DOI: 10.21917/ijivp.2010.0012

DEVELOPMENT OF 2D HUMAN BODY MODELING USING THINNING ALGORITHM

K. Srinivasan[1], K. Porkumaran[2] and G. Sainarayanan[3]

[1]*Department of Electronics and Information Engineering, Sri Ramakrishna Engineering College, Tamil Nadu, India*
E-mail: srineekvasan@gmail.com
[2]*Department of Electrical and Electronics Engineering, Dr.N.G.P Institute of Technology, Tamil Nadu, India*
E-mail: porkumaran@gmail.com
[3]*ICT Academy of Tamil Nadu, Tamil Nadu, India*
E-mail: sainarayanan@ictact.in

Abstract
Monitoring the behavior and activities of people in Video surveillance has gained more applications in Computer vision. This paper proposes a new approach to model the human body in 2D view for the activity analysis using Thinning algorithm. The first step of this work is Background subtraction which is achieved by the frame differencing algorithm. Thinning algorithm has been used to find the skeleton of the human body. After thinning, the thirteen feature points like terminating points, intersecting points, shoulder, elbow, and knee points have been extracted. Here, this research work attempts to represent the body model in three different ways such as Stick figure model, Patch model and Rectangle body model. The activities of humans have been analyzed with the help of 2D model for the pre-defined poses from the monocular video data. Finally, the time consumption and efficiency of our proposed algorithm have been evaluated.

Keywords:
Human Pose Analysis, 2D Human Body Modeling, Pre-processing, Frame Differencing Algorithm, Thinning Algorithm

1. INTRODUCTION

In the present decade, the human activity recognition has become a very important field for computer vision researchers due to its applications in many areas such as tracking of human, motion-based identification, computer-human interactions, and automated visual surveillance[1]-[4]. However, the activity analysis from videos is much more complex for the tracking, detection, modeling, and recognition of human. For the automated video surveillance, modeling based analysis is useful for all videos with different persons in real time applications. Marker based human tracking and modeling is a simple way of approach but it cannot reconstruct all the human poses in practical situations. This approach needs markers at every time in surveillance persons. So, the marker-less motion tracking and modeling has been very important in motion analysis. The marker-less human body modeling has been well suited for the motion analysis and improves the efficiency of the system. In this paper, a new approach has been proposed for 2D modeling using thinning algorithm for the motion analysis. Thinning is a morphological operation that can be used to remove the selected foreground pixels from binary images. Model based action analysis involves 2D and 3D human models representation. A 3D model [5]-[7] may have higher accuracy with higher computational complexity, whereas a 2D model [8]-[9] may have lower accuracy with lower computational complexity.

Feifei Huo [10] presents an approach to capture markerless human motion and recognize the human poses. Here, 2D model has been used for torso detection, tracking and 3D location of these body parts and are calculated for pose recognition. The simple image processing techniques can also be used for extracting feature points from 2D human figures with or without markers as in [11]. Antoni Jaume I Capo et. al [12]develop an algorithm for building a human body model skeleton in an automatic way. This algorithm is based on the curvature of a B-Spline parameterization of the human contour. In the previous work, the thinning algorithm is mostly attempted for several image processing applications like pattern recognition and character recognition [13]-[16]. The objective of this paper is to apply thinning algorithm to model the human body in 2D view and it finds the motion analysis in monocular video sequences in indoor environment without using any markers.

The organization of this paper is as follows. Section.1 overviews the introduction and applications of our work. The proposed approach is then discussed in section 2. The frame differencing algorithm that provides background subtraction is illustrated in section 3. Section 4 describes the Morphological operations used to demonstrate dilation and erosion operation. The thinning algorithm is introduced in section 5. In section 6, the determination of body feature points is given. The human activity analysis for different postures is explained in section 7. Section 8 deals the results and discussion of our proposed work. The paper is concluded in section 9.

