Human Body Extraction and Activities Observation from Still images for High Security systems

[a] **Y.B.T Sundari**, [b]**D.Veena**, [c]**K.V.Mohan**

[a] Assistant Professor, Dept of ECE, Holy Mary Institute of Technology, Hyderabad, India, 501301

[b] Assistant Professor, Dept of ECE, Holy Mary Institute of Technology, Hyderabad, India, 501301

[c]Assistant Professor,Dept of ECE,Holy Mary Institute of Technology and Science,Hyderabad,India,501301

Abstract

Now a days for high-security systems there were different Digital Image processing techniques are developed in that detection and recognition of human activities are difficult task due to problems of background clutter, occlusions, different scaling in view, illumination, intensity and appearance problems. The segmentation method of the human body from the video or still image is difficult task for getting accuracy and efficiency, these extraction and activities of the human body are beneficial methods in situation understanding for high-security regions such as robotics etc. For these various approaches are image processing proposed, until now, but those are all very complex and costly. And getting the complete extraction of a human body from the single image is not an easy thing. So here we evaluate the accuracy of human body extraction from still images using Viola Jones algorithm through three stages and some activities using Skeleton tracking kinect algorithms based on the segmentation process. In Viola Jones algorithm, this segmentation process starts with face extraction including skin color and in next stages, extracting of upper body with skin color through UBS method and lower body with skin through LBS method. Skeleton tracking kinect algorithm has been used for recognition of different activities such as walking, running, climbing, jogging and sitting etc. in this concentrated on sitting posture. The evaluation process on the above methods has been done through the MATALB 8.1(2013), finally the simulation results showed the best performance and high efficiency over a traditional state of art methods

KEYWORDS— EXTRACTION; TRAACKING; UBS, LBS, MATLAB8.1;

I.Introduction

Major contribution of these techniques are surveying on the people count effected by natural disasters and save them in span.Now a days security problems are there everywhere in the world, mainly at country borders, Industries, etc. and this problem is affecting the people with the Bomb basting's even though availability of CC camera's facilitation in some important places. To protect from these conditions, until now, there were so many security systems are developed in Biometric and non-biometric methods and in some security systems, Image Processing is also used. Actually, Digital image processing has a wide range of applications, In that segmentation of the human body from the image is challenging task for getting accuracy and efficiency, this extraction of the human body has a wide range of applications in real time such as scene understanding in high-security regions, robotics, etc. For these various approaches are proposed, until now, but those are all very complex and costly. And getting the complete extraction of the human body in a single attempt is difficult thing. So we are analyzed on an equipped human body extraction from still images using Viola Jones algorithm and Skeleton tracking kinect algorithms based on the segmentation process. With Viola Jones algorithm, this segmentation process has been done in extraction of the standing human body through three stages and first-stage starts with face extraction, including skin color and in next stages, we are extracting upper body with skin color through UBS method and lower body with skin through LBS method for the sitting posture extraction skeleton tracking kinect algorithm has been used and the continuous of the research is going on all positions and activities of the human body. The analysis process on the above methods has been done through the MATALB 8.1(2013). This paper gives a keen idea about the extraction of standing and sitting posture of the human body from still images. The image segmentation process is a challenging task, and it will be used in high-security systems, and the detail explanation is in following sections.

II. Contribution

The major contributions of this study address upright and not occluded poses.

A. Extraction of the standing human body (Viola jones)

1) We analyzed the process for automatic segmentation of human bodies in still images.
2) The total information collected from the different levels of image segmentation, which allows efficient and robust computations upon groups of pixels that are perceptually correlated.
3) Soft anthropometric constraints permeate the whole process and uncover body regions.
4) Without making any assumptions about the foreground and background, except for the assumptions that sleeves are of similar color to the torso region, and the lower part of the pants is similar to the upper part of the pants, we structure our searching and extraction algorithm based on the premise that colors in body regions appear strongly inside these regions (foreground) and weakly outside (background).[1][10].

