

Human Upper Body Post Estimation Based On Modified Multiple Importance Sampling Algorithm

^{*1}M.Mahalakshmi, ²D.Vijayakumar, ³S.Krishnaveni, ⁴Akshita Babburi

^{*1, 2, 3}Associate Professors, Department of ECE, CMR College of Engineering & Technology, Hyderabad, India

^{*1}strimaha@gmail.com

²vkkumarin@gmail.com

³krishnaveni@cmrcet.ac.in

⁴Illinois Institute of Technology, 10 W 35th St, Chicago, IL 60616, United States

⁴ababburi@hawk.iit.edu

Received: 25th October 2021, Accepted: 19th November 2021, Published: 31st December 2021

Abstract

Tracking of human motion and their body parts under dynamic environment has been the difficult task during video detection situations. The work is to propose the human object video tracking methods to estimate the movement of human upper body under dynamic environment. Considering the mixture of frequent moving object parts with occlusion problem, deciding of target object identification and upper body pose variations will end in improving the accuracy. It also helps with reduction in occlusion of the moving body parts detection. The Modified Multiple Importance Sampling filter (MMIS) has been proposed to trace the human poses with fast paced actions. Dynamic sampling filter tracks the upper a part of physical body with 2D image and 3D postures. The high accuracy of the system has been obtained for cluttered environment with occlusion problem by properly obtaining the sampling states of the filter as shown within the experimental result analysis part.

Keywords: Upper body Motion, Video tracking, Occlusion, MMIS, Dynamic environment

Introduction

The part estimation is an attracting research topic in various fields like, visual-interfering games, content oriented image analysis and retrieval, Generally, to trace or sense the movements of upper body parts of a person's , special sensors, clothes or specific environment is employed as shown in Figure 1. These may end in a greatest discomfort in many real-time applications. To avoid these disadvantages, the tracking is often done by processing the pictures obtained from a camera.

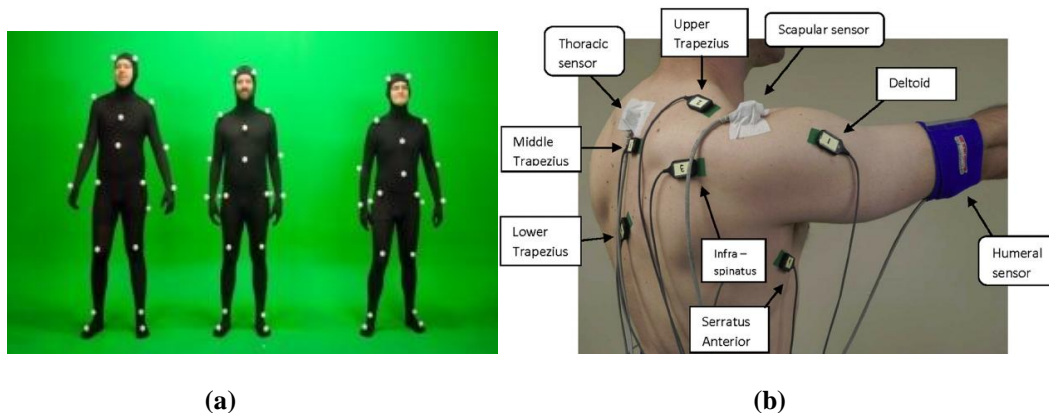


Fig 1: Tracking of human body parts based on (a) special environments and clothing (b) sensors attached to the body

Review of Literature

In current years, act recognition [1] has strained plenty consideration within the field of video investigation technology thanks to the developing demands from many applications, such as, entertainment environments, supervision environments and healthcare systems. In recent years, many researchers are wiped out the sector of detection and recognition of objects from the pictures. Suman Sedai et al. (2013) [2] proposed an Annealed Gaussian Process Guided Particle Filter (AGP-PF) for 3D physical body parts tracking in video sequences. Prediction uncertainty obtained from the Gaussian process has been wont to establish the discriminative method of tracking to realize more accuracy and stable tracking in every video frame for extended period of time. It lacks by means of a requirement of a person's body model, which makes tracking possible only with stationary human images.