2. PROPOSED APPROACH

Human activity analysis needs high level description of human actions from the video sequences. Human body modeling has been used in activity analysis of humans in indoor as well as outdoor surveillance environment. The features that are extracted from the human body are useful to model the surveillance of persons and it has been applied to recover the human poses and in finding their activities. In the proposed work, 13 feature points have been considered for a full body modeling and these features are Head, Neck, Left shoulder, Right shoulder, Left hand elbow, Right hand elbow, Abdomen, Left hand, Right hand, Left knee, Right knee, Left leg and Right leg as in Fig.1.

In the first stage of the proposed work, the Video sequence is acquired by the high resolution Nikon COOLPIX Digital Video Camera having 8.0 million effective pixels and 1/2.5-in.CCD image sensor which produces NTSC and PAL video output. The video sequence has been taken at a rate of 30 frames/second from the indoor surveillance environment for finding the human behaviour. After that, the video sequence is converted into individual frames with the help of the algorithm. The next step deals with the preprocessing of the acquired frames. The quality of video sequence may depend on lighting and environmental changes. The key function of preprocessing is to improve the image such that it increases the chances of success in the other processes.

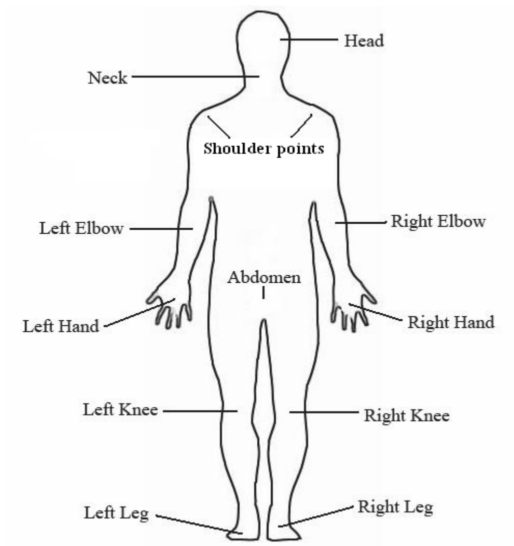


Fig.1.Thirteen feature points of human body model

In this application, the preprocessing techniques are used for enhancing the contrast of the frames, removal of noise and isolating the objects of interest in the frames. The output of the preprocessing technique is shown in Fig.2.

(a) (b)

Fig.2. Image enhancement using preprocessing
(a) Input video frame, (b) Output of Pre-processing

Then, the background subtraction is used to remove the background of the video frame to highlight the foreground image. The next step of the background subtraction is that the morphological operation which helps for the representation and description of region shape, such as boundaries and skeletons. The Thinning algorithm produces skeleton of the human body. Thinning is normally applied only to binary images, and produces another binary image as an output. Then, the thirteen feature points are identified on the thinned image and the activities of human are analyzed for the pre-defined poses. Fig.3. explains the steps of the proposed work for 2D model based activity analysis.

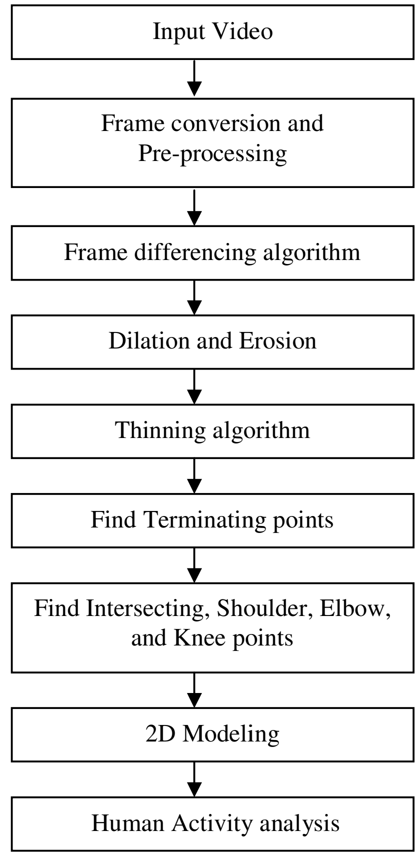


Fig.3. Proposed 2D modeling for activity analysis

3. FRAME DIFFERENCING ALGORITHM

In the proposed work, the background subtraction technique plays an important role in subtracting the foreground images from the background image and it is described in Fig.4. The frame differencing algorithm [17] has been proposed here to highlight the desired foreground scene which include,

Frame Differencing Algorithm

Step.1: Acquire the video sequences.