B. Sitting posture extraction (Skeleton Tracking)

For image access in each camera or image device has one Device ID. Because the Kinect for Windows camera has two separate sensors, the color sensor and the depth sensor. It shows how to create a video input object for the color sensor to acquire RGB images and then for the depth sensor to acquire skeletal data [11][2].

III. State of the art

The word “anthropometry’’ was coined by the French naturalist Georges Cuvier (1769–1832). It was first used by physical anthropologists in their studies of human variability among human races and for comparison of humans to other primates. Anthropometry literally means “measurement of man,’’ or “measurement of humans,’’ from the Greek words anthropos, a man, and metron, a measure. Although we can measure humans in many different ways, anthropometry focuses on the measurement of bodily features such as body shape and body composition (“static anthropometry’’), the body’s motion and strength capabilities and use of space (“dynamic anthropometry’’). [6][9].

Non-rigid object detection and articulated pose estimation are two related and challenging problems in computer vision. Numerous models have been proposed over the years and often address different special cases, such as pedestrian detection or upper

Body pose estimation in TV footage. This paper shows that such specialization may not be necessary, and proposes a generic approach based on the pictorial structures framework. We show that the right selection of components for both appearance and spatial modeling is crucial for general applicability and overall performance of the model. The appearance of body parts is modeled using densely sampled shape context descriptors and discriminatively trained Ada Boost classifiers.

The objective of this paper is to estimate 2D human pose as a spatial configuration of body parts in TV and movie video shots. Such video material is uncontrolled and extremely challenging. We propose an approach that progressively reduces the search space for body parts, to greatly improve the chances that pose estimation will succeed. This involves two contributions: (i) a generic detector using a weak model of pose to substantially reduce the full pose search space; and (ii) employing 'grab-cut' initialized on detected regions proposed by the weak model, to further prune the search space. Moreover, we also propose (iii) an integrated spatiotemporal model covering multiple frames to refine pose estimates from individual frames, with inference using belief propagation.

IV Methodologies

The human body extraction from still images in Standing and sitting posture

A. Extraction of Standing Human body

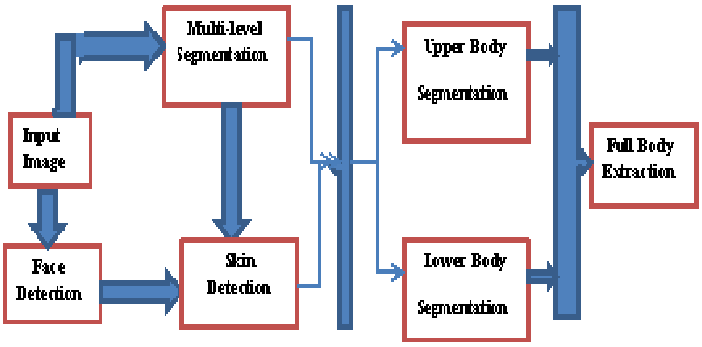


Fig1. Extraction of Standing Human Body

Face Detection

The face detection method is based on facial feature detection and localization using low-level image processing techniques, image segmentation, and graph-based verification of the facial structure. First, the pixels that correspond to skin are detected using the method. Then, the elliptical regions of the detected faces in the image found by the Viola–Jones algorithm are evaluated according to the probabilities of the inscribed pixels. More specifically, the average skin probability of the pixels X of potential face region FRi, for each person i, is compared with threshold T-Global Skin (set empirically to 0.7 in our experiments). If it passes the global skin test (greater than T-Global Skin), it is further evaluated by our face detector. If the facial features are detected, then FRi is considered to be a true positive detection. After fitting an ellipse in the face region, we are able to define the fundamental unit with respect to which locations and sizes of human body parts are Estimated, According To Anthropometric Constraints.[5][1]