Ankur Agarwal & Bill Triggs (2006) [3] used a learning-based method for 3D body pose tracking from single and monocular images by directly using shape descriptors extracted from silhouettes called as regression method, which makes use of an outsized training data set. Zhenget et al (2015) [4] proposed physical body part tracking under infrared based depth camera to work even in dark condition. Depth camera finds the connection between physical body motion and therefore the environment. Rincon et al (2011) [5] illustrated tracking of physical body parts supported Kalman and Particle filtering employing a single camera. Since initialization and re-initialization at every frame work isn't required, the main advantage of this method isn't constraint to specific activity and applicable to all or any sorts of motions.

Demerit shows that accuracy isn't maintained when the thing is occluded for longer duration. Navnnet Dalal & Bill Triggs (2005) [6] have proposed a system for robust visual visual perception. The disadvantage of this technique is that it still needed to be optimized and sped up for better detecting performance. Buehler et al (2008) [7] have worked on detecting and tracking the arms and hands in signing broadcasts supported generative model.

The tracking of physical body parts when the topic is distracted by various other objects or the other citizenry may be a major disadvantage of varied other techniques available for tracking. The clumsy use of sensors or special backgrounds and specialized clothes were used previously for tracking the physical body parts. The proposed technique avoids all such complex methods [8,9], [10] and makes use of single monocular camera leads us to use this system in real time visual surveillance applications. The effectiveness of tracking of the body parts even within the disturbed environment might be increased using the skeleton mapping using Modified Multiple Importance Sampling (MMIS) algorithm.

System Methodology

The image sequence is obtained from the only monocular camera in real time is taken into account because the input for processing and recognition of the thing. The technique avoids the distractions and disturbances caused by persons aside from the topic or the other objects which are the main disadvantage in many camera-based applications. the method consists of following steps as shown within the Figure 2. Initially, the input is given as video format. The video input is converted into frames by frame conversion process. The frame conversion process converts the video input into several frames. Therefore, the processing are often wiped out to individual frames. This frame rate up conversion increases the temporal rate during a progressive video and therefore the clear picture is obtained especially during a slow-motion video. The diagram of the method is explained as follows. to trace the motion of the parts, colour histogram is completed to calculate the distribution of colour of pixels in a picture. The output are going to be statistical graph sort of distribution of colour, which is employed to seek out the minute variation within the movements of the body parts. to trace a person the initialization of head and tracking with reference to the top is extremely important. So, because the first part, the top is initialized which is directly facing the camera and it's tracked continuously each and each frame. In cluttered environment to trace the arms, head and torso the automated initialization supported complexion detection on body parts and edges on hands are incorporated using complexion on edges and hands detector. After the initialization, the human upper body parts detection and

tracking has been done through MMIS algorithm to efficiently track the top and arms poses under occlusion. Finally, the tracked outputs are feed to the Convolutional Neural Network (CNN) classifier for individual body parts classifications and its corresponding action recognitions. The algorithms are explained within the next section.

