Real Time Face Detection System Based Edge Restoration and Nested K-Means at Frontal View

Hyun Jea Joo, Bong Won Jang, Md. Rezaul Bashar, and Phill Kyu Rhee

Dept. of Computer Science & Engineering
Inha University, Yong-Hyun Dong
Incheon, South Korea
{hj0235, bictory7, bashar}@im.inha.ac.kr, pkrhee@inha.ac.kr

Abstract. Bayesian technique is a popular tool for object detection due to its high efficiency. As it compares pixel by pixel, it takes a lot of execution time. This paper addresses a novel framework for head detection with minimum time and high accuracy. To detect head from motion pictures, motion segmentation algorithm is employed. The novelty of this paper carried out with the following steps: frame differencing, preprocessing, detecting edge lines and restoration, finding the head area and cutting the head candidate. Moreover, nested K-means algorithm is adopted to find head location and statistical modeling is employed to determine face or non-face class, while Bayesian Discriminating Features (BDF) method is employed to verify the faces. Finally, the proposed system is carried out with a lot of experiments and a recognizable success is notified.

1 Introduction

Many visual surveillance systems have adopted camera capture system to detect a lot of typical objects or analyze their motion and patterns. As huge size images, we need only objects active candidate regions. Because, the big sized images take long time to calculate and detect. Therefore, it is the key to find segmentation domains for using real time applications. Because of them, many systems adopt head detection and tracking methods [1][2][3][4][9][10][11]. And they adopt color – based methods to find head regions. [1][3][4][9] However many people have different face color and illumination in various environment. Therefore many systems do not define all face color values. They have solved the restricted condition. And they select the shape modeling like as ellipse. [12][14] But, these applications have disadvantage to solve noise effect. When the original image contains noise, the ellipse model turns to huge space region and image distortion. For overcoming this defect, some methods have used motion segmentation algorithm [7]. Motion segmentation in visual surveillance system falls into three divisions: environmental map, frame differencing and optical flow.

The algorithms using environmental map are very universal. First, they make statistical background. Then, they perform subtraction of current image frame from environment map [9][10]. This method is simple and gives good result in restricted

L. Wang et al. (Eds.): FSKD 2006, LNAI 4223, pp. 1192 – 1200, 2006.

environment. But it is very weak when the light circumstance is changed like as the direct rays of the sun. So, the first consideration in this case is to make accurate environment, which seems difficult to achieve.

Frame differencing approach adopts difference between two or three frames in image sequence. However, an improved performance is achieved by three frame differencing. Frame differencing is very robust method to continuing environmental changes [11]. Because of the differencing result, it is fragile to restore the original edges and to extract the active region. Therefore, the advantages of them are based on making activate candidate regions.

Optical flow selects flow vectors of moving objects in an image sequence. The results apply for gait and activity analysis [12]. But they have defects in complex and sensitive noise effects and do not perform in real time video.

Because of unrestricted surrounding changes in real time, frame-differencing method is used to improve accuracy and to minimize multi-resolution search. So the aim of this paper is to achieve high head detection rate employing minimum multi resolution searches and to reduce overall computational time. Edge detection and restoration challenge is prevailed over by using image evaluation, preprocessing and edge restoration. Then, active head regions are trimmed resulting in reduced search spaces in shorter time for using real time system. Head detection is employed on a predefined threshold value between the people and camera using interpolation method. Moreover, only head detection has been taken into account for this research. So we perform head detection using two clusters K-means algorithm followed by head and body histogram analysis. To detect heads of people, we adopt 13 frames per second frame rate for single person in indoor scene. The camera is fixed 1.6 meter above the ground. The people can move over 25 square meters area in front of camera with their face pointing towards it. Our approaches are divided into several steps. Section 2 presents system architecture. Section 3 deals with image evaluation, image preprocessing, edge detection and restoration. Section 4 presents head detection and face detection using Bayesian method. Section 5 brings up experiments. Finally, sections 6 describe the conclusion and future extensions.

