Essential Human Body Points Tracking Using Kalman Filter

Win Kong, Aini Hussain, *Member IAENG*, and Mohd. Hanif Md Saad

***Abstract*— Human body tracking has a wide spectrum of applications. Examples include video surveillance, motion and activity capture for medical analysis, just to name a few. In this paper, a human skeleton model is proposed for human body tracking in video surveillance system. The skeleton model is based on 8 important body points namely; head, neck, shoulder, pelvis, knees and, ankles. Basically two techniques are used to locate these points. The first technique uses gait study to find the knee and ankle points. The self occlusion of human legs itself was overcome using the gait information. The gait information is then used to overcome self-occlusion of human legs. The second technique involves Kalman filter to track the remaining proposed body points.**

***Index Terms*—Kalman filter, skeleton model, tracking, video surveillance**

I. Introduction

THE area of estimating and tracking a moving human body are inspired by their various applications in automated surveillance, traffic monitoring, sports science, medical analysis and computer animation. In tracking a moving object, only concerned features will be extracted for further investigations. For example, in tracking a human body, important body features are extracted based on the anatomical knowledge. This essential human body point is called as motion data. In biomechanics, the human body information is used to study and observe human movement. Nearly all the biomechanics movement starts from talking to walking, which involves the motion of all body parts. Then, the motion data can be used to identify the posture, pose, action and movement of an actor. Therefore, understanding the motion of body part is essentially needed for person identification.

Human body can be presented in numerous ways based on the body shape. It can be modeled as simple as the contours of the body, points with connected stick, skeletal and shape and even as complicated as 3D volumetric model [8]-[10]. Selecting the model is more important for tracking people and providing useful feature analysis. The performance of tracking is greatly influenced by modeling due to the interconnected joint and the complicated body shape.

In addition, occlusion and noisy background should be considered when doing tracking. Occasionally, the fast movement of human may cause the motion to be predicted incorrectly.

Human body tracking is currently widely used for intelligent video surveillance application. Video surveillance system is one of the most popular researches in computer vision. The rise in public concern about safety nowadays has gained many researchers attention to enhance the current video surveillance system. Intelligent Video Surveillance System (IVSS) is the most recent smart surveillance system which is capable to identify automated human activity and behavior from the video. The general framework applied for an IVSS is indicated in Figure 1.

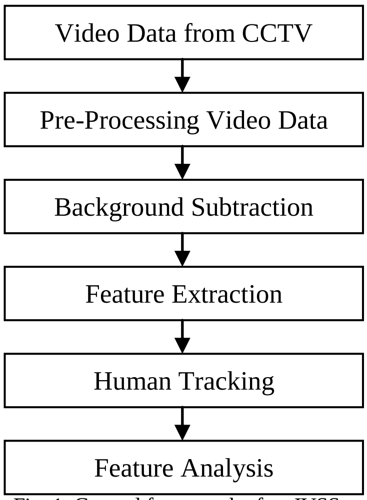


Fig. 1. General framework of an IVSS

Human detection and tracking are challenges tasks as human body is highly articulated and people tend to wear complex clothing texture. Embarking to this challenge, we have successfully developed an 8 points skeleton model. This 8 points skeleton model is an extended work from [1]. The body points are extracted from the silhouette based on anatomical knowledge and connect them by sticks. The skeletonization process allows the tracking of 8 points in each frame using Kalman filter. Kalman filter tracking technique is briefly introduced in this paper. We are able to implement our system for human body points tracking, given a video which consist of many frames with segmented binary images. The extracted human points extracted are later used for gait analysis.

Manuscript received April 26, 2013; revised May 15, 2013. This work was supported in part by the Universiti Kebangsaan Malaysia under Research University Grant DIP-2012-03 and MOHE Grant LRGS/TD/2011/UKM/ICT/04/02.

Win Kong is a PhD student at the Department of Electrical, Electronics and System Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, MALAYSIA (e-mail: kwin@eng.ukm.my).

Aini Hussain is with the Department of Electrical, Electronic and Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, MALAYSIA (tel:+603-89118393 fax:+603-89118359 e-mail: aini@eng.ukm.my).