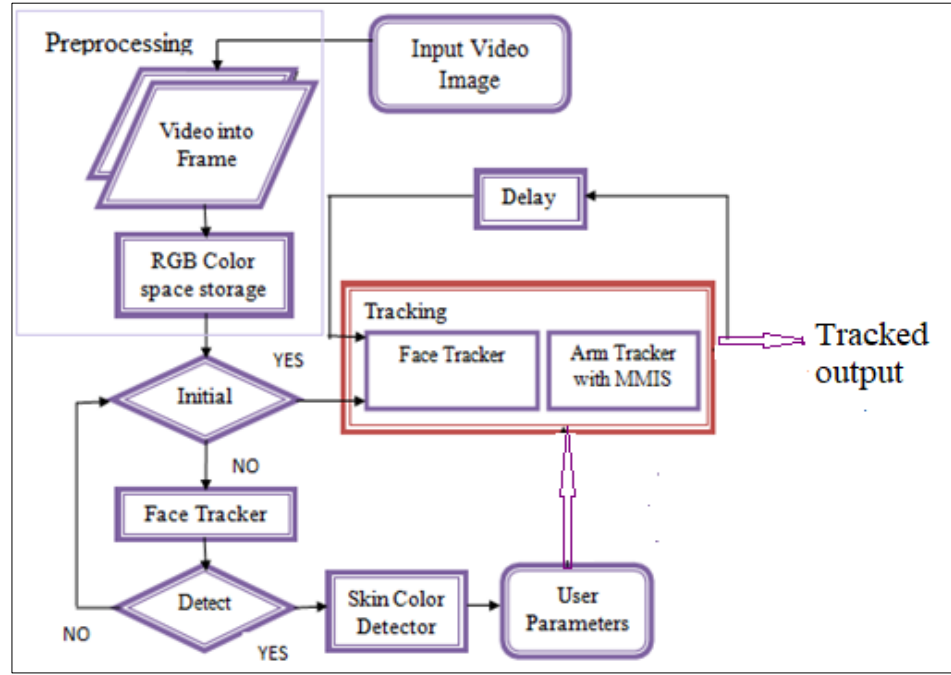


Fig 2: Human Body Part Tracking Technique

Histogram Fogram equalisation for the frame enhancement

This is the pre-processing stage of physical body parts detection and tracking. Here, the video input is read and converted into the frames. the 2 chrominance channels are parted. the amount of rectangular contextual tiles into which the image is split is obtained for every chrominance channel. The optimal value is set within the trial. The source for creating the contrast transform function springs from the uniform distribution. Let $i_{(c_min)}$ and $i_{(c_max)}$ be the minimum and maximum permissible intensity levels. Also optimal value of this clip limit should even be set. For the input contextual tile $i_{(c_in)}$, $F_k(i_{(c_in)})$ be the cumulative distribution function. The uniform distribution of the modified chrominance channel tile is given by the below equation. The flowchart for the tactic is given within the Figure 3.

$$\begin{bmatrix} Y \\ Cb \\ CR \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112.000 \\ 112.000 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

$$i_{c_out} = [i_{c_max} - i_{c_min}] * F_k(i_{c_in}) + i_{c_min} \quad (2)$$

Human upper body part tracking based on MMIS (Modified multiple importance sampling algorithm)

The main goal of the chapter is to offer a strategy to estimate the physical body parts accurately under occlusion through MMIS tracking algorithm. As shown within the Figure 3, the primary step here is to initialize the head that's appeared within the camera in order that the corresponding persons head, torso and arms of the upper body parts of the human has been tracked. After the top initialization to spot the various body parts complexion and hues on joints has been detected, then the human kinematic model has been went to find the angles associated with the locations of arms with reference to the shoulder for tracking the upper body parts of human.