Step.2: Convert it into video frames.

Step.3: Set the background image.

Step.4: Separate R, G, B components individually for easy computation.

bc_r = bg(:,:,1);bc_g = bg(:,:,2); bc_b = bg(:,:,3);

Step.5: Read the current frame from the video sequence.

Step.6: Separate R, G, B components individually for easy computation.

cc_r = fr(:,:,1);cc_g = fr(:,:,2); cc_b = fr(:,:,3);

Step.7: Subtract the corresponding colour components of the background and current images.

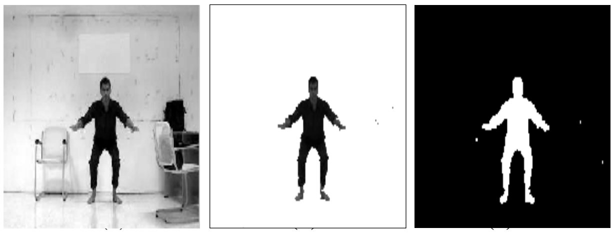


Fig.4. Results of frame differencing algorithm. (a) Input video, (b) Foreground human body, (c) Silhouette of body

4. ROLE OF DILATION AND EROSION

Once, the background is subtracted, the image is given to perform morphological operations. Morphological operations on an image are a type of processing in which the spatial form or structures of objects are modified. In a morphological operation, the each pixel in the input image and its neighbors are compared to find the output image. The size and shape of the neighborhood are to be considered for morphological operation. It includes two basic operations like dilation and erosion [18]. Finally, the noise is removed by using median filtering. Dilation adds pixels to the boundaries of the objects in an image. In binary images, the sets are members of the 2-D integer space Z^2, where each element of a set is a 2-D vector whose coordinates are the (x, y) coordinates of a black or white pixel in the image. Consider the sets A and B in Z^2, the dilation of A by B, is defined by the Eq.(1).

$$A \oplus B = \left\{ z / \left(\overline{B}\right)_Z \cap A \neq \varnothing \right\} \quad (1)$$

The dilation of A by B then is the set of all displacements, z, such that $\overline{B}$ and A overlap by at least one element. Based on this interpretation, the above Eq.(2) may be rewritten as,

$$A \oplus B = \left\{ z / \left[\left(\overline{B}\right)_Z \cap A \right] \subseteq A \right\} \quad (2)$$

Erosion removes the pixels on the object boundaries. To erode an image, *imerode* function has been used for our applications. The *imerode* function accepts two primary arguments. The first argument is that the input image to be processed as grayscale, binary, or packed binary image and the other is the structuring element object, returned by the *strel* function, or a binary matrix defining the neighbourhood of a structuring element. For the sets A and B in Z^2, the erosion of A by B is given by the Eq.(3).

$$A \otimes B = \{ z / (B)_z \subseteq A \} \quad (3)$$

At the end of this stage, the median filtering has been used to reduce the salt and pepper noise present in the frame. It is similar to using an averaging filter, in which each output pixel is set to an average of the pixel values in the neighborhood of the corresponding input pixel.

5. THINNING ALGORITHM

Thinning algorithm plays a vital role in finding the skeleton of human body. The thinning operation is performed by transforming the origin of the structuring element to each pixel in the image. Then it is compared with the corresponding image pixels. When the background and foreground pixels of the structuring element and images are matched, the origin of the structuring element is considered as background. Otherwise it is left unchanged. Here, the structuring element determines the use of the thinning operation. The structuring element determines the number of pixels added or deleted from the objects in an image. It is a matrix that has ones and zeros only with any arbitrary shape and size. In our proposed work, *strel ('line', 6, 90)* function has been used as a structuring element. It develops a flat, linear structuring element. Here, six specifies the length and 90 specify the angle (in degrees) of the line. And the length is approximately the distance between the centers of the structuring element members at opposite ends of the line. Fig.5. shows the structuring elements for Skeletonization by morphological thinning. At each iteration, the image is first thinned by the left hand structuring element, and then by the right hand one, and then with the remaining six *90°* rotations of the two elements. The thinning operation is achieved by the hit-and-miss transform. The thinning of an image *A* by a structuring element *B* is given by Eq.(4).