Mohamad Hanif Md Saad is a lecturer and researcher at the Department of Mechanical & Materials Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, MALAYSIA (e-mail: hanif@eng.ukm.my).

II. Related Work

Yu Huang et al [2] proposed a technique for tracking 2D articulated cardboard model from a monocular video sequence. They used Expectation-Maximization (EM) to solve body model parameters. The kinematics of the model is found in linear form. Nooritawati Md Tahir et al [11] used shock graph technique to identify the head and both feet of human body. Yilmaz et al [3] used silhouette contour to represent human body. They developed a new modeling approach based on the object shape and the features. The object shape helps in recovering the missing part during occlusion. Shape energy technique derived using bayesian is applied for contour tracking. Camaniciu et al [4] used shape-based model to represent the moving object that to be tracked. The kernel-based tracking approach is done by using the basic similarity function. Jang et al [5] proposed a modified version of Kalman filter which is called as Structure Kalman filter. The structure Kalman filter is then applied on the silhouette which has been partitioned into 3 parts.

III. Method Overview

A. Image Pre-processing

The existence of artifact hugely affects the point extraction performance. To deal with artifact, morphological operation is applied on the image. In this paper, the image is dilated twice and eroded once to reduce the unwanted and jagged edge. The edge of the silhouette is then extracted using canny edge detector. Once the edge detection process is done, its bounding box is created. From the bounding, we could actually measure the height and width of the body. The enhanced human silhouette is further processed to obtain the 8 important body points.

Fig 2(a). The original silhouette image of human body.

Fig 2(b). The dilated and eroded silhouette image.

Fig 2(c). The edge extracted human body image.

B. Points Extraction

We use the anthropometric knowledge of human individual. Firstly, we define the vertical position of each segment based on the anatomical knowledge. The dimensions of different body segment are proposed in [6]. We constructed an 8 points body model. The 8 points construct the point of interest to be detected and tracked from video stream images. The 8 points are defined as follow:

1) The Head (P_H | (X_H,Y_H))
2) The Neck (P_N| (X_N,Y_N))
3) The Shoulder (P_S| (X_S,Y_S))
4) The Pelvis (P_P| (X_P,Y_P))
5) Left Knee and Right Knee (P_{LK} & P_{RK}|(X_{iK},Y_{iK})
6) Left Ankle and Right Ankle (P_{LA} & P_{RA}| (X_{iA},Y_{iA}))

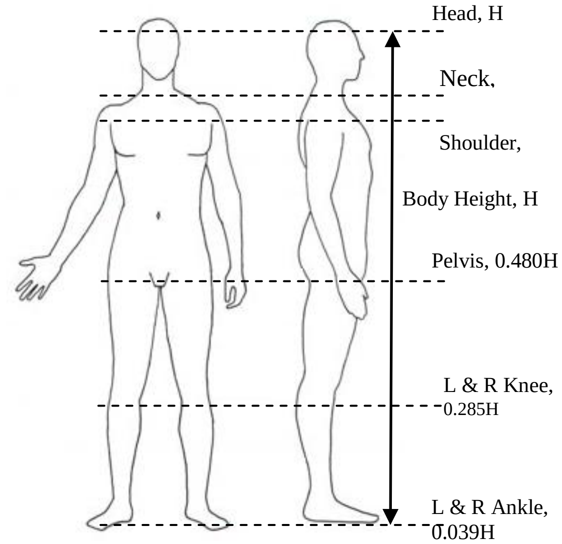


Fig 3. Body segment proportion based on [6]

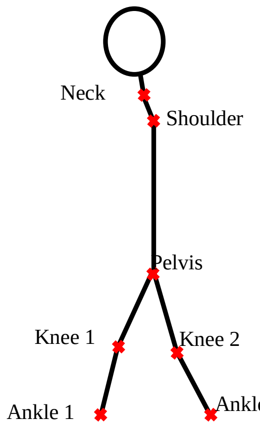


Fig 4. Skeleton model and important body points

We have initialized the height of body according to the respective created bounding box. To model the head, we constructed a circle to represent it. The radius, R_i is assumed to be half of vertical position from head to neck and the centroid, (X_c , Y_c) is found as below:

$$R_i = \frac{1}{2}(Y_N - Y_H)\,;\ \ i = 1,2,3\ldots, number\ of\ frame \qquad (1)$$

$$S_i = \{(X_i, Y_i) | I(X_i, Y_i) = 1\};$$
$$I = mage\ after\ dilation\ and\ erosion \qquad (2)$$

$$(X_c, Y_c) = mean\ (S_i) \qquad (3)$$

The neck, shoulder and pelvis points are found by calculating the mean of the interval starting from the first X_f to the last border point X_l.

