depth maps of a human activity, and second was a proposed term for motion joints, which displays the orientation of the body joints over time. Three CNN channels were instructed with different inputs in order to increase extraction for the precise action classification. Deep movement imaging (DMIs) was instructed for the first channel, and both DMIs and moving joint descriptors were trained for the second channel, and only moving joint descriptions trained on the third channel. For the final action classification, the action forecasts provided by three CNN channels have been combined. They recommend multiple fusion score operations to optimize the right metric. Experimental studies show that the results of the fusion of 3-channel output are better than the use of one or two channels.

Human action recognition (HAR) is commonly used to establish relationships between humans and machines in the field of environmental protection. It is not possible to ask people to act unnaturally in these applications. The algorithm must be modified and the communications must be made seamless as soon as possible. Sid Ahmed Walid et al. [22] suggested a new method, using skeleton information provided by RGB-D cameras, in order to improve the existing algorithms in relation to these points. Early estimation and more reliable viewing variation were accomplished and the method was successful. To achieve this objective, a new descriptor was proposed called body directional speed and a real-time classification was conducted. Experimental results on four criteria indicate that our method remains competitive with different HAR algorithms on the basis of a skeleton. They also demonstrate that their system for human behavior is ideal for early recognition.

Muhammad et al. in [22] has implemented a feasible multimodal fusion approach that uses data from various sensors, like RGB image, depth-sensor and portable inertial sensors to provide reliable human activity detection. They have extracted data from RGB-D video camera and inertial body sensors for computational efficiency. Those features include densely extracted histogram of oriented gradient (HOG) features from RGB-depth videos and wearable sensor data statistical signal attributes. A UTD-MHAD multimodal action data set comprised of 27 different human actions was used to evaluate the proposed HAR framework. For training and validation the proposed HAR fusion model K-nearest neighbor and supportive vector machine classifiers were used. The experimental findings revealed that greater identification outcomes than state-of-the-art tests were obtained in the suggested system. With a precision rate of 97.6 percent, the functional level fusion of RGB and inertial sensors represents the detailed best performance for the proposed system.

ISSN: 2233-7857 IJFGCN

3. PROPOSED METHODOLOGY

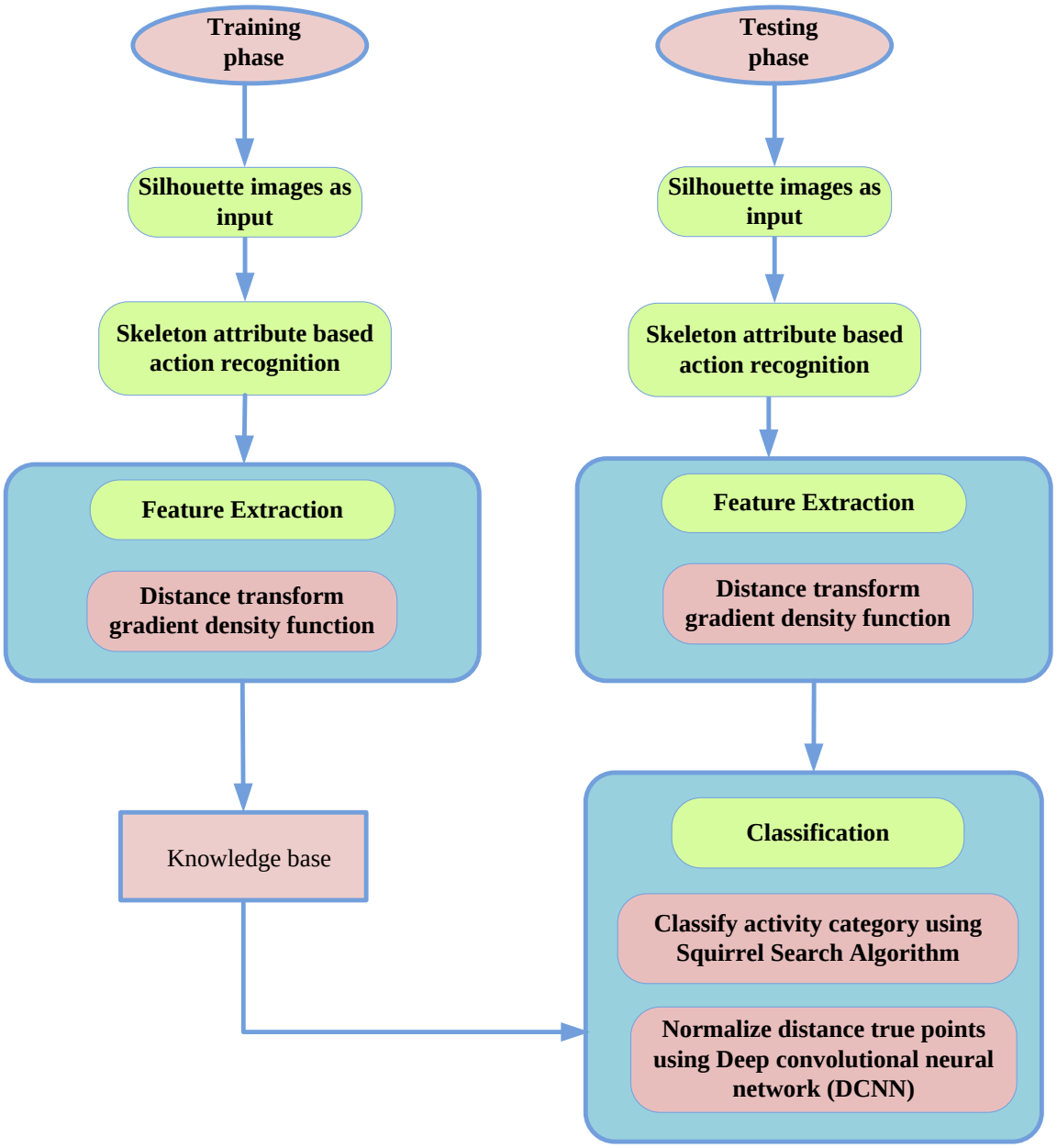


Figure 1: Overall structure Human recognition system

3.1 Skeleton based representations for action recognition

Skeleton data is a list of 3D coordinates of the aspects that make away the mutilated structure of the human body. When the body is moving intentionally, several movements occur during an action. The internal configuration of the body joints may be associated with these criteria. A graph with joints as points and the bones as edges can be viewed as a structure of the body. Figure 3 provides an outline of the creation of associated bones in 3D space during an operation. Given an individual, two geometric barriers are statistically significant for the skeletal system. First, the distance from two neighboring points along a connected segment is fixed as the width of the bone is relatively stable. Second, there are three perspectives that comprise two converging sectors on the same plane. Depending on the observations, the skeleton data exemplifies three types of information: the isolated joints of the human body, the edges denoting the connected segments, and the surfaces encompassing the intersecting segments. In order to optimize the capacity of deep networks to learn representations from raw data,

ISSN: 2233-7857 IJFGCN

function engineering approaches should be avoided and simplified primitive projections embraced. The description will be described as follows:

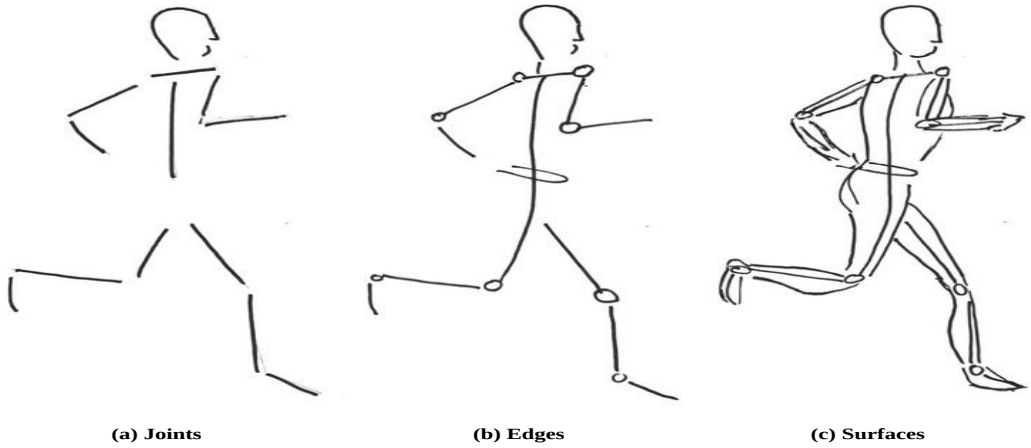


Figure 2: Physical structure of human body primitive data as input for DCNN

3.1.1 Joints

Figure 2(a) shows the isolated joints. Consider that for the structure of the human body there are S joints, and that the co-ordinates of points form a matrix $S \times 3$. When the length of a series is R , a tensor A with dimensions $R \times S \times 3$ can be defined as a skeleton. The joints' co-ordinates, which change over time, reflect time dynamics of actions. In reality, Johansson's experiments [31] demonstrate that many key joins provide the visual system with ample details on human activity in appropriate combinations of proximal movements. The majority of previous approaches actually rebuild A to a $R \times 3S$ matrix by reducing the second dimension. Several DCNN variants are utilized to recognize interpretations and human behavior as R changes for different sequences.