thin(A,B)=A-hit and miss(A-B) (4)

0	0	0
	1	
1	1	1

	0	0
1	1	0
	1	

Fig.5. Structuring element for thinning operation

This process is repeated in a cyclic fashion until none of the thinning produce any further change. Normally, the origin of the structuring element is at the center. The steps of the thinning algorithm include,

Step: 0 Partitioning the video frame into two distinct subfields in a checkerboard pattern.

Step: 1 Delete the pixel *p* from the first subfield if and only if the conditions *(5)*, (6), and (7) are satisfied.

$$X_H(p)=1 \quad (5)$$

where

$$X_H(p)=\sum_{i=1}^{4} b_i$$

$$bi=\begin{cases}1 \text{ if } X2i-1 = 0 \text{ and } (x2i = 1 \text{ or } x2i+1=1) \\ 0 \text{ otherwise}\end{cases}$$

x1, x2,....x8 are the values of the eight neighbours of *p*, starting with the east neighbour and are numbered in counter-clockwise order.

$$2 \leq \min\left\{n_1(p), n_2(p)\right\} \leq 3 \quad (6)$$

$$n_1(p) = \sum_{i=1}^{4} X_{2k-1} \vee X_{2k}$$

where

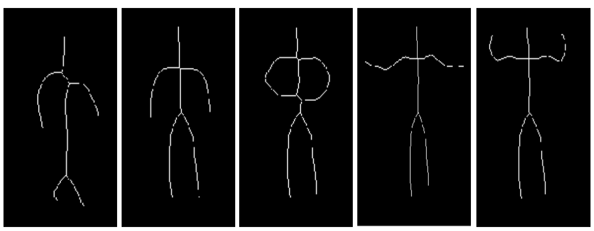
$$n2(p) = \sum_{i=1}^{4} X_{2k} \vee X2k+1$$

$$(X2 \vee X3 \vee \overline{X}8) \wedge X1=0 \qquad (7)$$

Step: 2 Deleting the pixel *p* from the second subfield if and only if the conditions *(5)*, (6), and (8) are satisfied.

$$(X6 \vee X7 \vee \bar{X}4) \wedge X5 = 0 \qquad (8)$$

The step 1 and step 2 together make-up one iteration of the thinning algorithm. Here, an infinite number of iterations ($n=\infty$) have been specified to get the thinned image. Fig.6. shows the results of the thinned algorithm for different activities such as walking, standing, hands in hip, both hands rise and both hands up.



(a) (b) (c) (d) (e)

Fig.6. Results of thinning algorithm for different poses (a) Walking, (b) Standing, (c) hands on hip, (d) Both hands rise, and (e) Both hands up

6. DETERMINATION OF BODY FEATURE POINTS

As in the proposed algorithm, thirteen feature points have been considered in the upper body as well as the lower body. The feature points include the Terminating points (5Nos),Intersecting points (2Nos), Shoulder points (2 Nos), Elbow joints (2Nos), and Knee joints (2Nos).Using the terminating points, the ends of the features such as head, hands and legs have been determined. The following are the steps involved in determining the terminating points.

Step.0: Input the thinned image.

Step.1: Initialize the relative vectors to the side borders from the current pixel.

Step.2: Select the current coordinate to be tested.

Step.3: Determine the coordinates of the pixels around this pixel.

Step.4: If this pixel is an island, then it is an edge to the island of 1 pixel. Save it.

Step.5: Default assumption: pixel is an edge unless otherwise stated.

Step.6: Test all the pixels around this current pixel.

Step.7: For each surrounding pixel, test if there is a corresponding pixel on the other side.

Step.8: If any pixels are on the opposite side of the current pixel, then this pixel is not a terminating pixel.