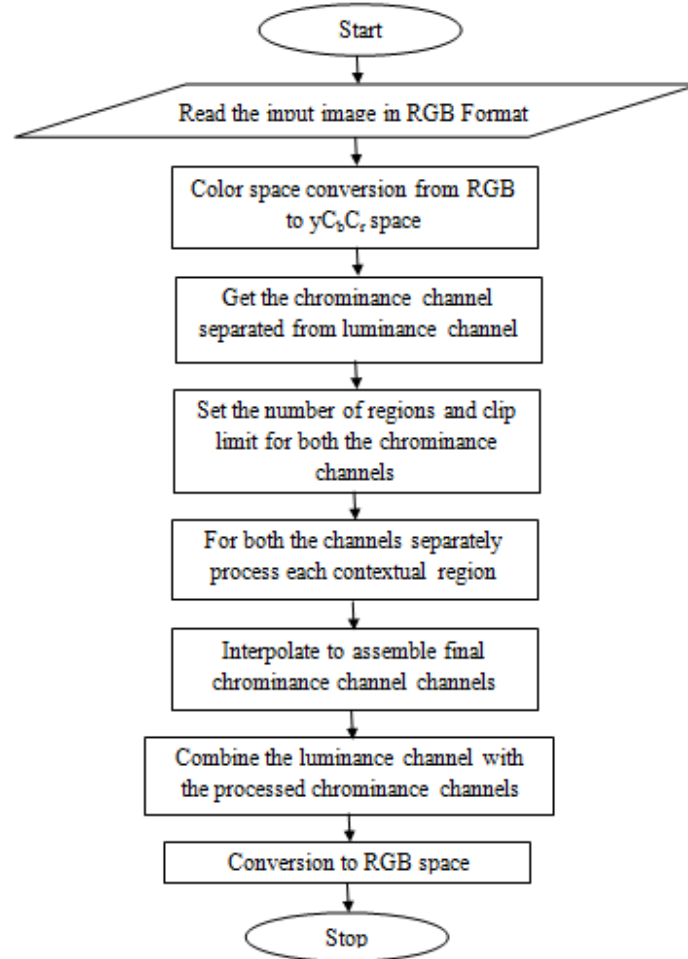


Fig 3: Flow chart of Histogram equalization

Results and Discussion

The experimental setup and results for the MMIS algorithm is described within the following section. The video arrangements caught by a Logitech webcam are handled by a private Computer (PC) with an Intel Core3, 2.10 GHZ and 4GB RAM processor. The input from the real-time video is converted into frames using video acquisition system for the pre-processing step. The frame rate up-conversion increases the temporal rate in progressive video. National Instrument Satellite Corporation (NTSC) format requires 60 fields per sec. The samples of comprise Conversion from a digitized movie recorded with a temporal rate of 24 frames per sec is taken into account because the frame rate conversion. Further, the converted frames undergo complexion and edge colour detections. Then supported kinematic model the locations of shoulder, upper arm and lower arm are obtained through Denavit-Hartenberg parameters. Next, the particle filtering process takes place so as to trace arm hypotheses and therefore the various cues obtained through the particle filter states are wont to obtain the body orientation track so as to yield the efficient tracking performance.

Root mean square error (RMSE): The basis Mean Square Error (RMSE) describes how accurately tracker tracks the objects through comparison between the anticipated values to the particular value. The RMSE gives variance of the model prediction error. A smaller of RMSE indicates better performance of the system. the basis Mean Square Error (RMSE) is giving a way of the anticipated values error and also indicates how close the anticipated values are to the particular values. The RMSE are often computed from the formula,

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (X_{\text{actual}, i} - X_{\text{practical}, i})^2}{N}} \quad (3)$$

Where $X_{\text{actual}, i}$ is actual values
 $X_{\text{predicted}, i}$ is predicted values,
 i represent the current predictor,
 N represents the number of predictors.

Standard deviation error (STDE): To validate the system performance in terms of valid contrast of the image the quality deviation error has been used. the quality Deviation Error are often computed from the formula,

$$\sigma_x = \left(\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2 \right)^{1/2} \quad (4)$$

Where μ_x is mean of an sample image X, and
 N is Number of samples predicted.

And, the mean intensity μ_x is expressed as,

$$\mu_x = \bar{x} = \frac{1}{N} \sum_{x=1}^N x_i \quad (5)$$

Although the background and contextual information is beneficial and will be taken into consideration, its contribution to the ultimate representation should be but the act. Since the act recognition has evoked considerable interest within the various applications due to its potential use within the proactive computing, act recognition has induced required interest in various applications. To handle emergency situations at health care or life care these technologies are widely used. A system which is capable of recognizing various human activities has many useful applications like automated surveillance systems, human computer interaction, smart home health-care systems and control free gaming systems etc. Thus, act recognition may be a very useful domain with many efficient applications in real time. So, the important time human object captured using the Logitech camera has been used the input for various analysis a part of the algorithm. because the initial step of the algorithm pre-processing for the input video has been done supported colour histogram equalization process and therefore the results are shown within the Figures 4, 5 and 6.