By using a rotation matrix, the joint coordinates of a given factor of perspective can be transformed into a slightly different perspective. When coordinate vector of a joint b_i is at a given R point in time the new coordinate vector can be acquired by:

$$\tilde{b}_i = Vb_i \tag{1}$$

Where, V is the 3-to-3 dimensional revolving matrix.

According to an input set, V is the similar to various joints and time - series. The new tensor experienced from a different viewpoint for tensor joint A can therefore be mathematically formulated as:

$$\tilde{A} = A \times_3 V^R \tag{2}$$

If $\times_3$ denotes tensor multiplied by a 3-mode, the magnitude of $\tilde{A}$ and A is equal.

3.1.2 Edges

The bone activity patterns differentiate between behaviors in relation to the spatial structure of the joints. The actual relations between body joints are described in a graphic representation. The articulations are labeled with edges, and the edges are labeled with borders. $S-1$ Edges are available considering a graph of S nodes. The edge shows primarily the position of the bone. For simplicity, as seen in Figure 2(b), it indicates the orientation of the edges. Every node has a vector of the coordinates, and any edge is defined by extracting the vector from the end point. For computing,

ISSN: 2233-7857 IJFGCN

$$d_i = b_k - b_l \tag{3}$$

Where d_i is the edge coordinate vector, b_l is the start point coordinate vector, b_k is the end point coordinate vector.

The skeleton edges can be denoted by tensor B with measurements $R \times (S-1) \times 3$. In Figure 1(b) we can also use the edge vectors to represent the edges. In fact, the vector of the edge terminating at the node will define a node. We indicate it by a zero vector for a node without end points of edges (e.g. hip-spine joint in Figure 2(b)). Therefore, the second dimension of B is expanded by one and the dimensions of A and B is equal. The edge coordinate vectors of a certain perspective can be synchronized and changed even by rotating into another perspective matrix. The change can comfortably be ascertained based on equations (1) and (3) for the coordinate vector of an edge at a given time sequence:

$$\tilde{d}_i = \tilde{b}_k - \tilde{b}_l = V_{d_i} \tag{4}$$

Where $\tilde{d}_i$ represents the converted edges of d_i. Subsequently, it can be represented in 3-mode tensor multiplication form:

$$\tilde{B} = B \times_3 V^R \tag{5}$$

Where $\tilde{B}$ refers to converted tensor edges. By relating equations (2) and (5), it is found that they have similar rotation matrices of joints and edges.

3.1.3 Surfaces

The edges style the contiguous relation of the joints. It is difficult for two joints to identify the condition that are parallel to the same joint that has two consecutive boundaries. In addition, the relative motion of neighboring bones also helps to distinguish behavior. We use the usual vector to define the plane, since two neighboring boundaries form a level surface.

Assume d_k and d_l be the vectors corresponding to the consecutive edges, the normal vector p_i is:

$$p_i = d_k \times d_l \tag{6}$$

Where $\times$ refers the 3D space cross product. We are not normalizing the vector here since the magnitude represents the passing angle of the respective sides. We easily calculate it by a constant of 100 to preserve the scale of the standard vector similar to the coordinate vector.There are $S+2$ surfaces for human body with S joints overall. We remove two surfaces with similar details (the normal vector that correspond to other normal vectors) to allow a rational analogy with joints and edges. The S surfaces described in Figure 2(c) are therefore functional. Thus, a tensor D in dimensions $R \times S \times 3$ can also be described as the normal vectors of a series.

The normal function can be viewed from a single perspective from a certain angle. The new normal surface vector is formulated at a certain point on the basis of equations (4) and (6) as:

$$\tilde{p}_i = (Vd_k) \times (Vd_l) = Cf(V)(d_k \times d_l) = Cf(V)\, p_i \tag{7}$$

Where, $Cf(V)$ is the cofactor matrix of V ·The matrix with cofactor is the transposition of the matrix. With a matrix R invertible, can possess:

ISSN: 2233-7857 IJFGCN

$$Cf(V) = (dt(V))(V^{-1})^R \quad (8)$$

Where, $(V^{-1})^R$ is the inverse transposition of V . The factor in determining of a rotation matrix is 1 and its transpose is the contrary of a rotation matrix. We can infer the formula further as follows:

$$Cf(V) = (V^{-1})^R = V \quad (9)$$

Thus, it also follows the surface tensor representation for:

$$\tilde{D} - D \times_3 V^R \quad (10)$$

Where D is the tensor transformation of surface normal vectors. It concludes by directly correlating Equations (2), (5) and (10) proves that the same rotation matrix is normal in joint, edge or surface.

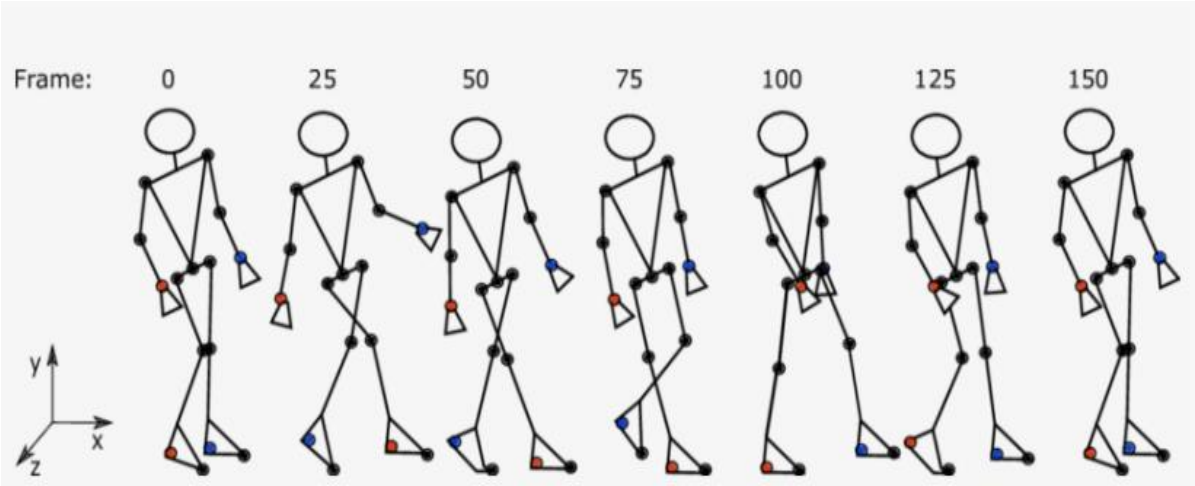


Figure 3: Human skeleton plotted in 3-D coordinate system

3.2 Feature Extraction

3.2.1 Histogram of Oriented Gradient (HOG)

The descriptors provided in the Histogram of Oriented Gradient (HOG) [27] provide outstanding efficiency with respect to other known functions such as wavelets. The fundamental hypothesis is that, even without profound comprehension of the appropriate gradient or edge position, the occurrence and form of the local objects can often be described as rather well by the utilization of the local intensity gradient / edge direction.In [27] each sensing window is segregated into 8 x 8-pixel cells and each group of 2 x 2 cells sliding in a row, therefore blocks overlap. The gradient size $a(m,n)(3)$ and inclination $\theta(m,n)(4)$ is determined for each pixel $P(m,n)$ of these cell forms. The gradient direction of the sample points inside a cell then forms a central, one-dimensional histogram of orientation of the gradient. The gradient angle range is divided into each histogram number of bins standardized (e.g. 9 bins). The size of the gradients vote on the histogram of orientation.A convolutional vector of all its cells is encompassed in each block. This means that each block has a 36-D characteristic vector that is standardized into the UL2 unit length (5). There is a total of 3780 characteristics per detection window, every 64 x 128 detection window is represented by 7 x 15 frames. This abstraction of features is evidently a complex representation, which maps local image areas to high-dimensional spaces. This is used for the formulation of a linear DCNN classification.

$$fm = P(m+1,n) - P(m-1,n) \quad (11)$$

$$fn = P(m,n+1) - P(m,n-1) \quad (12)$$

ISSN: 2233-7857 IJFGCN

$$a(m,n) = \sqrt{fm^2 + fn^2} \tag{13}$$

$$\theta(m,n) = \tan^{-1}\left(\frac{fm}{fn}\right) \tag{14}$$

$$r \leftarrow \frac{r}{\sqrt{\|r\|_2^2 + \varepsilon^2}} \tag{15}$$

The original histogram measurement method is not accurate. We adopt an Integral Image[28] to calculate histogram efficiently over cells in this paper. Nearly the entire picture at location (m,n) contains the sum of the above and left pixels of (m,n) including

$$in(m,n) = \sum_{m' \le m, n' \le n} o(m',n') \tag{16}$$

Where $in(m,n)$ denotes integral image and $o(m,n)$ denotes original image. Utilizing the same recurrence combination,

$$h(m,n) = h(m,n-1) + o(m,n) \tag{17}$$

$$in(m,n) = in(m-1,n) + h(m,n) \tag{18}$$

Where $h(m,n)$ denotes total sum row, $h(m,-1) = 0$ and $in(-1,n) = 0$, with one pass over the original image the integral value can be determined.