Step.9: If the current pixel does not satisfy the above condition, then it is an edge.

Then the two intersecting points are found after finding the terminating points and these points are used to join the hands and legs. Then, the two shoulder points are determined. The left shoulder co-ordinate is plotted at the pixel where the iteration encounters a white pixel. Similarly the right shoulder co-ordinate is plotted using the same technique. Fig.7. gives an idea of determining the Shoulder, Elbow and Knee of the human body in graphical way.

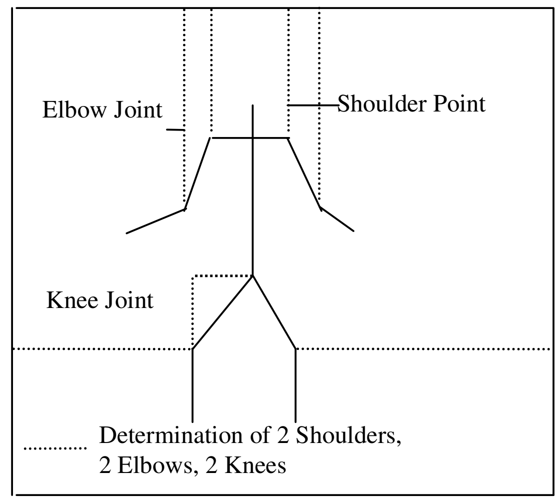


Fig.7. Determination of Shoulder, Elbow and Knee points

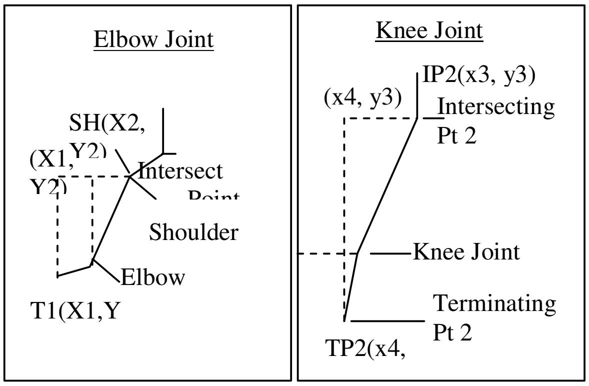
The elbow point is approximately halfway between the shoulder and the terminating points of the two hands. The problem arises when the hand is bent. In order to get the accurate elbow joint, a right angle triangle is constructed as in Fig.8(a).The (x1, y2) point of the right angled triangle is determined by obtaining the x-axis of the terminating point-1 (x1) and the y-axis of the shoulder point (y2). The distance between the point (x1, y2) and (x2, y2) is calculated by using the Eq.(9).

$$\text{Distance between points} = \sqrt{(x_1-x_2)^2+(y_1-y_2)^2} \qquad (9)$$

$$\text{Elbow Joint} = \frac{\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}}{2} \qquad (10)$$

Using the available distance as the x-axis reference, a *for loop* is iterated from the first point of the same x-axis. The point at which the iteration encounters a white pixel is plotted as the elbow joint. Similarly, the other elbow joint is also determined using the same technique. The process of determining the knee joints is similar to the technique adopted for determining elbows. Fig.8(b) shows the graphical method to determine the knee joint. But, in this case the loop is iterated with a constant y-axis and a varying x-axis. The elbow joint is identified using the Eq.(10).

After the determination of the thirteen points, it has been displayed.



(a) (b)

Fig.8. Graphical idea to find Elbow joint and Knee joint (a) Determination of Elbow joint, (b) Determination of Knee joint