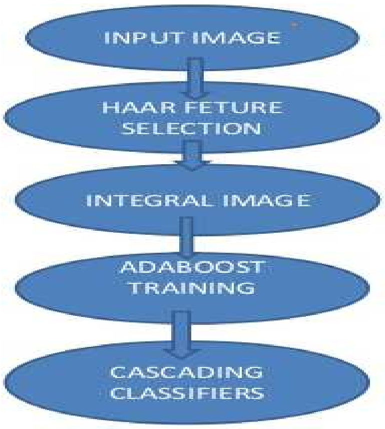


Fig2. Viola-Jones algorithm

Multiple-Level Image Segmentation

In this study, we propose using an image segmentation method, in order to process pixels in more meaningful groups. However, there are numerous image segmentation algorithms, and the selection of an appropriate one was based on the following criteria. First, we require the algorithm to be able to preserve strong edges in the image, because they are a good indication of boundaries between semantically different regions. Second, another desirable attribute is the production of segments with relatively uniform sizes.[1][9][10].

Skin Detection

In this study, we propose combining the global detection technique with an appearance model created for each face, to better adapt to the corresponding human's skin color. The appearance model provides strong discrimination between skin and skin-like pixels, and segmentation cues are used to create regions of uncertainty. Regions of certainty and uncertainty comprise a map that guides the Grab-Cut algorithm, which in turn outputs the final skin regions. False positives are eliminated using anthropometric constraints and body connectivity.

Each image pixel's probability of being a skin pixel is calculated separately for each channel according to a normal probability distribution with the corresponding parameters. We expect true skin pixels to have strong probability response in all of the selected channels. The skin probability for each pixel X is as follows:[6][7].

$$P_{Skin_i}(X) = \prod_{j=1}^{6} \mathcal{N}\left(X, \mu_{ij}, \sigma_{ij}\right) \qquad (1)$$

The adaptive model in general focuses on achieving a high score of true positive cases. However, most of the time it is too "strict" and suppresses the values of many skin and skin-like pixels that deviate from the true values according to the derived probability distribution. At this point, we find that an influence of the skin global detection algorithm is beneficial because it aids in recovering the uncertain areas.

Upper body segmentation (UBS)

In this section, we present a methodology for extraction of the whole upper human body in single images, extending, which dealt with the case, where the torso is almost upright and facing the camera. The only training needed is for the initial step of the process, namely the face detection and a small training set for the global skin detection process. The rest of the methodology is mostly appearance based and relies on the assumption that there is a connection between the human body parts.

Processing using super-pixels instead of single pixels, which are acquired by In this section, we present a methodology for extraction of the whole upper human body in single images, extending, which dealt with the case, where the torso is almost upright and facing the camera. The only training needed is for the initial step of the process, namely the face detection and a small training set for the global skin detection process. The rest of the methodology is mostly appearance based and relies on the assumption that there is a connection between the human body parts. Processing using super-pixels instead of single pixels, which are acquired by an image segmentation algorithm, yield more accurate results and allow more efficient computations.

Here, we use two segmentation levels in this stage of 100 and 200 super-pixels, because they provide a good tradeoff between perceptual grouping and computational complexity [1], [8]

Upper Body Segmentation (UBS)

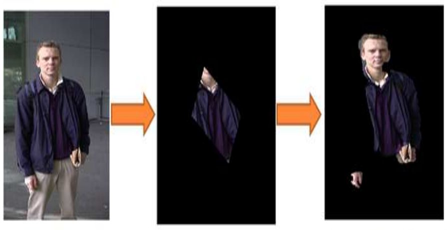


Fig1:Input image

Fig2: Rectangular method for upper body segmentation

Fig 3:collaboration of face and upper body segmentation

$$P_{simIm_{li}}(X) = \prod_{j=1}^{3} \mathcal{N}\left(X, \mu_{ij}, \sigma_{ij}\right) \qquad (2)$$

Sequentially, a searching phase takes place, where a loose torso mask is used for sampling and rating of regions according to their probability of belonging to the torso. Since we assume that sleeves are more similar to the torso colors than the background, this process combined with skin detection actually leads to upper body probability estimation.

Our approach has the advantages of taking different perceptual groupings into account and being able to alleviate the need for accurate torso mask estimation, by conjunctively measuring the foreground and background potentials. The fact that we use super pixels in the computations makes comparisons more meaningful, preserves strong boundaries, and improves algorithmic efficiency. Results may be improved by adding more segmentation levels and masks at different sizes and locations, but at the cost of computational complexity.