$$X_{iw} = mean(BP)\ ;\ BP : [X_f, X_l];$$
$$w = Neck, Shoulder, Pelvis \qquad (4)$$

The process of extracting tracking the two lower limbs is a challenging task due to possible overlapping of the two legs. When the two legs are overlapping, we must be able to differentiate and identify them. To overcome this problem, we

ISBN: 978-988-19252-3-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

first detect the legs whether both are overlapping or separate. The technique and equation used to detect the crossover of the two legs are as follows:

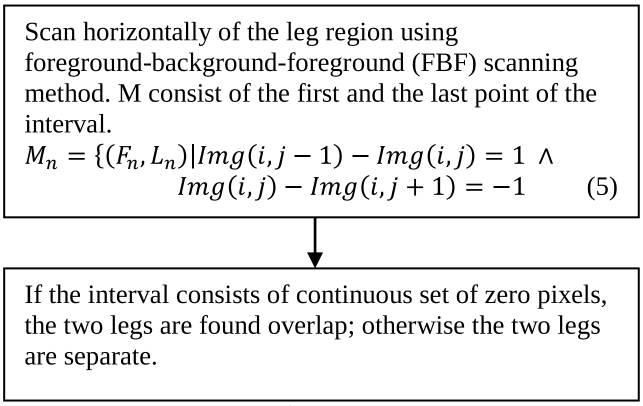


Fig 5. Algorithm for leg crossover detection

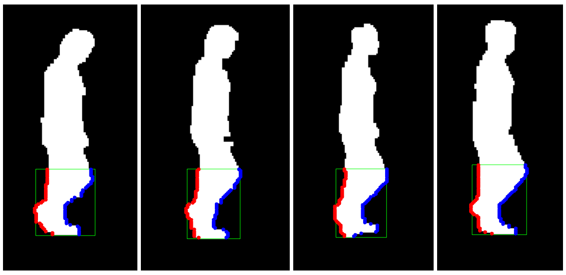


Fig 6. Detected image when two legs are overlapped

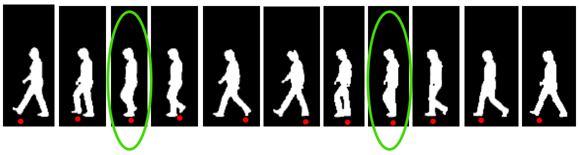


Fig 7. Silhouette of a complete gait cycle

Figure 6 indicated the detected image sequences when two legs are overlapped. For the purpose of identifying two legs individually, gait knowledge is studied. A complete gait cycle is the period of time from one movement to following movement of the same pose with the same foot. It means we need to have twice detected overlap per to complete a gait cycle. Referring to Figure 7, two legs interchange when two legs are exactly overlapped each other. The red dot in the image sequence change alternatively. The green circle indicates both legs are overlapped. The next step is to extract the knee and ankle points. It can be observed from the collected set of image sequences in Figure 7, extracting knee and ankle points is a difficult and challenging work as the lower limb has significant changes from frame to frame.

To obtain knee and ankle points, we divides the lower limb into two parts:

1) upper part (Vertical position from Y_P to Y_K)
2) lower part (Vertical position from Y_K to Y_A)

The upper part is for knee point extraction purpose whereas lower part is for ankle point extraction purpose. The border points of the upper part are extracted. We obtained two set of points which is leftmost border points, $LK_i(X_{Li}, Y_{Li})$ and rightmost border points, $RK_i(X_{Ri}, Y_{Ri})$. Furthermore, the mean of each set of x and y values, $\overline{X_L}$, $\overline{X_R}$, $\overline{Y_L}$, $\overline{Y_R}$ are computed. Once all this data is available, we calculate the joint θ_{LKi} and θ_{RKi}angle using the linear regression. Finally, we calculate the joint coordinate using trigonometry function.

