Significance Of Effective Local Gradient Distribution Technique With Eft On Multi-View And Cloth Invariant Gait Recognition

1tejas K. Rayangoudar ,2dr. H. C. Nagraj

1.Asst. Professor, Dept. Of ECE, K.L.E. Institute Of Technology,Hubballi, Karnataka, India
2.Principal, NITTE Meenakshi Institute Of Technology, Bangalore, Karnataka, India,

Abstract

The gait recognition is gaining a growing interest from computer vision researchers as a trusted second generation biometrics. Change in walking speed of the subject, variation in cloths and variation in view angle of the gait poses serious challenges to gait biometrics. The current research has not yet addressed these challenges in gait recognition. This research work establishes the significance of spatial and frequency domain feature extraction on cloth and view invariant gait recognition. It considers the feature extraction algorithms CHOG and EFT and their individual output response to the SVM classifier is analyzed for cloth and view invariant gait recognition. The circular histogram of oriented gradient captures the decimated angular local energy gradient feature, the Elliptical Fourier Transform is applied on CHOG feature gives the accurate geometry of the gait. The fusion of this two-feature extraction algorithm implicitly captures geometrical structure and dynamically changing characteristics of gait under different view angle and with different cloths. The CHOG feature effectively contributes for the gait classification with different view angle but showcases weak features with changing cloths resulting in a poor classification rate. The combination of CHOG features with EFT gives an excellent geometric structural details of inter variation gait pattern and has given the classification rate via SVM of 97% and above for 10 different view angle and 11 different cloths. The combination of CHOG and EFT gives very encouraging and stable results versus the previously proposed spatial domain techniques like regression model, entropy feature, deterministic learning, GEI with static and dynamic approach.

Keywords – Circular Histogram of Oriented Gradient (CHOG), Gait cycle, view invariant, Elliptical Fourier Transform (EFT)

1. INTRODUCTION

The unique pattern in which the human walks is called the Gait. Gait biometrics traces the walking pattern over one complete gait cycle to identify the person or the subject. It involves the leg movement of the subject in two phases swing phase and stance phase respectively. The literature depicts two viewpoints for the human gait identification i.e. model based analysis and motion based analysis [1, 2, and 3]. In model based method, the gait walking pattern is found through the pre defined articulation points during gait cycle thereby it requires implementation of strong computational algorithms. In motion based method the recognition is done based on analysis of gait image frame sequences or based on the analysis of gait templates.

The current research is concentrated on the underlying challenges in gait like subject appearing in different cloths[4,5], loading effect on the subject, walking speed effects, varying illumination conditions, variation in view angle[6, 7]. In real world scenario, when the subject walks in various directions to the camera, the gait pattern changes non linear leading to significant changes in gait pattern. Also when the subject appears on camera with different cloths, there is a possibility of gait mismatch. To address this multi view gait invariance and the multi cloth invariance, we apply evaluate and discuss the significance of salient feature extraction technique namely Circular Histogram of Oriented Grading (CHOG) and Elliptical Fourier transforms with SVM classifier.

2. PREVIOUS WORK

WorapanKusakunniran, et. al., has described gait features for different view angles that are normalized into a normal view using learned VTM(s). The elastic net is used in the regression function, which is problem free of over fitting and stable regression models in VTM construction.

P.B. Shelke, P.R. Deshmukh[5] proposes a method based on body parts that features as subject identification on the basis of Rectangular Region developed based on Silhouette Analysis (RRSA) algorithm mainly analyze the significance of body parts individuallyfor the correct identification of the subject. This experiment uses CASIA dataset with 20 subjects.

SoumabhaBhowmick, Anup Nandy, Pavan Chakraborty, G. C. Nandi [6] propose a path on how different apparel worn by a people reflects on the extraction feature of an individual's gait pattern. A computer vision based method is applied to acquire gait behavioral feature like gait entropy image to extract the feature vector. A statistical based Naïve Baye's condition probability function is the classifier used with OU – ISIR dataset considering 15 subjects with 16 different cloths for each subject.

Shiqi Yu, Daoliang Tan, Tieniu Tan[6] propose to use three different data sets namely A, B and C which has normal data walking set sequence, walking data set with different clothes and walking data set with different load carry respectively. Each data set is captured using 11 different cameras at a varying multiple angle of 18°.GEIand standard deviation is captured as feature and is used to analyze the end results. The experiment has considered 124 subjects with each of image size 320 x 240 at 25 frames per seconds and is tested on CASIA gait dataset.