2 Real Time Head Detection System Architecture

Our purposed system architecture is shown in Fig. 1. Input images are of 320 × 240 sizes. Then the difference between the current image (N) frame and previous frame (N-1) is calculated and evaluated. The difference is then compared with certain threshold. The system proceeds towards edge detection only when the difference is significant enough to find the body segment. But when the result is not changed i.e. difference is less than the threshold value (person is not moving), current frame (N) is compared with the second previous (N-2) and difference is evaluated. Therefore the maximum difference is measured between two frames by comparing three image frames. If again the difference is not significant then previously segmented head part is used for face detection. Otherwise, other steps proceeded out which results in new head segmentation. The steps include edge detection or restoration. Here accurate

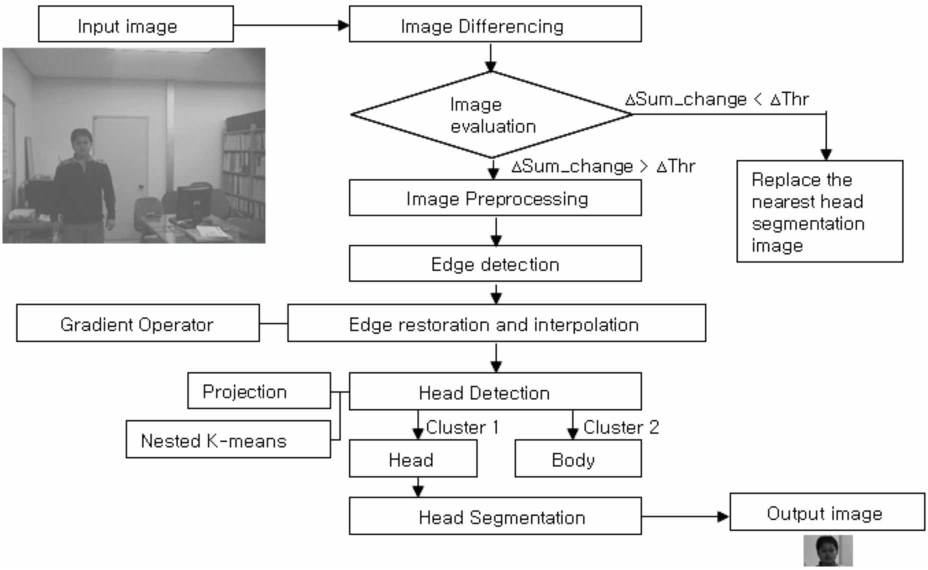

Fig. 1. Architecture of the proposed methodology

edge line across the body is found by performing gradient operators and interpolation, detail of which is discussed in section 3. Next, the projection of horizontal distance between the edges is calculated for each row. Local minima and maxima point is calculated on the projection vector to find the head and body location respectively. To overcome the error due to noise, the projection vector is clustered around the two regions (one for head and one for body) using K-means algorithm. After the head detection region is extracted from the image, Bayesian Discriminating feature method is used to search the target's face in the region to verify.

3 Movement Detection, Edge Detection and Restoration

3.1 Movement Detection

Our proposed method adopts image evaluation for movement detection, i.e. finding people that are moving or static. When the people get far from camera, human's head is very small, for example 25 pixels × 25 pixels, this is why, people can move 25 square meters within target tracking region. However, head size in end point varies on hairstyle, bald headed, man and woman, and other factors. Since we use Bayesian discrimination method for face detection and it uses minimum 16×16 pixel window for statistical analysis. Threshold is calculated assuming that the person should at least move his head for the noticeable change in the successive image frames. The borderline of 16×16 window contains 16×4 pixels and the approach finds the accurate human edges. Therefore the threshold is found to be $\Delta Thr_{total} = 16 \times 3$ (i.e. discarding the lower edge of window).

$$\text{Comp(x,y)}=\begin{cases}255 & Set(x,y) < \Delta Thr_{in} \\ 0 & otherwise\end{cases} \tag{1}$$

where

$$\text{Set(x,y)}=\begin{cases}F_N(x,y)-F_{N-1}(x,y) & \Delta Sum_{change} \geq \Delta Thr_{total} \\ F_N(x,y)-F_{N-2}(x,y) & otherwise\end{cases}$$

$$\Delta \text{Sum}_{\text{change}}=\frac{1}{255 * Total}\sum_{n=1}^{total} Comp(x,y)$$

Comp(x, y) is the threshold operator and it sets only two extreme gray color values 255 or 0. FN(x, y) is the image sequence indicating the Nth intensity value of (x, y). These values are resolved into white and black for finding edge line. However when $\Delta\text{Sum}_{\text{change}}$ value is less than $\Delta\text{Thr}_{\text{total}}$ current frame is regarded as static. In this case the previous nearest head segmentation image is used as active region.