We can achieve accurate and robust results without imposing computational strain. The obvious step is to threshold the aggregated potential torso images in order to retrieve the upper body mask. In most cases, hands or arms' skin is not sampled enough during the torso searching process, especially in the cases, where arms are outstretched. Thus, we use the skin masks estimated during the skin detection process, which are more accurate than in the case they were retrieved during this process, since they were calculated using the face's skin color, in a color space more appropriate for skin and segments created at a finer level of segmentation. These segments are superimposed on the aggregated potential torso images and receive the highest potential (1, since the potentials are normalized). Instead of using a simple or even adaptive thresholding, we use a multiple level thresholding to recover the regions with strong potential according to the method described, but at the same time comply with the following criteria: 1) they form a region size close to the expected torso size (actually bigger in order to allow for the case, where arms are outstretched), and 2) the outer perimeter of this region overlaps with sufficiently high gradients. The distance of the selected region at thresholdt (Region t) to the expected upper body size (Exp Upper Body Size) is calculated as follows:

$$ScoreSize = \frac{-|Region_t\_ExpUpperBodySize\ |}{ExpUpperbody} \quad (3)$$

Where Exp Upper Body Size=11×PL 2. The score for the second criterion is calculated by averaging the gradient image (GradIm) responses for the pixels that belong to the perimeter (PRegiont) of Region as

$$ScoreGrad = \frac{1}{|PRegion_t|} \sum^{|PRegion_t|} GradIm \cap PRegion_t \qquad (4)$$

Segmentation (LBS Lower Body)

The algorithm for estimating the lower body part, in order to achieve full body segmentation is very similar to the one for upper body extraction. The difference is the anchor points that initiate the leg searching process. In the case of upper body segmentation, it was the position of the face that aided the estimation of the upper body location. In the case of lower body segmentation, it is the upper body that aids the estimation of the lower body's position. More specifically, the general criterion we employ is that the upper parts of the legs should be underneath and near the torso region. Although the previously estimated UBR provides a solid starting point for the leg localization, different types of clothing like long coats, dresses, or color similarities between the clothes of the upper and lower body might make the torso region appear different (usually longer) than it should be. To better estimate the torso region, we perform a more refined torso fitting process, which does not require extensive.

Lower Body Segmentation (LBS)

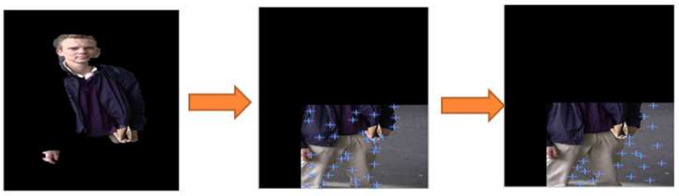


Fig1: Input image

Fig 2:Foreground selection

Fig 3:Background selection

Computations, since the already estimated shape provides a very good guide.[1]

The expected dimensions of the torso are again calculated based on anthropometric constraints, but in a more accurate model. In addition, in order to cope with slight body deformations, we allow the rectangle to be constructed according to a constrained parameter space of highest granularity and dimensionality. Specifically, we allow rotations with respect to rectangle's center by angleφ, translations in x-andy-axes,τx andτy and scaling inx- andy-axes,sx andsy. The initial dimensions of the rectangle correspond to the expected torso in full frontal and upright view and it is decreased during searching in order to accommodate other poses. The rationale behind the fitting score of each rectangle is measuring how much it covers the UBR, since the torso is the largest semantic region of the upper body, defined by potential upper body coverage (UBC), while at the same time covering less of the background region, defined by potentials(for Solidity). Finally, in many cases, the rectangle needs to be realigned with respect to the face's center (Face Center) to recover from misalignments caused by different poses and errors. A helpful criterion is the maximum distance of the rectangle's upper corners (L Shoulder, R Shoulder) from the constrained. Thus, fitting of the torso rectangle is formulated as a maximization problem [1][7][9]

$$\theta maxf(\theta) = \alpha_1 \times UBC(\theta) + \alpha_2 \times s(\theta) + \alpha_3 \times D_{sf}(\theta) \quad (5)$$

where Torso Mask(θ) is the binary image, where pixels inside the rectangle r Torso Mask(θ) are 1, else 0; UBR is the binary image, where pixels inside the UBR are 1, else 0; a1,a2,a3 are weights, set to 0.4, 0.5, and 0.1, respectively[1][3] [5].