7. HUMAN ACTIVITY ANALYSIS

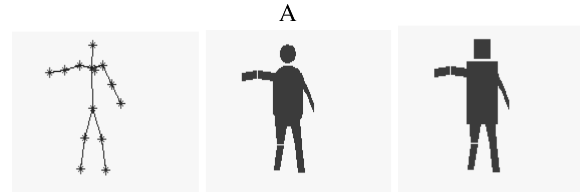
Human activity analysis is achieved with the help of template matching method. Here, the silhouette extracted from the human body is compared with the data base. Once the video is converted to frames, then the human body is segmented and it is converted into silhouettes. The silhouette has been used as a template for further processing. This can be used as a data base for pre-defined twelve activities. So, the current image silhouette is compared with database with the help of correlation function (*corr2*). The correlation is a function which is used to find the relation between two frames. By the experimentation with the various videos, the correlation threshold value is set to 0.9. This function matches each pixel of both the images and outputs the similarity value between two images. If there is any similarity between the images then it shows value near and approximate to 1 else 0. The algorithm of post processing include,

Post processing Algorithm	
Step.0:	Create a data base having pre-defined human silhouettes.
Step.1:	Read the video sequence and segment the human body.
Step.2:	Convert it into silhouette form.
Step.3:	Apply correlation function (*corr2*) between the data base and current frame.
Step.4:	If the correlation of activity is greater than the threshold value of 0.9, then the algorithm outputs the corresponding activity.

8. SIMULATION RESULTS

To test the efficiency and effectiveness of the proposed approach, more than 50 videos are considered. In each sequence, different human poses are taken for the analysis. The proposed algorithm has been developed using MATLAB 7.6(2008a) on Intel dual core processor, 2 GB RAM and Windows XP SP2. The algorithm is implemented in the indoor surveillance video with straight poses for twelve activities such as Standing, Right hand rise, Left hand rise, Both hands rise, Right hand up, Left hand up, Both hands up, Right leg rise, Left leg rise, Right salute, Left salute, and Crouching as in Fig.9.