3.2 Edge Restoration

Edge is an important issue to detect a particular shape of an object. Human is symmetric shaped especially for frontal view. As frame differencing creates a lot of noise and there is no predefined way to detect noise. To overcome this, edge restoration is employed.

After preprocessing and evaluation, the image edge is searched for the left and right row point to find the first edge pixel .The column points for left and right edge is stored as $\text{Left}_{\text{first}}$(x,y) and $\text{Right}_{\text{first}}$(x,y) respectively. Some successive points for both columns are considered determining the presence or absence of noise. Noise is detected in two ways: distance and median. For distance method, if there is abrupt change between successive points, noise is detected; otherwise, noise free as shown in equation (2). For median method, the center point between left and right position is determined. If this center point much differs from threshold, we consider this as noise as shown in equation (3). If noise is detected noise-removing technique is applied. Fig2. Shows edge restoration.

$$\text{E}_{\text{left}}(\text{x}_{\text{i}},\text{y})=\begin{cases}(Left_{first}(x_i,y)=Left_{first}(x_{i-1},y)) & Dev(x,y) \geq Thr_{out} \\ (Left_{first}(x_i,y) & otherwise\end{cases} \tag{2}$$

where

Dev_{left} (x,y)= $(Left_{first}(x_i,y)-Left_{first}(x_{i-1},y)$

$\text{E}_{\text{right}}(\text{x}_{\text{i}},\text{y})$, $\text{Dev}_{\text{right}}$ (x,y) are the same ways.

$$\text{Center}_{\text{n-1}}\text{(x,y)}=\frac{1}{n-1}\sum_{i=1}^{n-1}\frac{1}{2}\left(Left_{first}(x_i,y_i)+Right_{first}(x_i,y_i)\right) \tag{3}$$

$\text{J}_{\text{n}}(\text{x,y}) \geq \text{Thr}_{\text{out}}$

$$\text{J}_{\text{n}}\text{ (x,y)}=\frac{1}{2}\left(Left_{first}(x_n,y_n)+Right_{first}(x_n,y_n)\right)-\text{Center}_{\text{n-1}}\text{(x,y)}$$

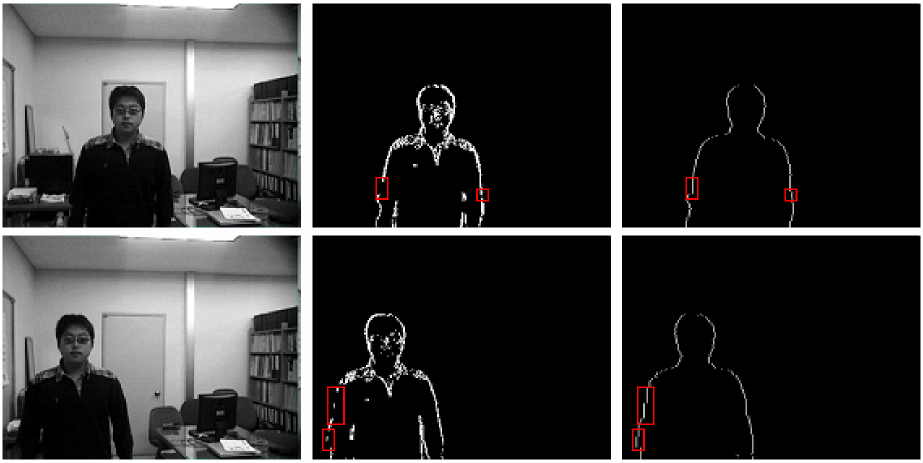

Fig. 2. Original images (left), Differencing images and preprocessed images (center), Edge restoration images (right)

4 Head Region Segmentation Using Nested K-Means Algorithm

To reduce the search space for face detection, first active region of head is detected by the method of head segmentation using vertical projection of edge contours and K-means algorithm.