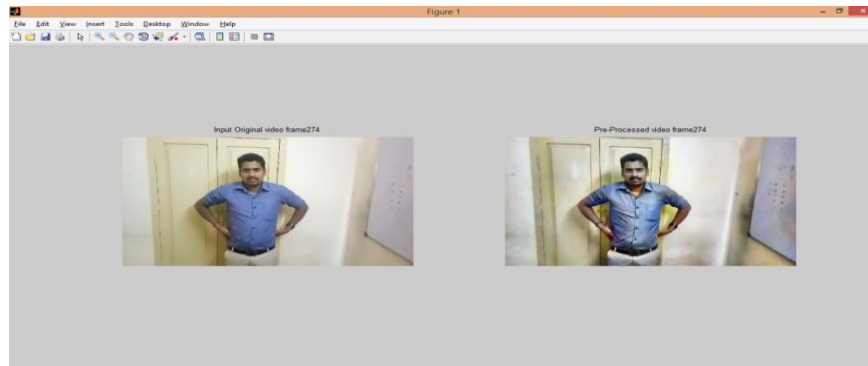


Fig 4: Pre-processed Image of a person with hands holding on hip

The input video with 25 seconds is converted into 500 frames with 30 seconds of frame rate. The video is subjected into different actions like arms are holding hip, straight, towards up and therefore the corresponding actions must be tracked using the proposed tracing and classification algorithms further.

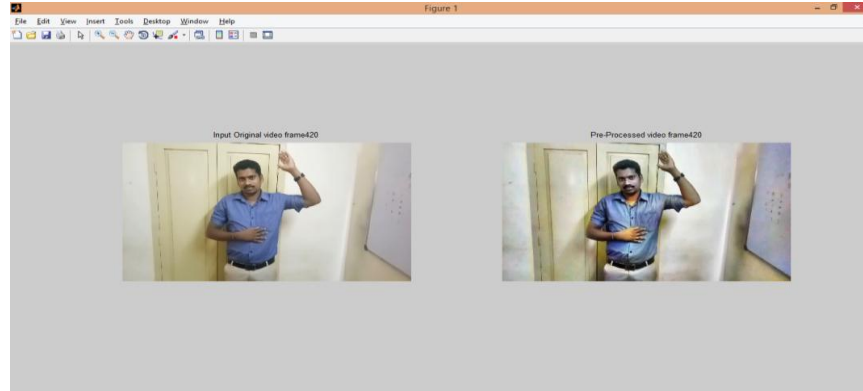


Fig 5: Another pre-processed Image of a person with different arm actions

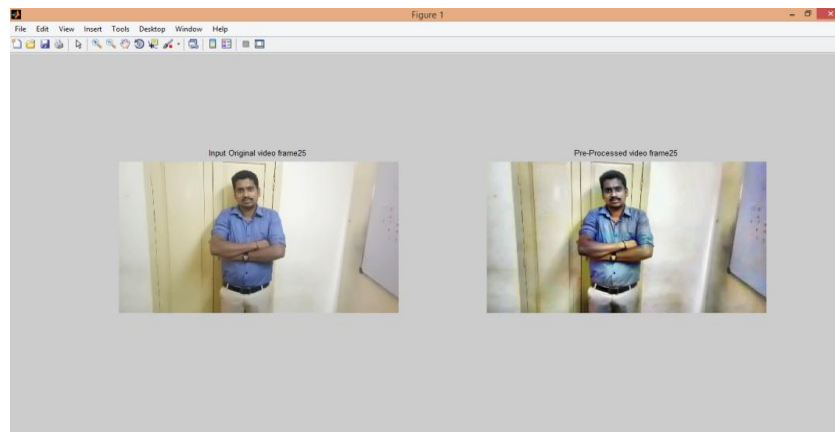


Fig 6: Third action of persons pre-processed image

The frame numbers 420 and 425 defined with other different actions captured from the video and are pre-processed and indicated through the Figures 7, 8 and 9. To research an action the essential interest is within the movements of the body of the topic. The real-time video input obtained from the camera is processed for pre-processing and face has been initialized to trace the actual physical body poses. Next, the skin colours are identified for face and arms to follow their movement and therefore the MMIS tracking algorithm particle filtering outputs over the moving parts are verified. to spot the precise locations of the arms for his or her accurate movement tracking the variations of arms and limbs with reference to the shoulders are found using the length and angle variations.