$$\theta_K = \tan^{-1}\left(\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}\right) \quad (6)$$

$$X_K = X_P - L_K * (\sin(\delta - \theta_K)) \quad (7)$$

$$Y_K = Y_P + L_K * (\cos(\delta + \theta_K)) \quad (8)$$

C. Kalman Filter Tracking

Kalman filter is used to estimate and track the body points. Kalman filter estimates the coordinate of the body points in every frame of the image sequence. The input parameters of the Kalman filter are centroid and radius of head and coordinate of neck, shoulder, pelvis, left and right knee, left and right ankle which can change significantly during the movement. Every point is tracked individually. The calculated parameters discussed above are state vector and measurement vector. As an example, for tracking head point, we have (P_H | (X_H,Y_H)). We measured X_H and Y_H from the image. We defined the measurement vector, Z as the measured value (the points to be tracked). The vector is shown by Equation(9).

$$Z_i = (X_H, Y_H) \quad (9)$$

We randomly initialized the estimate state and estimate covariance which are respectively X_{est} and P_{est} as a set of zero. The value of $X_{est_{(i-1)}}$ and $P_{est_{(i-1)}}$ will passed to the next frame. And X_{est} is used as state vector in the next frame. To start tracking, we defined the state model matrix, A, observation model matrix, H, and noise, Q and R as below:

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (10)$$

$$H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (11)$$

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (12)$$

$$R = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} \quad (13)$$

In general, the prediction stage use the previous state estimated from the previous frame to define the state of the current frame. The equations used are shown as below:

$$Z_{prd_{(i)}} = A \times Z_{est_{(i-1)}} \quad (14)$$

$$P_{prd_{(i)}} = A \times P_{est_{(i-1)}} \times A^T + Q \quad (15)$$

To improve the estimated state, we update it by using the Kalman gain calculated as following:

$$S = H \times P_{prd_{(i)}}{}^T \times H^T + R \quad (16)$$

$$B = H \times P_{prd_{(i)}}{}^T \quad (17)$$

$$\text{Kalman Gain, } G = B \times S^{-1} \quad (18)$$

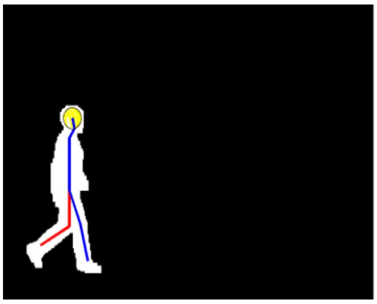
$$Z_{est_{(i)}} = Z_{prd_{(i)}} + G \times \left(Z - H \times Z_{prd_{(i)}}\right) \quad (19)$$

$$P_{est_{(i)}} = P_{prd_{(i)}} - G \times H \times P_{prd_{(i)}}) \quad (20)$$

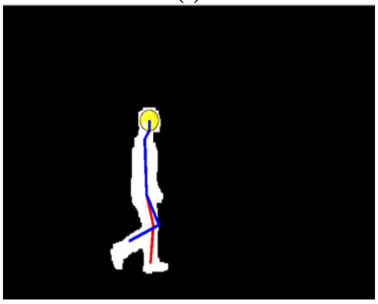
IV. Results And Discussion

The algorithm discussed above is realized in MATLAB. There are two main results which is vital to the success of the algorithm. First, the proposed approach is tested on a few sets of silhouette image from CASIA [7] as shown in Figure 8(a-e).