Here edge contours of the person are extracted by the method described in section 3.2. Then Body width set projection of the edge contours is calculated. Then local minima and maxima are calculated on the projection vector in order to find the expected location of head and body respectively.

$$\mathrm{RP}(\mathrm{x}_{\mathrm{i}}) = \mathrm{Dev}_{\mathrm{right}}(\mathrm{y}) - \mathrm{Dev}_{\mathrm{left}}(\mathrm{y}) \tag{4}$$

The first and second derivative of projection vector is calculated to find local minima and maxima. The definitions about second derivative [5] states that; "If the first derivative *RP*'(xi) is positive (+), then the function RP(xi) is increasing (↑) and RP'(xi) is negative (-), then the function RP(xi) is decreasing (↓). Also, If the second derivative RP'' (xi) is positive (+), then the function RP (xi) is concave up (∪) and RP'' (xi) is negative (-), then the function RP (xi) is concave down (∩)." It is assumed that y=RP(x) is a twice-differentiable function with RP'' (c) =0.

$$\begin{aligned} &f''(x) : y'', \frac{d^2y}{dx^2}, \frac{d^2}{dx^2} f(x) \\ &\mathrm{RP'}(\mathrm{x_i}) = \frac{f(x_i + h) - f(x_i)}{h} \\ &\mathrm{RP''}(\mathrm{x_i}) = \frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) = \lim_{h\to 0} \frac{f'(x_i + h) - f(x_i)}{h} \end{aligned} \tag{5}$$

Therefore if RP'' (d) <0, then RP (x_i) has a relative maximum value at x=d. It means that body center ($Max_d(y)$) is in body cluster. In same ways, if RP'' (c) >0, then RP (x_i) has a relative minimum value at x=c. So, it is near to head center (Minc(y)) and in head cluster, too. Maxd(y), Minc(y) points, that is the center of clusters is used to calculate the similarity:

$$D_b = \|Max_d(y) - RP(x_i)\|^2$$
$$D_h = \|Min_c(y) - RP(x_i)\|^2 \tag{6}$$

$$G(x,y) = \begin{cases} 0 & D_b \geq D_h \;\; Head \quad part \\ 1 & otherwise\, Body \quad part \end{cases} \tag{7}$$

$$\mu_{head} = \frac{\sum x}{count(x)} \qquad forallx : G(x,y) = 0 \tag{8}$$

$$\mu_{body} = \frac{\sum x}{count(x)} \qquad forallx : G(x,y) = 1$$

$$d_b = \|x - \mu_{body}\|^2$$

$$d_h = \|x - \mu_{head}\|^2$$

$$g(x,y) = \begin{cases} 0 & d_b \leq d_h \;\; Head \quad part \\ 1 & otherwise\, Body \quad part \end{cases}$$

Nested k-means algorithm is implemented for the purpose of head and body segmentation. Clustering is done in two direction of projection vector. At first local minima and maxima is found in the projection vector RP (x). The operator is defined in equation (4). Then the vector is clustered to search head and body location on the width value using initial centroid as calculated minima and maxima in (5). The distance between the head cluster's center and body cluster's center point is calculated for every entry in projection vector. If distance to the head cluster is greater than that of body cluster, then the point is shifted to body cluster. The first stage of clustering separates head and body parts but error due to noise still exists. That is there are some of the body parts, which are still mis-clustered as head part. To reduce this noise, second level of clustering is needed. Second level of clustering is done on x direction of RP(x), i.e. on the index value of the projection vector taking centroid result of first level clustering as initial centroid. In the same way, the distance between the head cluster's center and body cluster center's point is calculated for every entry in projection vector. If distance to the head cluster is greater than that of body cluster, then the point is shifted to body cluster. So, in second stage of clustering any outline point (x, y) (related to body), which mis-clustered to head part, is corrected. It is found that this method increases the detection rate, because second level of clustering corrects most of the noise occurred in first level of clustering. Fig 4. shows the result of finding head and body regions. The left image is the neck line and body line with the first k-means method using object widths. However, this image contains the noise effect. Because of the nearest Euclidean distance, the first k-means method contains this area to head parts. But it must be a body part. According to my algorithm, the second k-means method eliminates the rectangle using Euclidean distance. The right image is the grouping result about head and body parts using second k-means.