3.2.2 Haar-like feature

In the Viola and Jones algorithm the basic haar-like features utilized so-called the haar wavelet transforming coefficients are determined similarly. Instead of raw pixel values there are two reasons for the use of haar-like features. The first is that the haar-like functionality can encrypt adhoc domain information with a small amount of training data, which is impossible to articulate. The haar-like features will significantly decrease / boost the class / out-of-class volatility effectively in relation to raw pixels, thereby easing the classification.

The haar-like functionalities show the ratio of darkness to light within a kernel.The location of the eye on the human face is, for example, darker than the cheek region and one rugged skin can capture the other easily. Secondly, a haar-like feature-based method is designed to employ far more rapidly than a pixel system. In order to measure the gray-level disparity between white and black rectangles, the haar-like characteristics often give a fairly stable response against vibration and illuminative shifts. The alterations in noise and light significantly influence the pixels of the entire area of the feature and can ameliorate this significant impact.

Each feature is made of 2 or 3 "black" and "white" rectangles connected with each haar. The significance of haar is the discrepancy between the integrals of the black and white rectangular pixel metrics, i.e.

$$t(m) = \sum_{black}(pixelvalue) - \sum_{white}(pixelvalue) \tag{19}$$

Nearly the integral image at pixel location (m,n) contains the sum of the above and left pixel value including

$$A(m,n) = \sum_{m' \le m, n' \le n} a(m',n') \tag{20}$$

ISSN: 2233-7857 IJFGCN

The sum of the pixel values within the desired area can be determined by the description of the "integral image".

The image is scanned by a sub window with a special haar-like feature in order to identify a subject of interest. A corresponding weak $g_k(m)$ classification is described according to each haar-like feature t_k as,

$$g_k(m) = \begin{cases} 1, & if a_k t_k(m) < a_k \theta_k \\ 0, & otherwise \end{cases} \quad (21)$$

Where m denotes sub window and θ denotes threshold value. The inequality sign direction is indicated by a_k.

3.2.3 Distance transform gradient density function

The geometry of the transformation of space refers to a group of intersection cones and the roots in the centers of Voronoi. Except at the cone crossing and at the roots, the distance gradients transform, which occur worldwide, into unit vectors. Therefore, a single-dimensional gradient density coefficient is represented over the orientation space. The directions along each ray of each cone are constant and unique. Its distribution of probability function is indicated as

$$\Gamma(\theta \le \Theta \le \theta + \Delta) \equiv \frac{1}{R} \int \int_{\theta \le \arctan\left(\frac{H_n}{H_m}\right) \le \theta + \Delta} fmfn \quad (22)$$

Where R denotes area of bounded domain Ω. The orientation random variable is represented as $\Theta = \arctan\left(\frac{H_n}{H_m}\right)$. The probability distribution function also leads to a simple analytical interpretation, as described below, for its density function.

Assume that $\Omega \subset \Re^2$ signifies polynomial grid such that its boundary $\partial\Omega$ is procured of finite range of straight line segments. Let's assume that the set $B = \{B_i \in \Re^2, i \in \{1,...., I\}\}$ be the point set locations determined and then $B_i = (m_i, n_i)$. At certain point $A(m,n) \in \Omega$ the Euclidean distance transform is determined by

$$H(A) \equiv \min_i \|A - B_i\| = \min_i \left(\sqrt{(m - m_i)^2 + (n - n_i)^2} \right) \quad (23)$$

At B_i centered, let G_i signify the i^{th} Voronoi region relating to input B_i. G_i denoted by Cartesian product $[0,2\pi] \times [0, V_i(\theta)]$ where $V_i(\theta)$ denotes the ray length of i^{th} cone at an orientation θ. When a grid point is $A(m,n) \in (B_i + G_i)$, then $H(A) = \|A - B_i\|$. A convex polygon G_i whose boundary $\partial\Omega$ for each is procured of finite range of straight line segments.

Notification that the distance transformation is well defined even for points that lie on the Voronoi border with the radial length equal to $V_i(\theta)$. Area R is specified for the polygon grid as

$$R \equiv \sum_{i=1}^{I} \int_0^{2\pi} \int_0^{V_i(\theta)} dfdf\theta = \sum_{i=1}^{I} \int_0^{2\pi} \frac{V_i^2(\theta)}{2} f\theta \quad (24)$$

ISSN: 2233-7857 IJFGCN

Equation (22) can be simplified with the above set-up after recognition of the cone geometry at every Voronoi centre B_i as

$$\Gamma(\theta \le \Theta \le \theta + \Delta) \equiv \frac{1}{R}\sum_{i=1}^{I}\int_{\theta}^{\theta+\Delta}\int_{0}^{V_i(\theta)} dfdf\theta = \frac{1}{R}\sum_{i=1}^{I}\int_{\theta}^{\theta+\Delta}\frac{V_i^2(\theta)}{2}f\theta \tag{25}$$

After the dramatic simplification the shapes of the unit vector distance can be compiled as gradients for the density function as

$$A(\theta) \equiv \lim_{\Delta\to 0}\frac{\Gamma(\theta \le \Theta \le \theta + \Delta)}{\Delta} \equiv \frac{1}{R}\sum_{i=1}^{I}\frac{V_i^2(\theta)}{2}f\theta \tag{26}$$

It is convenient to have this depending on the expression R in Equation (24)

$$\int_0^{2\pi} A(\theta)f\theta = 1 \tag{27}$$

As the cells of Voronoi are convex polygons, each cell gives the density function by positioning precisely one conical ray.

4. Deep Convolutional neural network (DCNN)

The recognition rate is increased by a beneficial training process, because data is gathered when new functions are added to the Human body recognition system. The training process is based on the principle of physiology that relates the relationships between neurons or characteristics in the training space. It seems that every feature has an important role to play so that each feature pays attention to the scenario with its input features. Mostly during training phase, the network uses various layers such as input, hidden and output between the three layers; the input layer has several sub-layer levels, namely convolution, pooling, fully connected and normalized layers used to operate attributes viably. Initially, the specified features are used as an input to the next hidden layer. The information gathered are analyzed in a convolution network with respect to the propagation region. The metadata is collected by way of the in-depth learning process, which helps to recognise compatibility, compared to these source information.

The convolution layer process generates the responsive field, which results in the output of the cluster and demonstrates the same inputs as the same cluster. For each cluster, the maximum grouping function applies to the concentration layer. This ensures that the full value of features is chosen from each task cluster. The network has long scrutinized the over-adjustment of the data by boosting the file capacity. The extracted data can be integrated into the function space in the pooling layer and the highest value is selected for the fully connected layer, which computes the intrinsic worth of the output.In the totally connected layer the output importance is determined as authorization vector by matrix multiplication. The algorithm iterates until the selected functions are learned, which are kept for recognition in the database. Ultimately, different individual properties should be tested in a deep learning convolution neural network to model regular and irregular attributes. The extracted characteristics are synthesized to the input level, which is transferred from the weighted input to a hidden layer to estimate the output of the hidden layer.

Undertake an operation of analogy and perform equation (40) for the input vector access convolution.

$$M_a^x = \left(\sum_{t=1}^{z}\sum_{r=-b}^{b}\sum_{f=-b}^{b} N_t^{x-1}(i-r, j-f) * W_{t,a}^x(r, f) + L_a^x\right) \tag{40}$$

Where, z signifies last layers number of maps feature, t denotes feature map indices of current layer and d signifies feature map indices of previous layer, l denotes the layer, * denotes convolution operation, L and b represents the bias and size of filter respectively. Initially N_t^0 represents the input

ISSN: 2233-7857 IJFGCN

image on which first convolution is to be performed and N_t^1 represents the input on which second convolution is to be performed, which can be obtained after applying pooling on N_t^0.

$$K_a^x(i,j) = \left(\frac{1}{4} \sum_{r=0}^{y} \sum_{f=0}^{y} M_a^x(2i-r, 2j-f) \right) \quad (41)$$

Where y denotes window width of the enclosure. Pursue a large number of convergence and pooling iterations when considered necessary. Relocate the extracted outcome from the last layer of pooling into a fully connected classification layer and calculate the efficiency of the equation 42.

$$Actual\ Output = \sigma(wht \times ott + L) \quad (42)$$

Where, ott is the final output vector obtained after last pooling operation, wht is the weight vector of fully connected layer.

For sigmoid activation function,

$$\Delta \hat{W}(i) = (\hat{W}(i) - W(i)) \bullet \hat{W}(i)(1 - \hat{W}(i)) \quad (43)$$

Computation of Δwht,

$$\Delta wht = \Delta \hat{W}(i) \times ott(j) \quad (44)$$

Computation of ΔL,

$$\Delta L = \frac{\partial Q}{\partial L(i)} \quad (45)$$

The result is determined on the basis of Equation (42) to accurately predict the scenario of new outputs of skeleton points within an expert model fully integrated with database-based training data. Throughout the classification process, weights and bias values are continuously monitored such that the detection error rate is mitigated. The element is used to modify new test characteristics to train following weight and procreative value testing.