Fig 7: Initialization of face for tracking the upper body parts through video



Fig 8: Right and Left arm angles calculation with respect to the shoulder

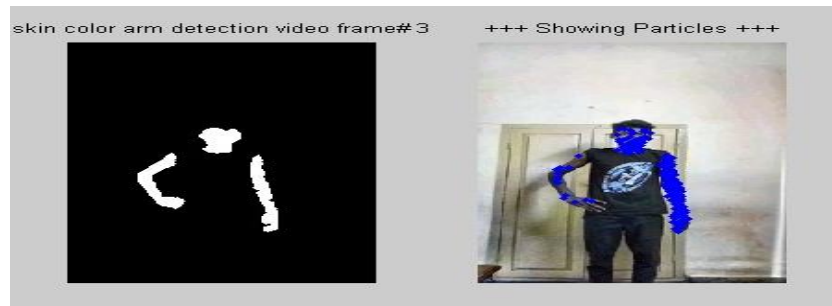


Fig 9: Upper body parts tracking through skin colour detection and MMIS particle filtering

The similar tracking output results are shown in the Figures 10, 11, 12 and 13, when the video is subjected onto the frame 2 and frame 6 respectively.

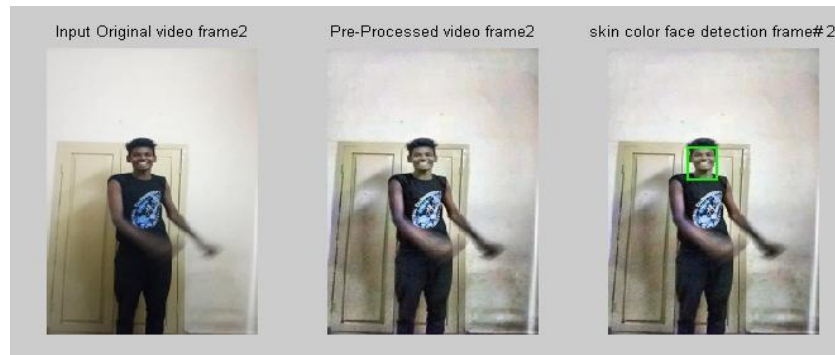


Fig 10: Initialization of face for tracking the upper body parts through video of frame 2

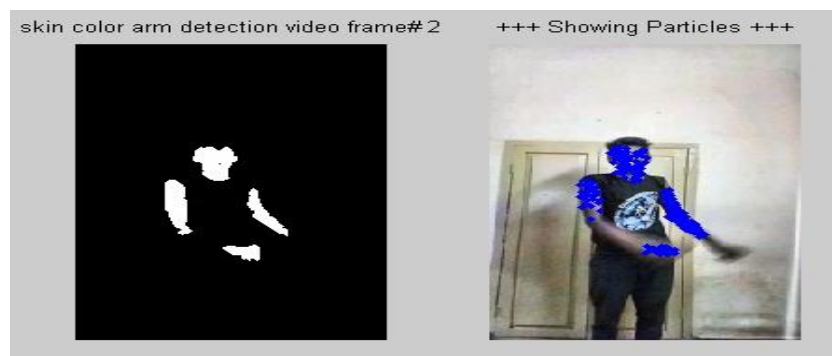
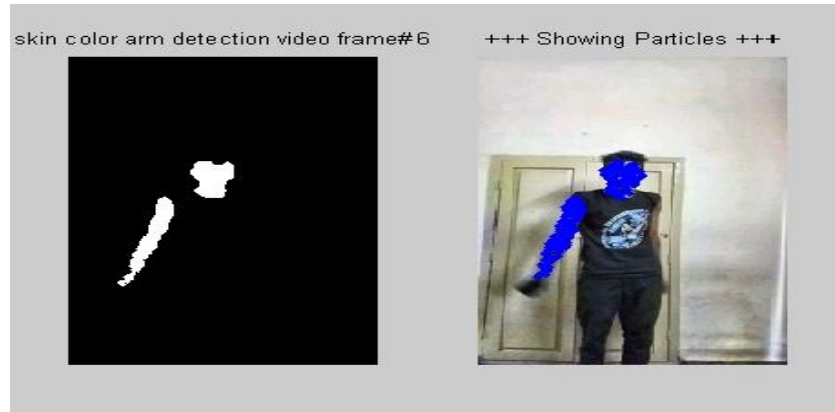