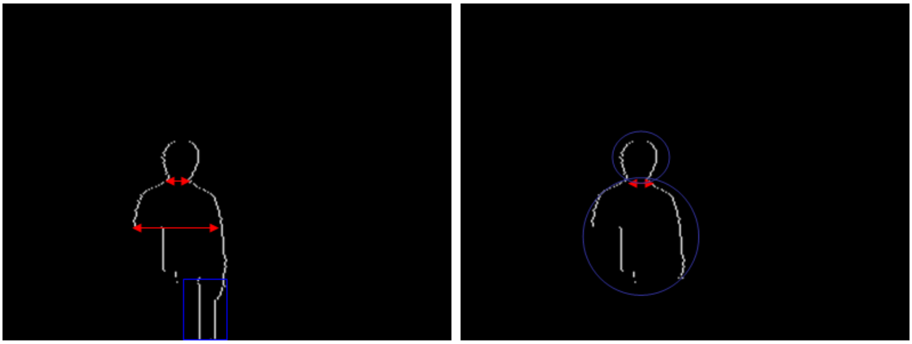

Fig. 3. First k-means method result (left) second k-means method result (right)

5 Experiments

Our proposed method is tested using two series of experiments; first face detection without Nested K-means, second is face detection with head segmentation with or without K means. A cheap LG USB camera is used as a video capturing device. The sequence frames of 320 x 240 pixels were acquired at 13 frames per second (fps) in 2.0GHx Pentium 4 PC computer and stored at the temporary folder in hard disk. Database is separated into 3 groups A, B and C. Group A contains normal people without any spectacles glasses and disguise form. Group B group includes people wearing glass. The last group C contains the people having beard and wearing glass in changing illumination condition. One group consists of 5 people dataset, and each

Table 1. Experiments about face detection rate about head region segmentation methods

	Experiment division	Database A group	Database B group	Database C group	Average total Database
BDF	**Face detection rate**	**92.46%**	**93.48%**	**83.31%**	**89.7%**
Non-including Nested k-means head detection +BDF		**73.29%**	**68.79%**	**65.87%**	**69.32%**
Nested k-means head detection +BDF		**97.29%**	**97.34%**	**97.21%**	**97.28%**
Non-including Nested k-means head detection +BDF	**Accurate Head Location rate**	**93.33%**	**92.14%**	**91.19%**	**92.22%**
Head seg +BDF (proposed method)		**98.07%**	**97.73%**	**99.23%**	**98.34%**

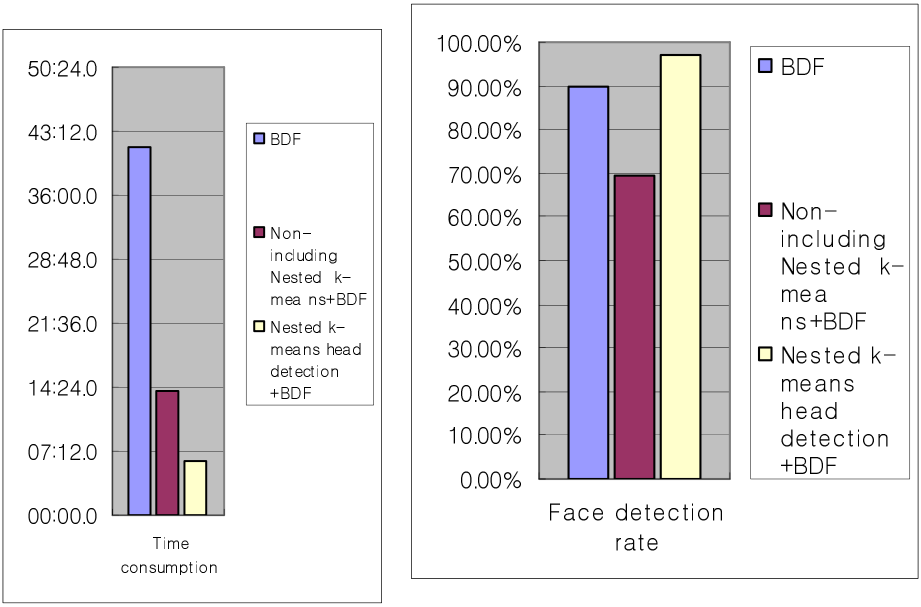

Fig. 4. Time consumption (left), Average face detection rate (right)

person has twenty-second movie clips at 13 fps. So each group has 1300 image sequences. Table 1 shows the experimental results. After the head segmentation, the system finds Bayesian classifier to discriminate face and non-face part of image. FERET database is used to learn the face Bayesian face model where as no-face Bayesian model is generated from the CMU database set.