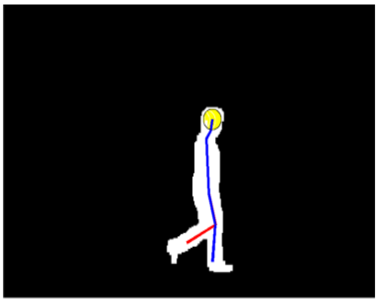
ISBN: 978-988-19252-3-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)



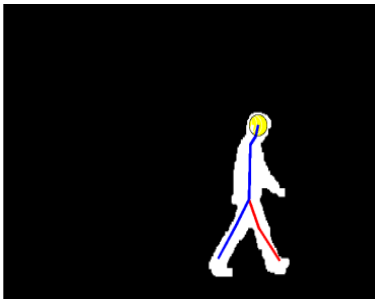
(a)



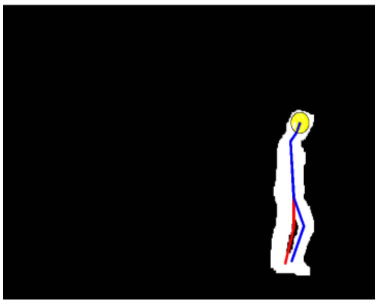
(b)



(c)



(d)

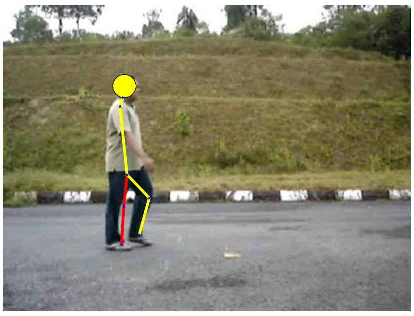


(e)

Fig 8. Sample results using silhouette image from CASIA

Subsequently, we input the video produced by InViSS group and the resulted output are given in Fig.9(a-d).

(a)



(b)

(c)

(d)

Fig 8. Sample results using image from InVISS

V. Conclusion And Future Work

The developed 8 points skeleton model is observed to be stable for human body tracking application across frame in a given video. The approach proposed in this paper is simple and provides foundation for further improvement in the method. In the future, we will be focusing in developing better view-invariant using advanced method.

References

[1] K. Win, A. Hussain, M. H. M. Saad, and N. M. Tahir, "Hand detection from silhouette for video surveillance application," in *Signal Processing and its Applications (CSPA)*, 2012 IEEE 8th International Colloquium on, 2012, pp. 514-518

[2] H. Yu and T. S. Huang, "Model-based human body tracking," in *Pattern Recgonition, 2002*. Proceedings. 16th International Conference on, 2002, vol. 1, pp. 552-555

ISBN: 978-988-19252-3-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

[3] Y. Alper, L. Xin, S. Mubarak, “ Contour based object tracking with occlusion handling in video acquired using mobile cameras,” *IEEE Transaction on Pattern Analysis and Machine Intelligence*. vol. 26, No. 11, pp.1101-1113, Nov. 2004.

[4] C. Dorin, R. Visvanathan, M. Peter, “Kernel-based object tracking. *IEEE Transaction on Pattern Analysis and Machine Intelligence*,” vol. 25, No. 5, pp. 564-577, May 2003.

[5] J. Dae-Sik, J. Seok-Woo, C. Hyung-II, “ 2D human body tracking with structural Kalman filter,” *Pattern Recognition*, vol.35, pp. 2041-2049, 2002.

[6] W. T. Dempster and G. R. L. Gaughran. Properties of Body Segments Based on Size and Weight. *American Journal of Anatomy*, 120, pp.33-54, 1967.

[7] Chinese Academy of Sciences (CASIA) gait database, Institute of Automation, Center of Biometrics and Security Research.

[8] K. Rohr, “ Towards model-based recognition of human movements in image sequences,” *CVGIP: Image Understanding* , 59(1), 94–115, 1994.

[9] I. A. Kakadiaris and D. Metaxas, “ Model based estimation of 3D human motion with occlusion based on active multi-viewpoint selection,” in *Proc. IEEE CS Conf. on CVPR*, San Francisco, CA, 1996, pp. 81–87

[10] D. M. Gavrila and L. S. Davis, “ Towards 3-D model-based tracking and recognition of human movement: a multi-view approach,” in *Proceeding Intl. Workshop on Automatic Face and Gesture Recognition*, Zurich, Switzerland, June 1995.

[11] Nooritawati Md Tahir, Aini Hussain, Salina Abdul Samad and Hafizah Husain, “ Shock graph representation and modelling of posture,” *ERTI Journal*, 29(4),2007.

ISBN: 978-988-19252-3-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)