5. Squirrel Search Algorithm (SSA)

The use of the Deep Convolutional Neural Network (DCNN) algorithm to maximize the usefulness of an objective and efficient human actions recognition systems aim to correctly classify input data into its underlying activity category by Squirrel Search Algorithm (SSA) algorithm.The algorithm tries to emulate flying squirrel movements and gliding characters. The mathematical model primarily consists of where the food source is located and the visual appeal of predators. The entire process of optimization covers the summer and winter duration. However, the random summer search procedure shrinks convergence speed and also reduces convergence accuracy. This paper proposes an enhanced Squirrel Search Algorithm (SSA) in order to improve convergence accuracy and convergence speed. Depending on the current season, the conventional SSA updates the individual location, the type of people and if either predators occur.

5.1 Population Initialization

The upper and lower limits of the search space will be SP_U and SP_L provided the population number is N. N people are produced randomly by equation (28):

$$SP_k = SP_L + rand(1, M) \times (SP_U - SP_L) \quad (28)$$

ISSN: 2233-7857 IJFGCN

SP_i signifies the $k-th$ individual, $k=1...N$, $rand$ denotes the random number range between 0 and 1 and M is the problem dimension.

5.2 Population classification

SSA needs only one squirrel at each tree, supposing that the total squirrel number is N, so in the forest there are total N trees. All N trees are hickory trees and $N_{sp}(1<N_{sp}<N)$ acorn trees. The rest are normal, food-free trees. The hickory tree for squirrels is the best food resource and the acorn tree is second. Based on the multiple issues, N_{ss} may vary. The squirrels are categorized into three types by classifying the fitness values of the population in an increasing order: individuals in the hickory (S_y) trees position, those in the acorn trees (S_c) position and those in normal trees (S_m) position. S_y is an entity with the lowest fitness rating, S_c is an entity whose fitness varies from 2 to $N_{sp}+1$ and the remainder are classified as Fn. The destination of S_c is S_y in order to locate the best source of food; the destinations of S_m are arbitrarily identified as S_c or S_y.

5.3 Position Updation

The entities glide into their hickory trees or acorn trees to change their roles. Equations (29) and (30), respectively, indicate the following upgrade formulations as:

$$\begin{cases} SP_k^{q+1}=SP_k^q+w_a\times Y_d\times(S_y^q-SP_k^q) & ift>V_{we} \\ randomlocation & otherwise \end{cases} \tag{29}$$

$$\begin{cases} SP_k^{q+1}=SP_k^q+w_a\times Y_d\times(S_{ck}^q-SP_k^q) & ift>V_{we} \\ randomlocation & otherwise \end{cases} \tag{30}$$

t denotes random number ranging from 0 to 1; V_{we} determined to be 0.1 indicates a risk of dangerous appearance; when $t>V_{we}$, then no predator is emerging, forestry squirrels glide, humans are safe; if $t>V_{we}$, the predators appear; the squirrels are compelled to restrict the reach of the behaviors, the individuals are endangered, and their locations are arbitrarily moved. q notifies current iteration; Y_d denotes constant value 1.9; $S_{ck}(k=1,2,....N_{sp})$ denotes randomly selected individual from S_c; w_a denotes gliding distance that can be formulated by equation (31):

$$w_a=\frac{d_a}{\tan(\varphi)\times ps} \tag{31}$$

d_a denotes constant value 8; ps denotes constant value 18; $\tan(\varphi)$ denotes gliding distance that can be formulated by equation (32):

$$\tan(\varphi)=\frac{DF}{LF} \tag{32}$$

DF denotes drag force and LF denotes lift force that can be formulated by calculating equations (33) and (34) respectively;

ISSN: 2233-7857 IJFGCN

$$DF = \frac{1}{2\rho W^2 HE_{DF}} \tag{33}$$

$$LF = \frac{1}{2\rho W^2 HE_{LF}} \tag{34}$$

ρ, W, H and E_{DF} signifies constant; E_{LF} signifies random number.

5.4 Monitoring seasonal condition and random Updation

The SSA regulations ensures that the entire population at the start of each iteration to be updated during the winter. In other words, every individual seems to be updated. If the changes in season are determined by equations (35) and (36) then all individuals have been updated:

$$P_d^q = \sqrt{\sum_i^{DL} S_{ck,i}^q - S_{y,i}^{q\ 2}}\, i = 1,2,.....N_{sp} \tag{35}$$

$$P_{\min} = \frac{10e^{-6}}{365^{q/(Q/2.5)}} \tag{36}$$

E_{LF} denotes infinite number of cycles, when $P_d^q = P_{\min}$ winter is over and summer is the season, otherwise the season remains without any change. When the summer season arrives, every person who glides to S_y lives in the modified position and every person gliding to S_c and never encountering predators shifts based on equation (37):

$$SP_{wnew}^{q+1} = SP_L + Le'vy(x) \times (SP_U - SP_L) \tag{37}$$

From equation (37) formulate $Le'vy(x)$ which make people search for more opportunities within a brief range and search regularly in the lengthy range.

$$Le'vy(x) = 0.01 \times \frac{d_x \times \sigma}{\left|d_y\right|^{\frac{1}{\beta}}} \tag{38}$$

Where β denotes constant and σ is formulated as:

$$\sigma = \left(\frac{\Gamma\left(1+\beta \times \sin\left(\frac{\pi\beta}{2}\right)\right)}{\Gamma\left(\frac{1+\beta}{2} \times \beta \times 2^{\left(\frac{\beta-1}{2}\right)}\right)} \right) a^{\frac{1}{\beta}} \tag{39}$$

$where \Gamma(x) = (x=1)!$

The procedure of SSA is shown in figure 4.

ISSN: 2233-7857 IJFGCN

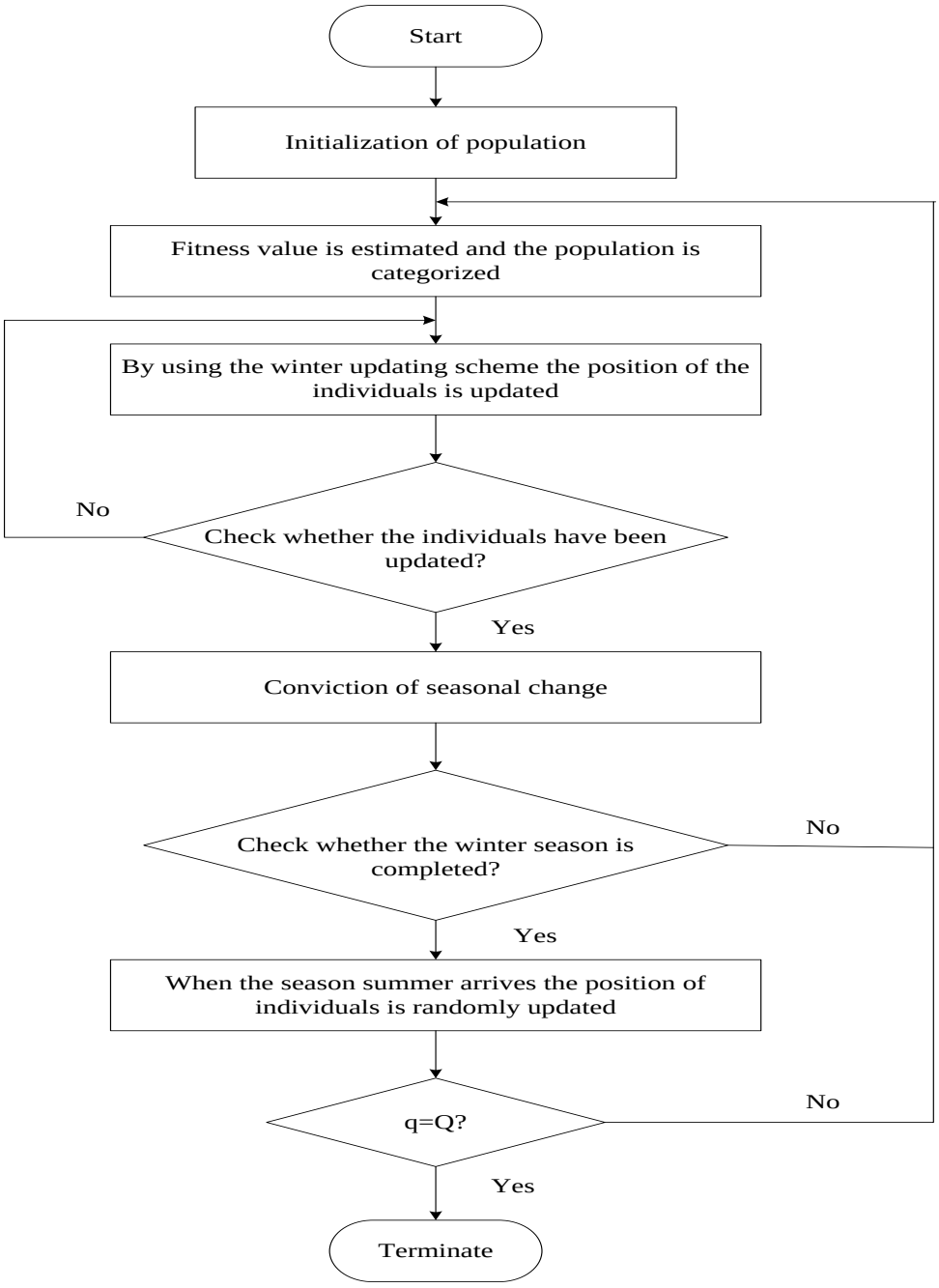


Figure 4: Flowchart of SSA

6. EXPERIMENTAL RESULTS

6.1 Parse 27-k Dataset

PARSE-27k is based on eight separate duration video sequences from a moving camera in an urban environment. The DPM pedestrian detector was used for each 15th frame of the series. The bounding boxes collected with 10 character labels were annotated manually. A robotics / automotive technology scenario motivates the choice of attributes and comprises of 2 orientation labels of 4 and 8 discretes and multiple binary attributes such as standing, walking, the left hand movement, right leg movement etc. PARSE-27k has a division of cautious train (5%), val (25%) and test (25%). This implies it is only segregated into series limits.In comparison, train-val or test sequences taken on the same day is utilized. It avoids very identical instances of differences. Therefore, the variance in the pose and crop of PARSE-27k is less because it includes only crops of pedestrian boundaries created by a pedestrian detector.