Table 1 shows the face detection result for different scenario. Here it can see that face detection without segmentation gives high accuracy rate but it is time consuming. The third row of table 1 shows the result of our proposed method, which uses head segmentation and k-means for noise reduction. The result of our proposed method is found to be superior in terms of accuracy 97% and very fast in terms of speed. The experiment was conducted, but that time not using k-means for noise reduction. The recognition rate in this case is found to be very poor. Hence the credit of higher recognition rate goes to the error reduction technique employed by k-means algorithm. And Table 1 shows the result of experiment that shows the accuracy of the head location. Same database and experiments were conducted. The table suggest that our method is almost perfect to find the head location with the accuracy higher than 97%. But, when we do not contain the nested K-means algorithm, the result is very low because our experiment in real-time visual surveillance system contains many noise effects like as changing illumination.

Fig. 4 shows the variation of time consumption and accuracy rates across various methods. It can be seen that the proposed method (Head segmentation +BDF [8]) method based face detection) have high detection rate with less response time.

6 Conclusion

We have proposed real time head region detection based on difference of image frames from active camera. The main component of the system is the use of vertical projection of edge contours to find head location and Nested K-means algorithm for error minimization during differencing and threshold. Overall the system shows low computational cost and high detection rate, which suits for the real time system. Accurate head detection rate of about 97% is achieved while minimizing the computational cost and time.

The future enhancement to the system can be to overcome the occlusion and to find automatic threshold for frames evaluation using multiple cameras.

References

[1] Maolin Chen and Seokcheol Kee, “Head tracking with shape modeling and detection”
[2] G. L. Foresti, “A Real-Time System for Video Surveillance of Unattended Outdoor Environments”, IEEE Transactions On Circuits And Systems For Video Technology, Vol. 8, No. 6, October 1998
[3] Tao Yang, Quan Pan, Jing Li, Yongmei Cheng, Chunhui Zhao, “Real-Time Head Tracking System With an Active Camera*, Proceedings of the 5" World Congress on Intelligent Control and Automation, June 15-19, 2004, Hangzhou, PR China
[4] Weimin Huang, Ruijing Luo, Haihong, Beng Hai Lee and Meneka Rajapakse “ Real Time Head Tracking and Face and Eyes Detection, Proceedings of IEEE TENCON’02
[5] Jing-Un Won; Yun-Su Chung; In-Soo Kim; Jae-Gark Choi; Kil-Houm Park, "Correlation based video-dissolve detection," Research and Education, 2003. Proceedings. ITRE2003
[6] How-Lung Eng, Junxian Wang, Alvin H. Kam, Wei-Yun Yau, "A Bayesian Framework for Robust Human Detection and Occlusion Handling using Human Shape Model," *icpr*, pp. 257-260, 17th International Conference on Pattern Recognition (ICPR'04) - Volume 2, 2004
[7] Weiming Hu, Tieniu Tan, Fellow, IEEE, Liang Wang, and Steve Maybank “A Survey on Visual Surveillance of Object Motion and Behaviors”, Ieee Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews, Vol. 34, No. 3, August 2004
[8] Chengjun Liu, Member, IEEE,” A Bayesian Discriminating Features Method for Face Detection”, IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 25, No. 6, June 2003
[9] I. Haritaoglu, D. Harwood, and L. S. Davis, “W : Real-time surveillance of people and their activities,” IEEE Trans. Pattern Anal. Machine Intell., vol. 22, pp. 809–830, Aug. 2000.
[10] S. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, “Tracking groups of people,” Comput. Vis. Image Understanding, vol. 80, no. 1, pp. 42–56, 2000.
[11] A. J. Lipton, H. Fujiyoshi, and R. S. Patil, “Moving target classification and tracking from real-time video,” in Proc. IEEE Workshop Applications of Computer Vision, 1998, pp. 8–14.
[12] D. Meyer, J. Denzler, and H. Niemann, “Model based extraction of articulated objects in image sequences for gait analysis,” in Proc. IEEE Int. Conf. Image Processing, 1998, pp. 78–81.
[13] S. Mulassiotis and M. G. Strintzis, “Real-Time Head Tracking And 3d Pose Estimation From Range Data”, 2003 IEEE
[14] Harsh Nanda and Kikuo Fujimura, “A Robust Elliptical Head Tracker”, Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR’04)