Sitting posture (Skeleton Tracking with kinect1 and kinect2**)**

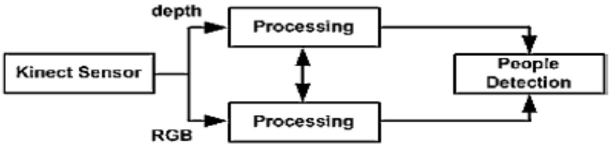


Fig 3: Simple methodology of Skeleton Tracking

In Figure 2 we show the basic ideas of two schemes. Here, scheme A employs the depth images only while scheme *B*

takes the advantage of complementary data emanating from the two vision sensors of Kinect. The algorithm presented in aims at detecting people based on depth information obtained by Kinect in indoor environments [2][4].

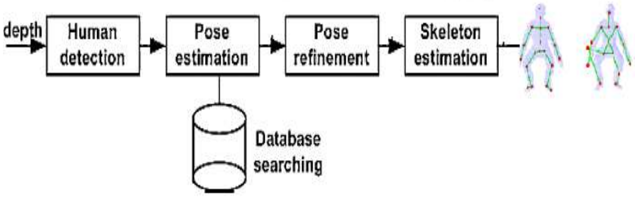


Fig 4: Human detection with activity

Analyzing human detection activities from video is an area with increasingly important consequences from security/surveillance to entertainment [2].

We will refer to this as pose estimation, because its goal is to achieve either a faster or a more accurate skeletal joints approximation. The research in the second category is called activity recognition, since it steps forward to recognize the semantic activity of a human in the context of various applications. Briefly speaking, pose estimation, in this scenario, provides the position of skeletal joints , while the activity recognition tells what the human is doing through analyzing temporal patterns in these joint positions.

Steps with Matlab

Create the video input object for the color sensor.

Observe the device-specific properties on the source device, which is the color sensor on the Kinect camera

Adjust the acquisition for this by setting the Backlight Compensation property to Low Lights Priority, which favors a low light level.

Preview the color stream by calling preview on the color sensor object created in step 1

Create the video input object for the depth sensor. Note that a second object is created (vid2), and Device ID 2 is used for the depth sensor.

Observe the device-specific properties on the source device, which is the depth sensor on the Kinect.

Start the second video input object (the depth stream.

Skeletal data is accessed as metadata on the depth stream

Observe any individual property by drilling into the metadata (Is Skeleton Tracked property)

Get the joint locations for the first person in world coordinates using the Joint World Coordinates property. Since this is the person in position 1, the index uses 1.

View the segmentation data as an image

The Body Posture property, in step 5, indicates whether the tracked skeletons are standing or sitting. Values are standing (gives 20 point skeleton data) and Seated (gives 10 point skeleton data, using joint indices 2 – 11

Activities Observations with Skeleton Tracking

NAME	POSE	DESCRIPTION
1. Shallow Squats	Sitting	Stand-to-sit movements without
2. Chair Stands	Sitting	Sit-to-stand movements..
3. Abs in, Knee Lifts	Sitting	Alternating knee lifts.
4. Lateral Stepping	Sitting	Alternating front and side stepping.