ISSN: 2233-7857 IJFGCN

6.2 Performance metrics

In comparison with the SVM[29], and NB[30], the achievement of the suggested technique of clinical cancer analysis is carried out using methods of awareness, accuracy, specificity, precision, reclamation, F-measurements, favourable predictive importance, PPV and MCC. Certain metrics are computed with True Negative (T_N), True positive (T_P), False Negative (F_N) and False positive (F_P).

6.2.1 Accuracy

The measure of overall usefulness/ effectiveness of the classification technique are called accuracy. The equation for accuracy is

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{46}$$

6.2.2 Specificity

To recognize patterns of a negative class, it is used to measure the classifier ability. It is computed as follows.

$$specificity = \frac{T_N}{T_N + F_P} \tag{47}$$

6.2.3 Sensitivity

To recognize patterns of a positive class, it is calculated to measure the classifier ability. It is computed as follows.

$$sensitivity = \frac{T_P}{T_P + F_N} \tag{48}$$

6.2.4 F-measure

To calculate the classification accuracy F-measure make use of recall and precision. It is defined as,

$$F_{measure} = 2 \times \frac{recall \times precision}{recall + precision} \tag{49}$$

$$recall = \frac{T_P}{T_P + F_N} \tag{50}$$

$$precision = \frac{T_P}{T_P + F_P} \tag{51}$$

6.2.5 Error rate

The number of all incorrect predictions divided by total number of the dataset is calculated as error rate.

$$Error_{rate} = \frac{F_P + F_N}{T_P + T_N + F_P + F_N} \tag{52}$$

The positive predictive value and negative predictive value is expressed as

$$PPV = \frac{T_P}{T_P + F_P} \tag{53}$$

ISSN: 2233-7857 IJFGCN

$$NPV = \frac{T_N}{T_N + F_N} \tag{54}$$

True positive rate $$TPR = \frac{T_P}{P} = \frac{T_P}{T_P + F_N} \tag{55}$$

True negative rate $$TNR = \frac{T_N}{T_N + F_P} \tag{56}$$

False positive rate $$FPR = \frac{F_P}{F_P + T_N} \tag{57}$$

False negative rate $$FNR = \frac{F_N}{F_N + T_P} \tag{58}$$

6.3 Performance analysis

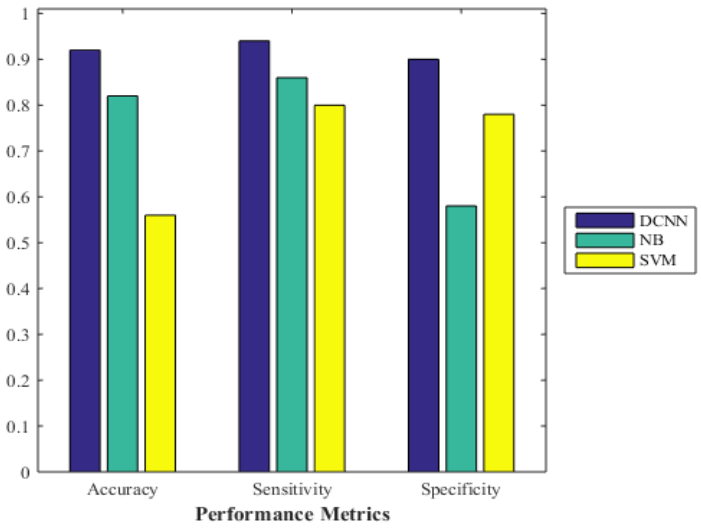


Figure 5: Comparison of performance metrics such as sensitivity, specificity and accuracy with various classifiers

In fig. 5, the performance metrics such as sensitivity, specificity and accuracy is compared with proposed DCNN, NB and SVM. In accuracy the proposed DCNN shows 11.56% and 44.34% higher compared to NB and SVM respectively. In sensitivity the proposed DCNN shows 9.86%and 13.43% higher compared to NB and SVM respectively. In specificity the proposed DCNN shows 42.56% and 14.98% higher compared to NB and SVM respectively.

ISSN: 2233-7857 IJFGCN

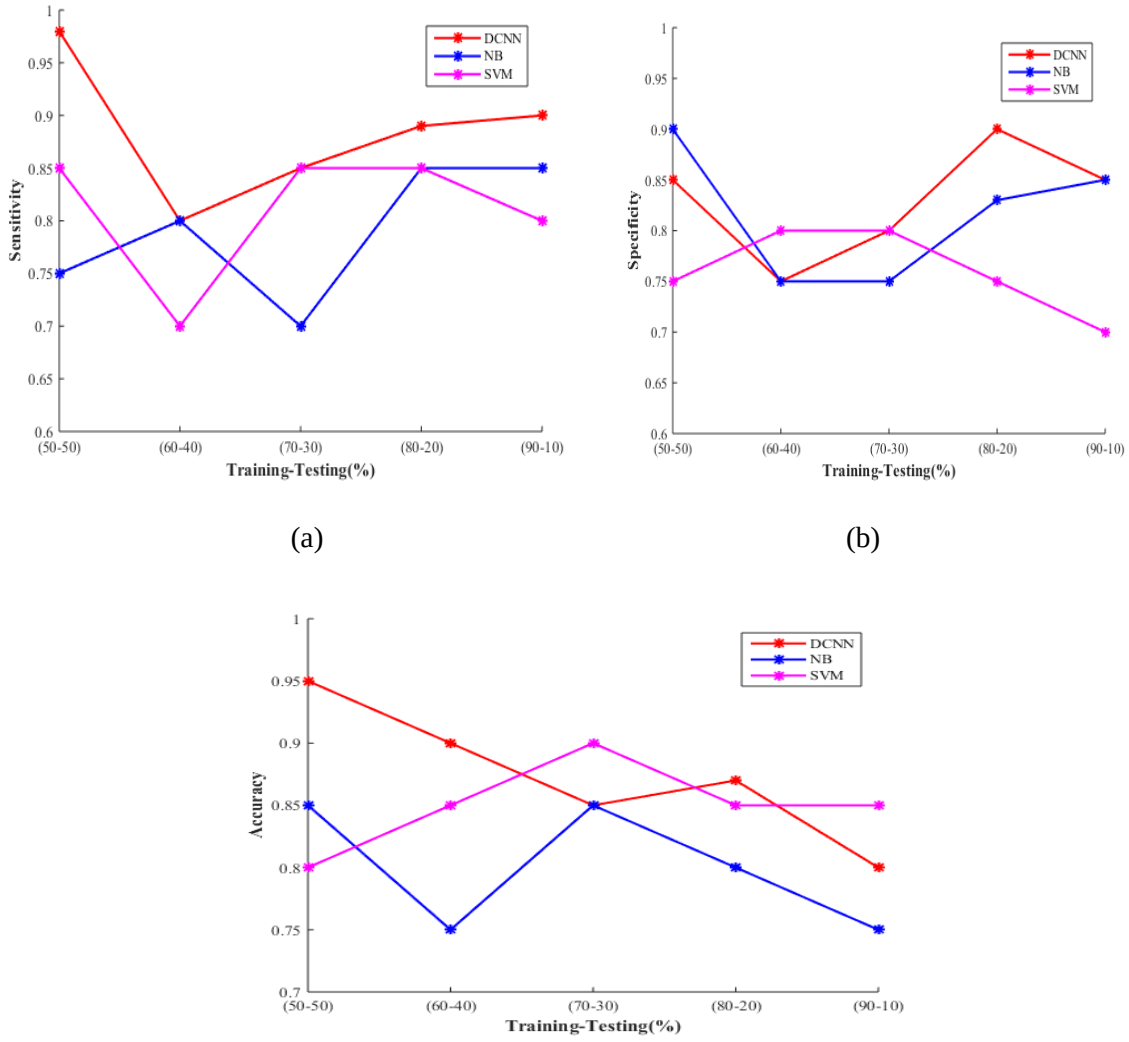


Figure 6: Comparison of performance metrics such as (a) sensitivity, (b) specificity and (c) accuracy with various classifiers using training and testing phase

In fig. 6, in sensitivity and accuracy, the DCNN performance is higher compared to NB and SVM with the values near to 1 at 50-50 and there is far more accuracy comparatively. But in specificity, the DCNN performance is higher compared to NB and SVM with the values near to 1 at 80-20.DCNN perpetrated efficiently.