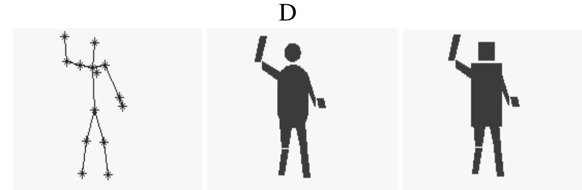
A



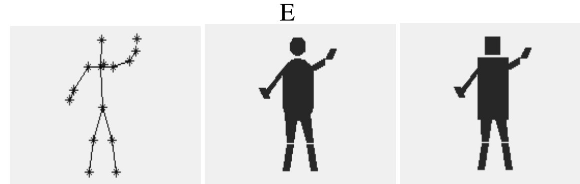
B

C

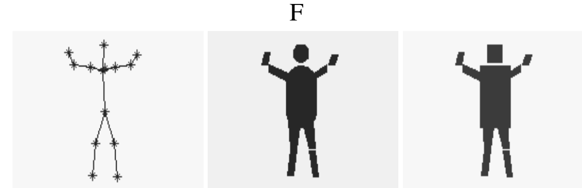
D



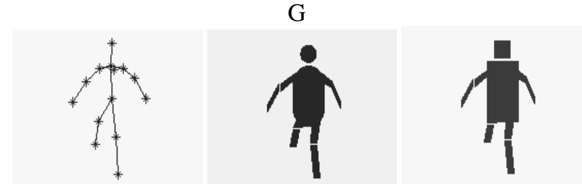
E



F



G

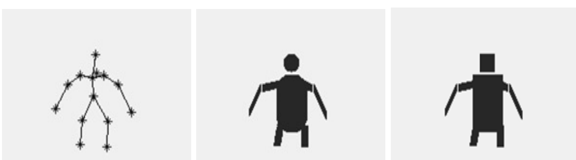


H

I

J

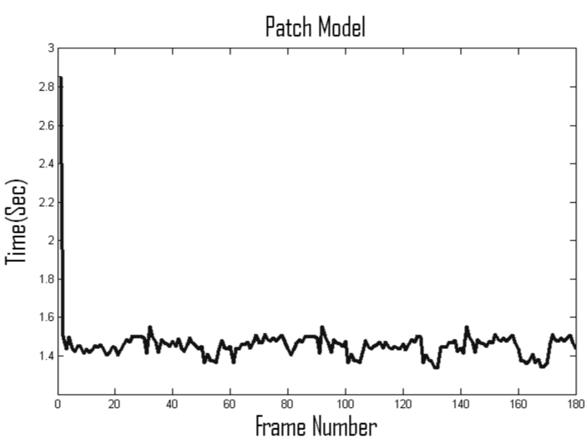
K



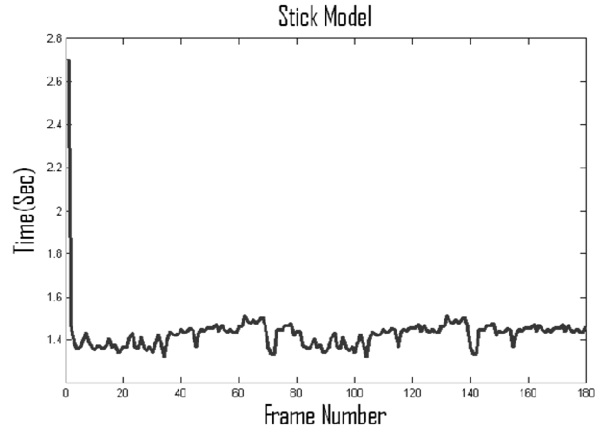
L

Fig.9. Results of 2D modeling based human activities (First column), Stick figure model (Second column), and Patch model (Third column) Rectangle model. A. Standing, B. Right hand rise, C. Left hand rise, D. Both hands rise, E. Right hand up, F. Left hand up, G. Both hands up, H. Right leg rise, I. Left leg rise, J. Right salute, K. Left salute, and L. Crouching

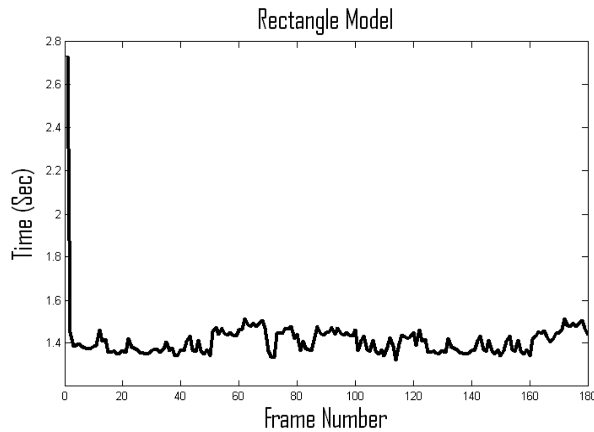
From the response shown in Fig.10, the time taken to compute our algorithm for the three kinds of 2D modeling such as Stick figure modeling, Patch modeling and Rectangle modeling for the steps of 10 frames in a video sequence are observed. For a first frame in a video sequence, it takes average of 2.7 seconds as high as compared to consecutive frames due to the computation of initial processing like frame conversion, background subtraction and preprocessing.



(a)



(b)



(c)

Fig.10. Time response of proposed models
(a) Patch Model (b) Stick Model (c) Rectangle Model

It is noticed that the proposed algorithm has taken approximately 1.5 seconds as an average for 2D models. The efficiency of our models has been found based on the True Positives (TP) and False Positives (FP). True Positives indicate the number of frames in which the output is correct in a video sequence. False Positive is the number of frames for which the output is incorrect. Table.1. shows the efficiency of our proposed algorithm for different videos.

Table.1. Efficiency of proposed algorithm for different Videos

Input	TP	FP	TP/(TP+FP)	Efficiency (%)
Video 1	929	77	0.9234	92.34
Video 2	1119	28	0.9755	97.55
Video 3	1301	81	0.9413	94.13
Video 4	802	62	0.9282	92.82
Video 5	1369	133	0.9114	91.14
Video 6	1137	120	0.9045	90.45
Video 7	1195	91	0.9292	92.92

9. CONCLUSION AND FUTURE WORK

Thinning algorithm is an algorithm which works on indoor video sequences to construct 2D human body model. This algorithm works on straight poses acquired by single static camera without using markers on the human body. Here, twelve activities of 2D models have been discussed based on thinning algorithm. We have considered 13 feature points for upper body modeling as well as for the lower body modeling. In this paper, time expenditure and efficiency of 2D models have also been presented. In the future work, this work can be extended to develop an algorithm for multiple persons tracking and modeling, Vehicle modeling, Occlusion problem, and also for Outdoor surveillance videos with side poses.