V. Simulation Results

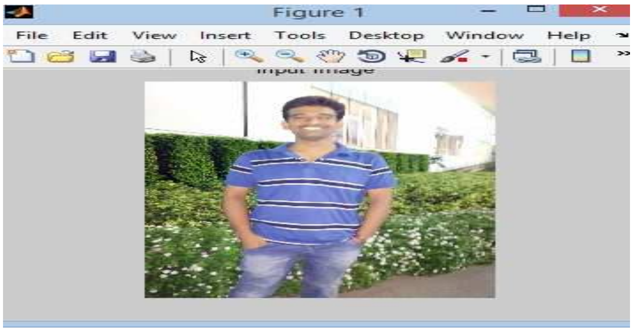
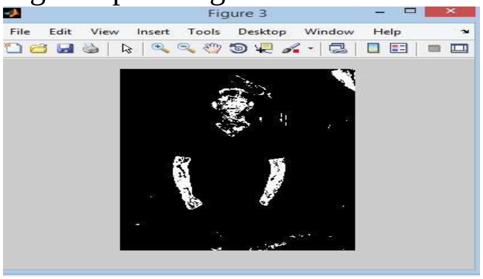


Fig 1. Input image



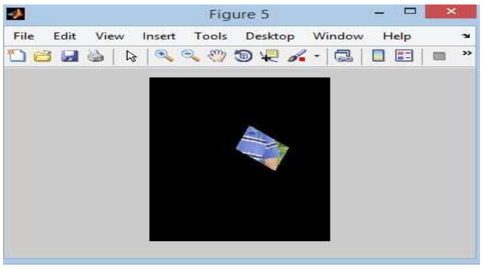


Fig 2. Face detection

Fig 3. Rectangular method for upper body detection

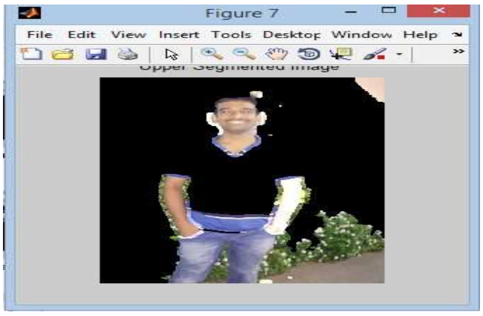


Fig 4. Collaboration of face and upper body Segmentation

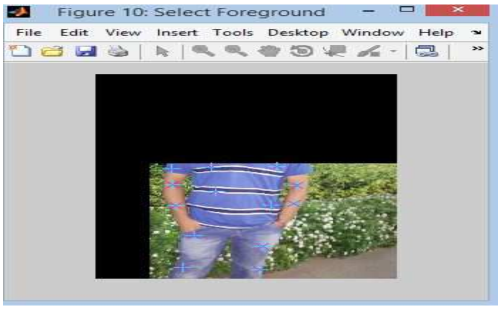


Figure 10: Select Foreground
File Edit View Insert Tools Desktop Window Help

Fig 5. Foreground selection

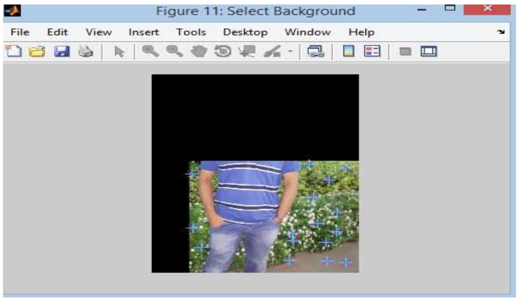


Figure 11: Select Background
File Edit View Insert Tools Desktop Window Help

Fig 6.Background selection

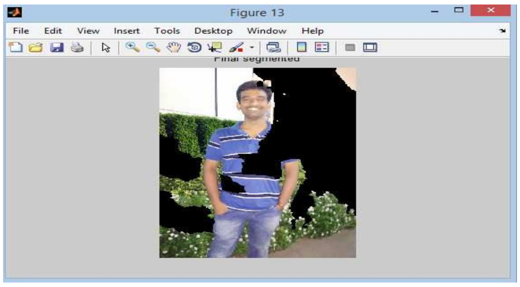


Fig 7 Final result

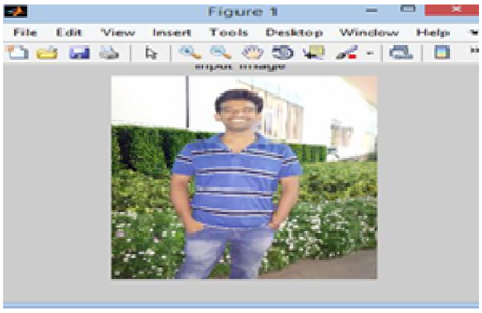


Fig 8. Input image (*spline regression*)