ISSN: 2233-7857 IJFGCN

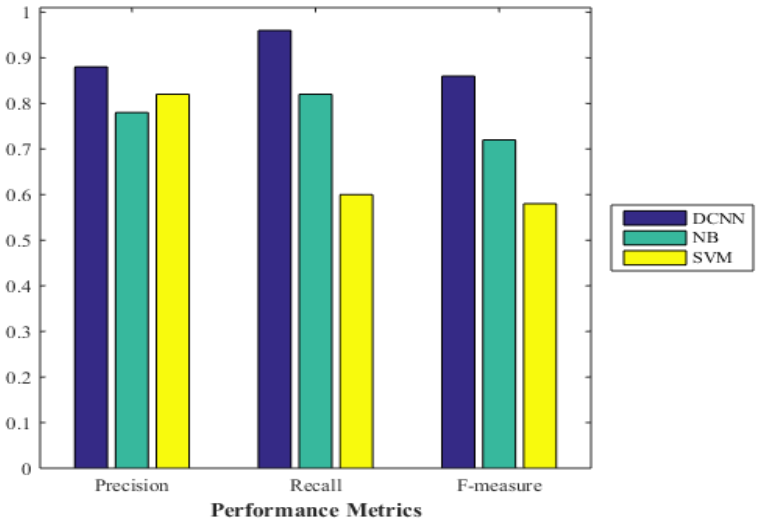
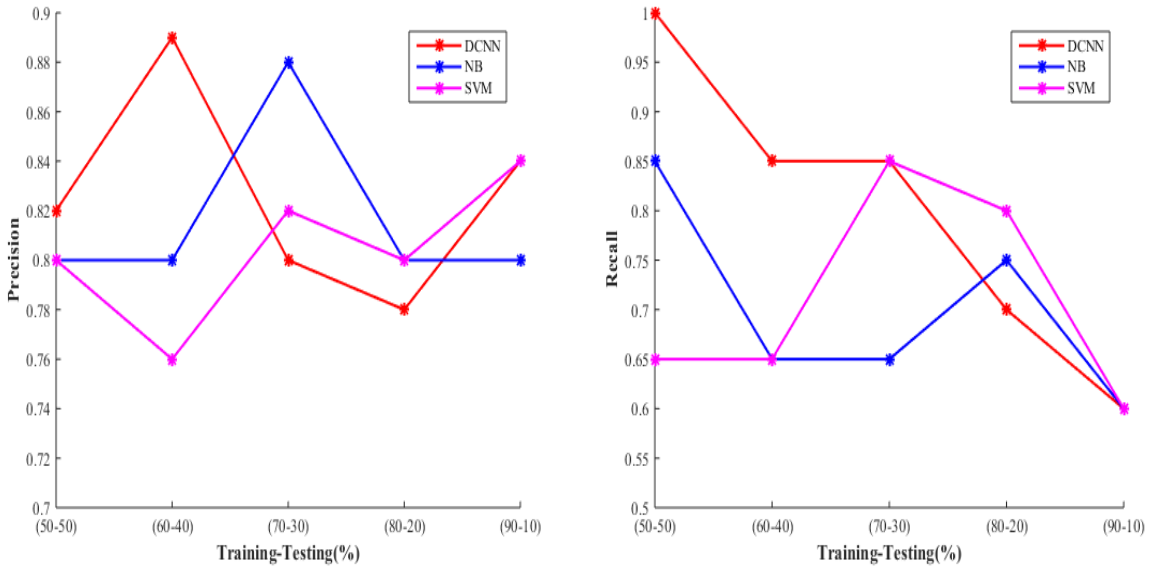


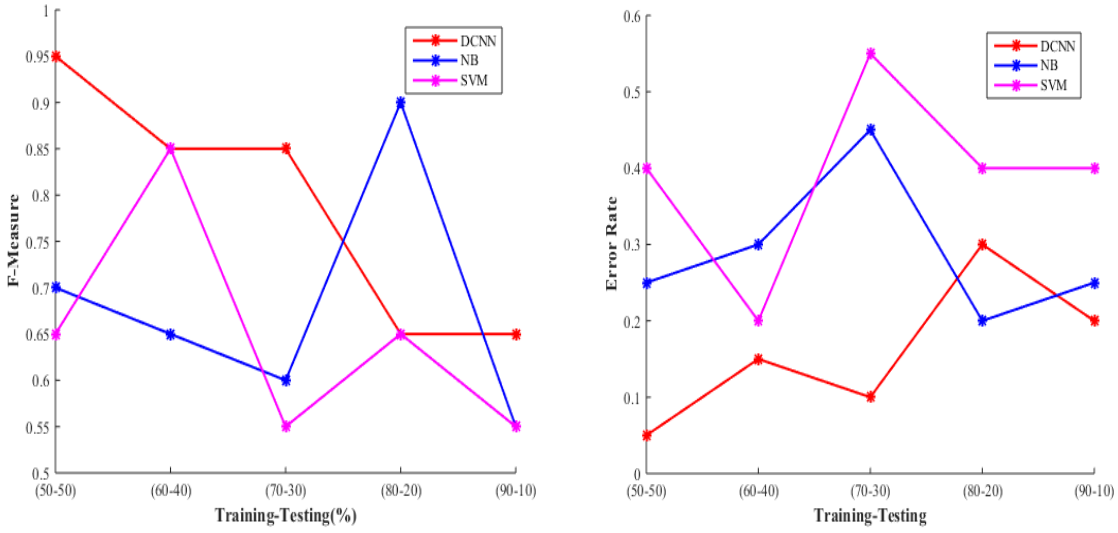
Figure 7: Comparison of performance metrics such as precision, recall and F-Measure with various classifiers

In fig. 7, the performance metrics such as precision, recall and F-Measure is compared with proposed DCNN, NB and SVM. In precision the proposed DCNN shows 11.24% and 6.13% higher compared to NB and SVM respectively. In recall the proposed DCNN shows 16.12% and 41.38% higher compared to NB and SVM respectively. In F-Measure the proposed DCNN shows 14.72% and 33.47% higher compared to NB and SVM respectively.



(a) (b)

ISSN: 2233-7857 IJFGCN



(c) (d)

Figure 8: Comparison of performance metrics such as (a) precision, (b) recall, (c) F-Measure and (d) error rate with various classifiers using training and testing phase

In fig. 8, in recall DCNN performance is higher compared to NB and SVM at 60-40 criteria. In precision DCNN performance is higher compared to NB and SVM at 50-50 criteria. In F-Measure DCNN performance is higher compared to NB and SVM at 50-50 criteria. In error rate DCNN performance is higher compared to NB and SVM except 80-20 criteria.

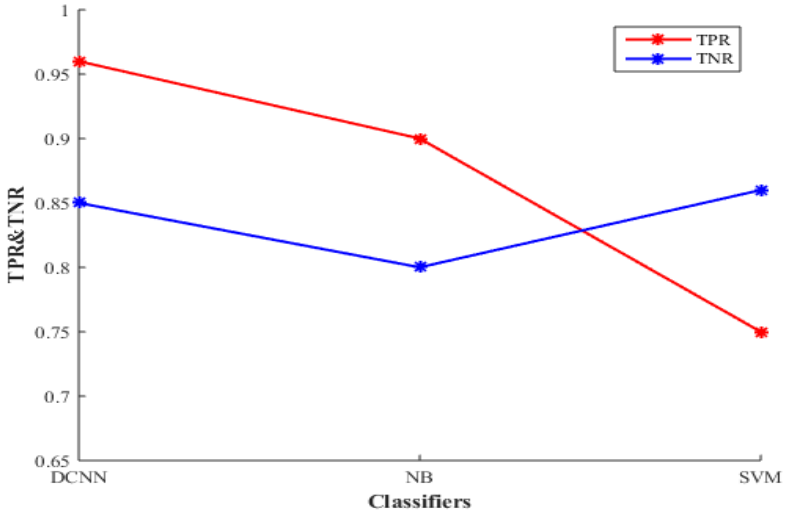
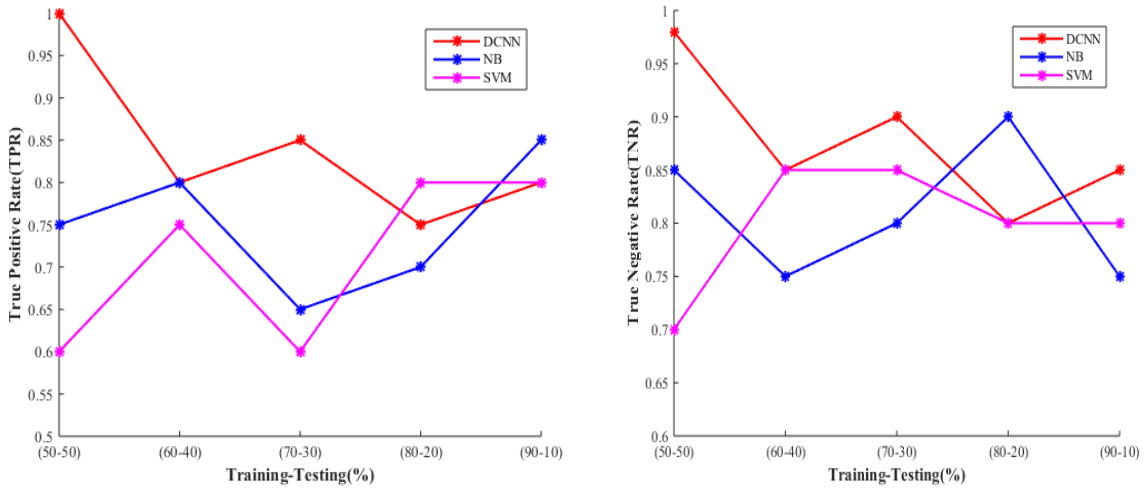


Figure 9: Comparison of performance metrics such as True positive rate (TPR) and True negative rate (TNR)with various classifiers

In fig. 9, the performance metrics such as True positive rate (TPR) and True negative rate (TNR) is compared with proposed DCNN, NB and SVM. In TPR, DCNN performance rate is higher compared to NB and SVM. In TNR, DCNN performance is least compared to NB and SVM.