Fig 11: Upper body parts tracking for frame 2**Fig 12: Initialization of face for tracking of frame 6****Fig 13: Human Upper body parts tracking for frame 6**

Conclusion

The implementation of proposed human object detection, tracking and classification algorithms are administered under cluttered environment so as to avoid occlusion problem and also to enhance classification accuracy. The physical body parts action training and tracking method is analyzed using MMIS tracking algorithm for physical body parts action tracking accurately under occlusion. So, the efficiencies of tracking parameters, detecting and classifying the human objects through dynamic object tracking, MVS and MMIS algorithms are computed under occlusion. Finally, the improved tracking and classification accuracies are obtained. Further the algorithm is often utilized in tracking of the human poses that are more complex involved fine variations and differentiations within the small body parts. Computational time might be reduced further by running the system in GPU based processor or FPGA processors.

References

- [1] Levi, K & Weiss, Y, 'Learning object detection from a small number of examples: the importance of good features', IEEE Conference on Computer vision and pattern recognition, 2013, vol. 1, pp. 53-60.
- [2] Suman Sedai, Mohammed Bennamoun & Huynh, Du, k, 'A Gaussian Process Guided Particle Filter for Tracking 3D Human Pose in Video', IEEE Transactions on Image Processing, 2013, vol. 22, no. 11, pp. 4286-4300.
- [3] Ankur Agarwal & Bill Triggs, 'Recovering 3D Human Pose from Monocular Images', IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, vol. 28, no. 1, pp. 44-58.
- [4] Zhenghao Xi, Dongmei Xu, Wanqing Song & Yang Zheng, 'A* algorithm with dynamic weights for multiple objects tracking', The Elsevier journal pattern recognition, 2015, vol. 1, pp. 1-8.
- [5] Jesus Martinez del Rincon, Dimitrios Makris, Carlos Orrite Urunuela & Jean Christophe Nebel, 'Tracking Human Position and Lower Body Parts Using Kalman and Particle Filters Constrained by Human Biomechanics', IEEE Transactions On Systems, Man, And Cybernetics, 2011, vol. 41, no. 1, pp. 26-37.
- [6] Dalal, N & Triggs, B, 'Histograms of oriented gradients for human detection', IEEE Conference on computer vision and patterns recognition, 2005, vol. 2, pp. 1-8.
- [7] Buehler, P, Everingham, M, Huttenlocher, DP & Zisserman, A, 'Long term arm and hand tracking for continuous sign language TV broadcasts', British Machine Vision Conference, 2008, vol. 1, pp. 110.1-110.10.
- [8] David Ribeiro, Nascimento, C, Jac, Alexander Bernardino & Gustavo Carneiro, 'Improving the performance of pedestrian detectors using convolutional learning', The Elsevier journal of pattern recognition, 2017, no. 61, pp. 641-649.
- [9] MunWai Lee & Isaac Cohen, 'A Model-Based Approach for Estimating Human 3D Poses in Static Images', IEEE Transactions On Pattern Analysis and Machine Intelligence, 2006, vol. 28, no. 6, pp. 906-916.
- [10] MunWai, L & Nevatia, R, 'Human pose tracking in monocular sequence using multilevel structured models', Pattern Analysis and Machine Intelligence, IEEE Transactions on Image Processing, 2009, vol. 31, pp. 27-38.