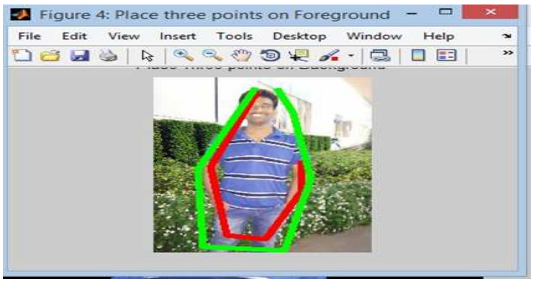


Figure 4: Place three points on Foreground
File Edit View Insert Tools Desktop Window Help
Place Three points on Background

Fig 9. Placing three points on foreground (*Spline Regression*)

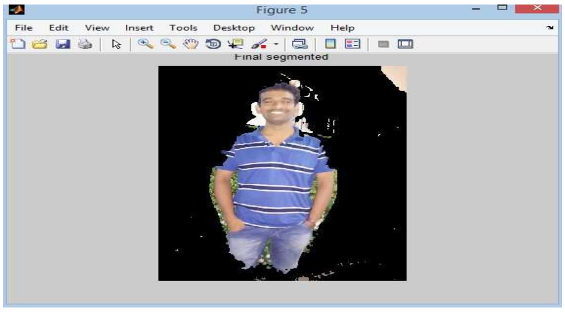


Fig 10 Final segmentation (Spline Regression)

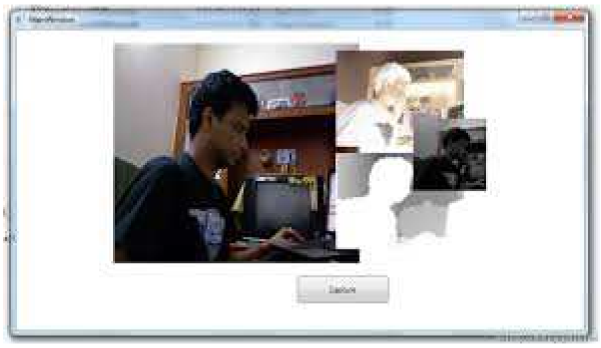
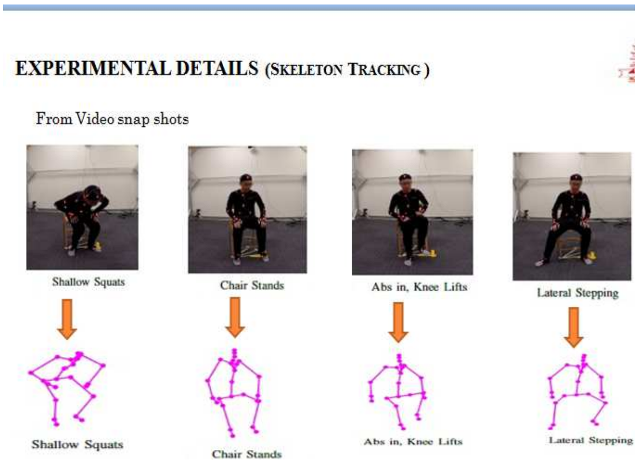


Fig11. Side view of the Human in sitting posture (*skeleton tracking kinect)*

SKELETON TRACKING (By video snapshots)



EXPERIMENTAL DETAILS (SKELETON TRACKING)

From Video snap shots

AVERAGE JOINT POSITION OFFSETS WITHOUT OUTLIERS (Sitting). Mean and Stnddard Deviation(SD) analysis as shown in below