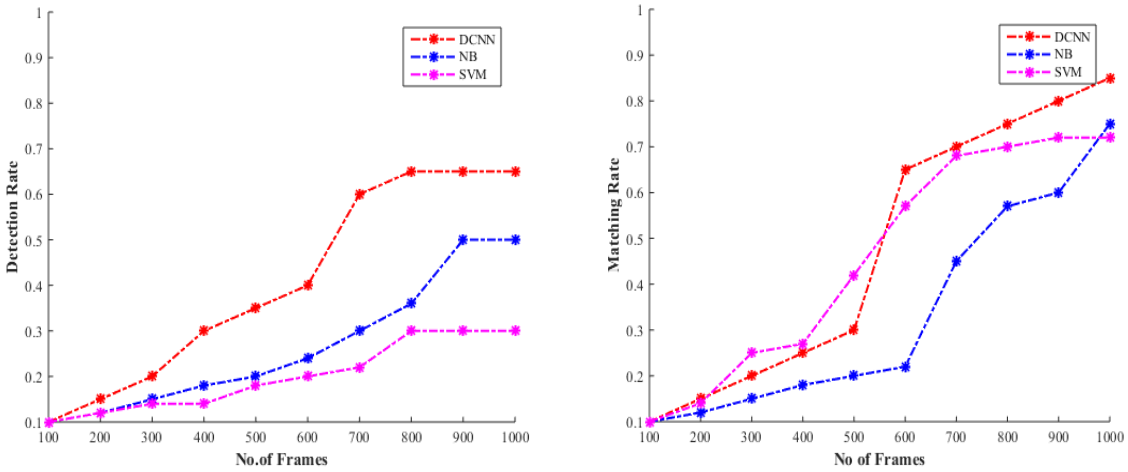
ISSN: 2233-7857 IJFGCN



(a) (b)

Figure 10: Comparison of performance metrics such as (a) True positive rate (TPR) and (b) True negative rate (TNR) with various classifiers using training and testing phase

In fig. 10, in True positive rate (TPR) and True negative rate (TNR), DCNN performance is higher compared to NB and SVM.



(a) (b)

Figure 11: Performance of detection rate and matching rate based on number of frames

In fig. 11, in detection rate there is a gradual increase rate as the number of frames increases. In matching rate there is a gradual increase rate from frames 100-500 then there is a steep rise of 0.7 rates from 0.3 at frame 600 and then there is again gradual increase rate from 600-1000 frames.

ISSN: 2233-7857 IJFGCN

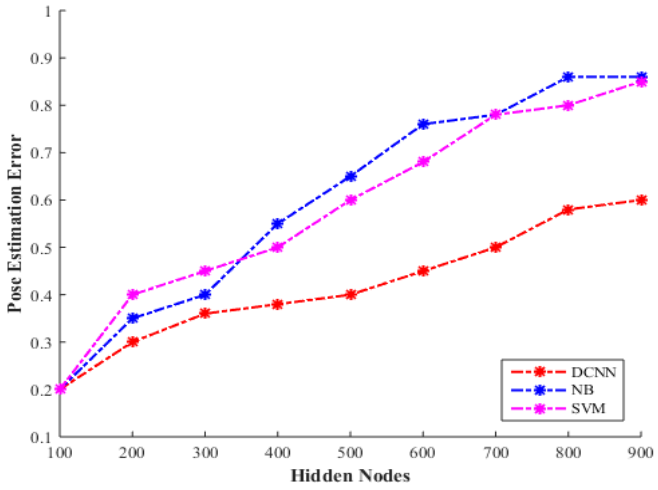


Figure 12: Pose Estimation Error Vs Hidden node with various classifiers

In fig. 12, the proposed DCNN performs minimum error rate compared to NB and SVM based on the hidden nodes.

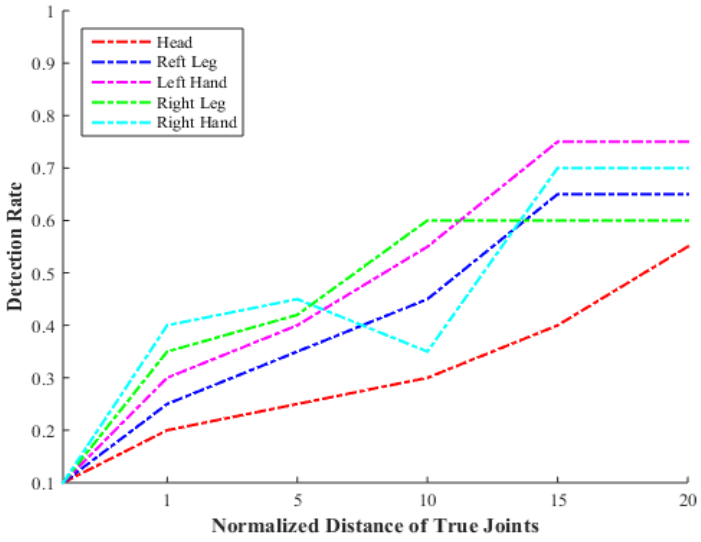
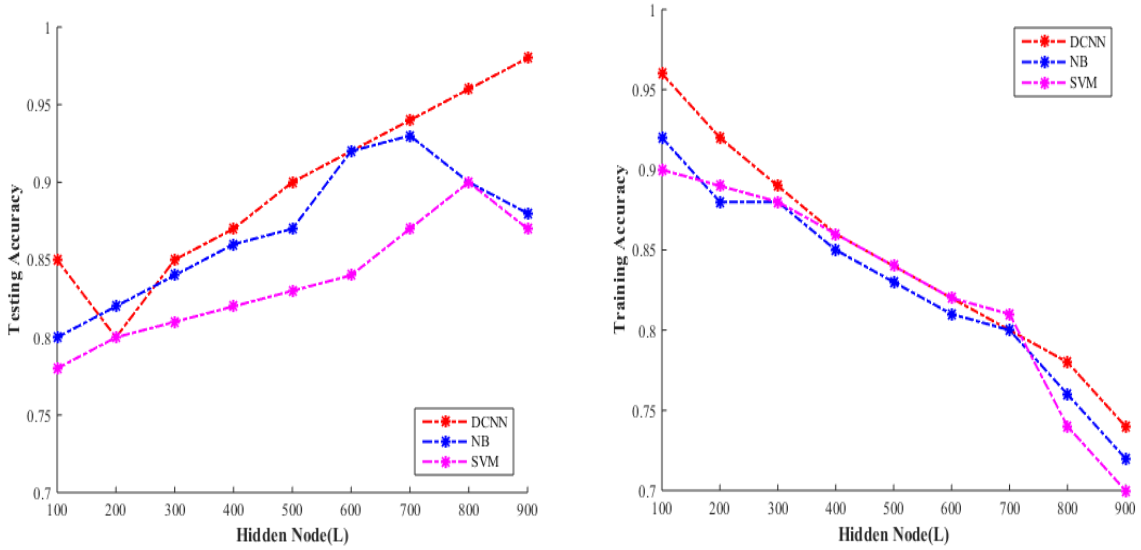


Figure 13: Detection rate Vs Normalized distance of true joints

In fig. 13, at normalized distance of true joints rate 1, the head detection rate is 0.2, the left leg detection rate is 0.25, the left hand detection rate is 0.3, the right leg detection rate is 0.35 and the right hand detection rate is 0.25. At normalized distance of true joints rate 5, the head detection rate is 0.26, the left leg detection rate is 0.32, the left hand detection rate is 0.37, the right leg detection rate is 0.39 and the right hand detection rate is 0.41. At normalized distance of true joints rate 10, the head detection rate is 0.26, the left leg detection rate is 0.4, the left hand detection rate is 0.58, the right leg detection rate is 0.61 and the right hand detection rate is 0.32. At normalized distance of true joints rate 15, the head detection rate is 0.4, the left leg detection rate is 0.65, the left hand detection rate is 0.77, the right leg detection rate is 0.6 and the right hand detection rate is 0.71.At normalized distance of true joints rate 20, the head detection rate is 0.58, the left leg detection rate is 0.65, the left hand detection rate is 0.77, the right leg detection rate is 0.6 and the right hand detection rate is 0.71. From 15 to 20 the distance is normalized for true joints.