REFERENCES

[1] H. S. Chen, H. T. Chen, Y. W. Chen, and S. Y. Lee, 2006, "Human action recognition using star skeleton," in Proc. of the 4th ACM Int. Workshop on Video Surveillance and Sensor Network, pp. 171-178.

[2] H. Su, F. Huang, 2005, "Human Gait Recognition Based on Motion Analysis," in Proc. of MLC, pp. 4464-4468.

[3] M. Ahmad and S. W. Lee, 2006, "Human action recognition using multi-view image sequences features," in Proc. of IEEE Int. Conf. on Automatic Face and Gesture Recognition, pp.523-528.

[4] Tao Zhao, Ram Nevatia and Bo Wu, 2008, "Segmentation and Tracking of Multiple Humans in Crowded Environments", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.30, No.7, pp.1198-1211.

[5] K.Onishi, T.Takiguchi, Y.Ariki, 2008, "3D human posture estimation using the HOG features from monocular image," Proc. of 18th IEEE Int. conference on Pattern Recognition, Tampa, FL, pp.1-4.

[6] Mun Wai Lee, and Isaac Cohen, 2006, "A Model- Based Approach for Estimating Human 3D Poses in Static Images", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, No. 6, pp. 905-916.

[7] Mun Wai Lee, and Ramakant Nevatia, 2009, "Human Pose Tracking in Monocular Sequence Using Multilevel Structured Models", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 31, No. 1, pp.27-38.

[8] D.Y. Chen, S.W.Shih, and H. Y. Mark Liao, 2007, "Human action recognition using 2-D Spatio-temporal templates", Proc. of IEEE Int. Conf. on Multimedia and Expo, pp. 667-670.

[9] A. Gritai, A. Basharat, and M. Shah, 2008, "Geometric constraints on 2D action models for tracking human body", in Proc. of 18th international conference on Pattern recognition, Tampa, FL, pp.1-4.

[10] Feifei Huo, Emile Hendriks, Pavel Paclik and A.H.J. Oomes, 2009, "Markerless human motion capture and pose recognition", Proc. of 10th Workshop on Image Analysis for Multimedia Interactive Services, pp.13-16.

[11] Jianhui Zhao, Ling Li and Kwoh Chee Keong, 2004, "Motion Recovery Based on Feature Extraction from 2D Images", in Proc. of Computer vision and Graphics, ICCVG 2004, Warsaw, Poland, pp. 1075-1081.

[12] Antoni Jaume i Capo et.al, 2006, "Automatic Human Body Modeling for Vision-Based Motion Capture," in Proc. of Short Communication, Plzen, Czech Republic, WSCG'2006.

[13] S. Veni, K. A. Narayanan kutty and M. Kiran Kumar, 2009, "Design of Architecture for Skeletonization on Hexagonal Sampled Image Grid," ICGST-GVIP Journal, Vol.9, No.1, pp.25-34.

[14] L. Huang et al., 2003, "An Improved Parallel Thinning Algorithm," in Proc. of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003), Vol. 2, pp.780-783.

[15] V.Vijaya Kumar., et.al, 2008, "A New Skeletonization Method Based on Connected Component Approach", IJCSNS International Journal of Computer Science and Network Security, Vol.8, No.2, pp.133-137.

[16] S. Schaefer and C. Yuksel, 2007, "Example-Based Skeleton Extraction", in Proc. of Eurographics Symposium on Geometry Processing, pp. 1–10.

[17] K. Srinivasan, K. Porkumaran and G. Sainarayanan, 2009, "Improved background subtraction techniques for security in video applications," in Proc. of 3rd Int. Conference on Anti-counterfeiting, Security, and Identification in Communication, pp.114-117.

[18] William K. Pratt, 2002, "Digital Image Processing", Jhon Wiley & Sons, Inc., Third edition, Singapore.