Kinect 1						Kinect 2					
Mean (%)			SD (%)			Mean (%)			SD (%)		
0^0	30^0	60^0	0^0	30^0	60^0	0^0	30^0	60^0	0^0	30^0	60^0
5	8	6	35	38	35	2	5	3	23	36	23

VI. CONCLUSION

In these advanced methodologies, an ease way is observed throughout a segmentation process to extract the standing human bodies and with the skeleton tracking algorithm sitting posture is also examined. Here the multi levels of segmentation in bottom-up approach helped in finding salient regions of high potential of belonging to the human body. The main thing of this methodology is the face detection here we can assume the rough location of the body, and also we can make a rough anthropometric and skin color model. These models directed the most visible parts namely upper and lower body such as pose of the body. And the sitting posture of the side view observed using skeleton tracking kinect these processes examined on a challenging set of data in the MATLAB.

The results are shown that algorithms can well performed the state-of-the-art segmentation algorithms, and cope with various types of standing everyday poses. In the continuing research extraction of the human body in all positions, the restricted poses such as unusual pose, Human daily activities and occlusion problems (missing extreme regions, such as hair, shoes, and gloves can be solved by incorporation of more masks) will be solved because this extraction of the human body has a wide range of applications in real time such as in medical fields providing unhuman medical surgeries with robotic touch. Surveying for people recovery within less span of time after the occurance of natural disaster and other some other real time applications like military etc.

References

[1] Athanasios Tsitsoulis, Member, IEEE, and Nikolaos G. Bourbakis, Fellow, IEEE "A Methodology for Extracting Standing Human Bodies From Single Images" in IEEE Transactions on Human-Machine Systems Volume: 45, Issue: 3, June 2015.

[2] Jungong Han, Member, IEEE, Ling Shao, Senior Member, IEEE, Dong Xu, Member, IEEE, and Jamie Shotton, Member, IEEE "Enhanced Computer Vision with Microsoft Kinect Sensor: A Review" in IEEE TRANSACTIONS ON CYBERNETICS, VOL. 43, NO. 5, OCTOBER 2013

[3] M. Andriluka, S. Roth, and B. Schiele, "Pictorial structures revisited: People detection and articulated pose estimation," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2009, pp. 1014–1021.

[4] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (VOC) challenge,"Int. J. Comput. Vis., vol. 88, no. 2, pp. 303–338, 2010.

[4] V. Ferrari, M. Marin-Jimenez, and A. Zisserman, "Progressive search space reductions for human pose stimation," inProc. IEEE Conf. Comput. Vis. Pattern Recog., 2008, pp. 1–8.

[5] M. P. Kumar, A. Zisserman, and P. H. Torr, "Efficient discriminative learning of parts-based models," in Proc. IEEE 12th Int. Conf. Comput. Vis., 2009, pp. 552–559.

[6] V. Delaitre, I. Laptev, and J. Sivic, "Recognizing human actions in still images: A study of bag-of-features and part-based representations," in Proc. IEEE Brit. Mach. Vis. Conf., 2010.

[7] A. Gupta, A. Kembhavi, and L. S. Davis, "Observing human-object interactions: Using spatial and functional compatibility for recognition,"IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 10, pp. 1775–1789, Oct. 2009.

[8] B. Yao and L. Fei-Fei, “Grouplet: A structured image representation for recognizing human and object interactions,” inProc. IEEE Conf. Comput. Vis. Pattern Recog., 2010, pp. 9–16.
[9] P. Buehler, M. Everingham, D. P. Huttenlocher, and A. Zisserman, “Long term arm and hand tracking for continuous sign language TV broadcasts,” inProc. 19th Brit. Mach. Vis. Conf., 2008, pp. 1105–1114.
[10] A. Farhadi and D. Forsyth, “Aligning ASL for statistical translation using a discriminative word model,” inProc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog., 2006, pp. 1471–1476.
[11] L. Zhao and L. S. Davis, “Iterative figure-ground discrimination,” inProc. 17th Int. Conf. Pattern Recog, 2004, pp. 67–70.
[12]http://in.mathworks.com/help/imaq/acquiring-image-and-skeletal-data-using-the-kinect.html