ISSN: 2233-7857 IJFGCN



(a) (b)

Figure 14: Comparison of (a) testing accuracy and (b) training accuracy based on hidden nodes with various classifiers

In fig. 14, in testing accuracy as the node increases the accuracy also increases but in training accuracy as the node decreases the accuracy increases.

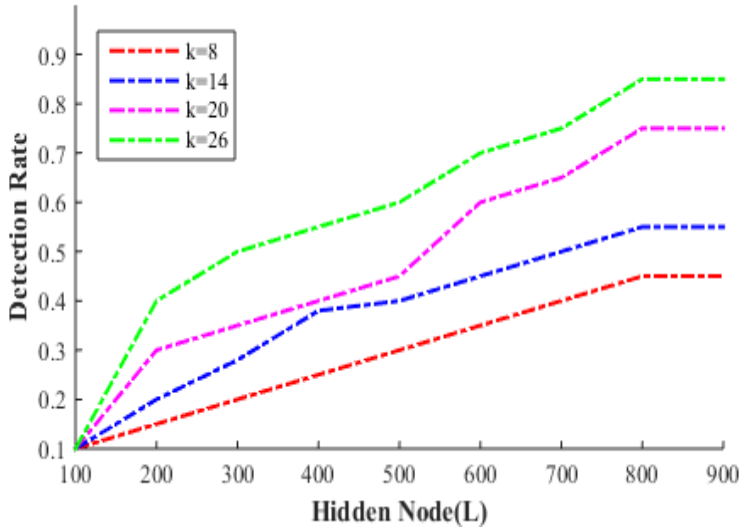


Figure 15: Detection rate Vs Hidden node based on number of frames

In fig. 15, the number of frames selected is 8, 14, 20 and 26. When the number of frames increases the detection rate increases.

ISSN: 2233-7857 IJFGCN

7. CONCLUSION

We intend to research the representations of human skeleton basic geometries, i.e. joints, edges and surfaces in this paper. We propose a new architecture based on DCNN to accommodate the three inputs to the action recognition. For the detection of actions, we first define objects, and then develop a multiple sliding search algorithm for sliding windows to produce detection results. We aim to enhance the efficiency of recognizing human actions by improving the classification method from the raw input images and Human Body Skeleton feature extracted Based on Silhouette (Pose) images such as walking, running, dancing etc. using Squirrel Search Algorithm. We present a new dataset PARSE 27-K focused on practical outdoor video clips, which includes over 27,000 people, with 10 characteristics, to help further assessments and use it to perform extensive evaluation to examine the performance relevant factors of our model.

8. REFERENCES

1. Bosch, A. Zisserman, and X. Muñoz, "Scene classification via pLSA," in Computer Vision—ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7–13, 2006, Proceedings, Part IV, vol. 3954 of Lecture Notes in Computer Science, pp. 517–530, Springer, Berlin, Germany, 2006. View at Publisher · View at Google Scholar
2. J. C. Niebles, H. Wang, and L. Fei-Fei, "Unsupervised learning of human action categories using spatial-temporal words," International Journal of Computer Vision, vol. 79, no. 3, pp. 299–318, 2008. View at Publisher · View at Google Scholar · View at Scopus
3. A. Bissacco, M. H. Yang, and S. Soatto, "Detecting humans via their pose," in Advances in Neural Information Processing Systems, vol. 19, pp. 169–176, MIT Press, Boston, Mass, USA, 2007. View at Google Scholar
4. J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore,A. Kipman, and A. Blake, "Real-time human pose recognition in partsfrom single depth images," in IEEE Conference on Computer Visionand Pattern Recognition, 2011.
5. I. Laptev, "On space-time interest points," IJCV, vol. 64, no. 2-3, pp.107–123, 2005.
6. G. Willems, T. Tuytelaars, and L. Van Gool, "An efficient denseand scale-invariant spatio-temporal interest point detector," in ECCV.Springer, 2008, pp. 650–663.
7. I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld, "Learningrealistic human actions from movies," in IEEE Conference on ComputerVision and Pattern Recognition. IEEE, 2008, pp. 1–8.
8. Q Zhu et al., in Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. Fast human detection using a cascade of histograms of oriented gradients (IEEE, 2006)
9. N Dalal, B Triggs, in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Histograms of oriented gradients for human detection (IEEE, 2005)
10. A Satpathy, X Jiang, H-L Eng, Human detection by quadratic classification on subspace of extended histogram of gradients. IEEE Trans. Image Process. 23(1), 287–297 (2014)
11. S Zhang, C Bauckhage, AB Cremers, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Informed haar-like features improve pedestrian detection (2014)
12. P Viola, MJ Jones, D Snow, in null. Detecting pedestrians using patterns of motion and appearance (IEEE, 2003
13. WR Schwartz et al., in Computer Vision, 2009 IEEE 12th International Conference on. Human detection using partial least squares analysis (IEEE, 2009)
14. W Gao, H Ai, S Lao, in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. Adaptive contour features in oriented granular space for human detection and segmentation (IEEE, 2009)
15. B Leibe, A Leonardis, B Schiele, Robust object detection with interleaved categorization and segmentation. Int. J. Comput. Vis. 77(1–3), 259–289 (2008)

ISSN: 2233-7857 IJFGCN

16. Zhang, Lun, Rufeng Chu, Shiming Xiang, Shengcai Liao, Stan Z. Li. “Face detection based on multi-block lbp representation.” In International Conference on Biometrics, pp. 11–18. Springer, Berlin, Heidelberg, 2007.
17. Chen, Chen et al. "A Real-Time Human Action Recognition System Using Depth and Inertial Sensor Fusion". IEEE Sensors Journal, vol 16, no. 3, 2016, pp. 773-781. Institute of Electrical and Electronics Engineers (IEEE), 2019.
18. Tu, Zhigang et al. "Action-Stage Emphasized Spatiotemporal VLAD for Video Action Recognition". IEEE Transactions on Image Processing, vol 28, no. 6, 2019, pp. 2799-2812. Institute of Electrical and Electronics Engineers (IEEE), 2019.
19. Wang, Hongsong, and Liang Wang. "Beyond Joints: Learning Representations from Primitive Geometries for Skeleton-Based Action Recognition and Detection". IEEE Transactions on Image Processing, vol 27, no. 9, 2018, pp. 4382-4394. Institute of Electrical and Electronics Engineers (IEEE), 2019.
20. Wei, Ping et al. "Learning Composite Latent Structures For 3D Human Action Representation and Recognition". IEEE Transactions on Multimedia, vol 21, no. 9, 2019, pp. 2195-2208. Institute of Electrical and Electronics Engineers (IEEE), 2019.
21. Kamel, Aouaidjia et al. "Deep Convolutional Neural Networks for Human Action Recognition Using Depth Maps and Postures". IEEE Transactions on Systems, Man, And Cybernetics: Systems, vol 49, no. 9, 2019, pp. 1806-1819. Institute of Electrical and Electronics Engineers (IEEE), 2019.
22. Talha, Sid Ahmed Walid et al. "Features and Classification Schemes for View-Invariant and Real-Time Human Action Recognition". IEEE Transactions on Cognitive and Developmental Systems, vol 10, no. 4, 2018, pp. 894-902. Institute of Electrical and Electronics Engineers (IEEE), 2019.
23. Ehatisham-Ul-Haq, Muhammad et al. "Robust Human Activity Recognition Using Multimodal Feature-Level Fusion". IEEE Access, vol 7, 2019, pp. 60736-60751. Institute of Electrical and Electronics Engineers (IEEE), 2019.
24. Hinton, Geoffrey E. "Deep belief networks." Scholarpedia 4, no. 5 (2009): 5947.
25. H. Shin et al., "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning", IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1285-1298, 2016.
26. M. Jain, V. Singh and A. Rani, "A novel nature-inspired algorithm for optimization: Squirrel search algorithm", Swarm and Evolutionary Computation, vol. 44, pp. 148-175, 2019.
27. D. Navneet, and T. Bill, "Histograms of Oriented Gradients for Human Detection, "Computer Vision and Pattern Recognition, vol.l, pp. 886-893, June 2005.
28. P Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features, "Computer Vision and Pattern Recognition”, 2001.
29. Shankar, S. K. Lakshmanaprabu, DeepakGupta, AndinoMaseleno and Victor Hugo C. de Albuquerque 2018, Optimal feature-based multi-kernel SVM approach for thyroid disease classification, Journal of Supercomputing,pp.1-16.
30. Masud Karim and Rashedur M. Rahman, "Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing", Journal of Software Engineering and Applications, Vol.6 No.4, pp. 1-11, 2013.
31. G. Johansson, “Visual perception of biological motion and a model for its analysis,” Perception & Psychophysics, vol. 14, no. 2, pp. 201–211, 1973.

ISSN: 2233-7857 IJFGCN