2. Working Principle

The proposed method is indicated in figure 1, 32 subject's gait cycles are considered from standard data set CMU MOBO with 11 different apparels each and with 10 different view angle for experiments. Three different feature extraction algorithms namely CHOG (The circular variant of HOG Technique) and EFT are computed as gait behavioral feature set to train the system, these features are tested with SVM classifier for gait identification efficiency analysis with respect to multiple clothes and view angles.

Further at the input side for every subject's four different gait cycles are considered for training the system with respect to each view angle. The gait cycle video is considered with 16 image frame sequences of gait image. Each of these images is pre processed to apply background removal and retain only the foreground subject in the image. The subjects behavioral gait feature is now obtained using Radon transform, CHOG feature coefficients in spatial domain and Elliptical Fourier co-efficient features in frequency domain. These multiple features extracted and are averaged over one complete gait cycle image frame sequences and are stored as a gait.

3.1 Background Subtraction

Statistical method is employed to attain background subtraction of gait cycle video frame sequences. Background subtraction is used to produce high quality silhouettes which is further used for human identification in gait recognition. By obtaining the gradient differences between current image with the modeled background image pixel by pixel, locomotion in the image can be identified. The method is effective against noise, shadow and change in lighting conditions. Initially, background frames from different gait cycles are collected and the average background frame is modeled using equation1.

$$B_{s,\,t+1} = (1-\alpha)\, B_{s,\,t} + \alpha\, I_{s,\,t+1} \quad (1)$$

Where, $\alpha = 1/t+1$, $I_{s,t}$ = current background frame,

A = .5 gives average

In an Average background frame each pixel will have average intensity of red, green and blue intensities of all the frames for same pixel using equation 2.

ISSN: 2233-7857 IJFGCN

Distance = $|I^{R}_{s(t)} - B^{R}_{s,t}|+|I^{G}_{s(t)} - B^{G}_{s,t}|+ |I^{B}_{s(t)} - B^{B}_{s,t}|$
(2)

The average background is found by computing the average Red, Blue and Green intensities for each pixel. The Red, Blue and Green gradient intensity variation between the median average image and each individual image background results into the pixel that has intensity gradient diminishing to zero this intensity of differenced gradient pixel enters into a highly correlated ellipsoidal shaped trivariate normal distribution.

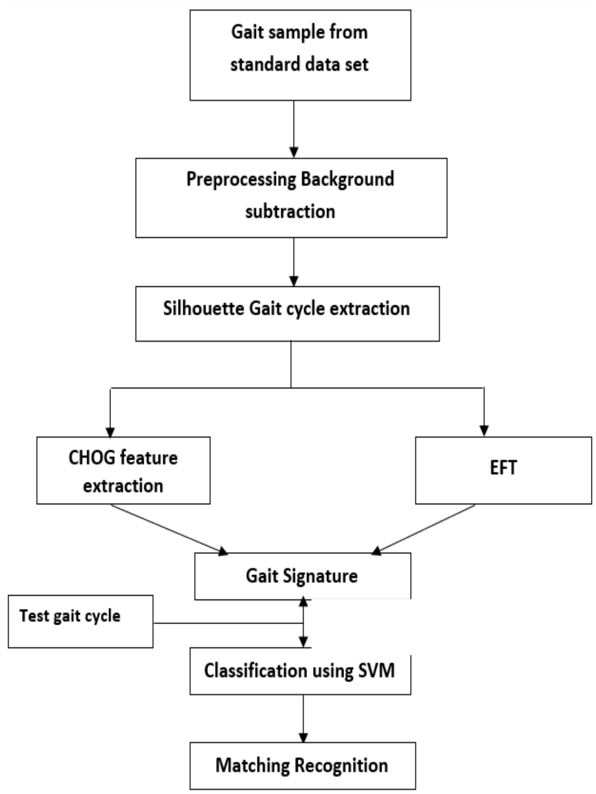


Figure 1. Flow chart of the proposed work methodology

After differencing, all the pixels related to background become a part of the trivariate normal distribution. A contour created using specified number of standard deviations at distance of mean values of RGB intensities, produces an ellipsoid. If the value of differenced pixel inside ellipsoid is treated as background pixel and the one which falls outside, it will be regarded as apart of the walking person.

$X = R^2 / (k \times \rho(R)) + G^2 / (k \times \rho(G)) + B^2 / (k \times \rho(B))$ (3)

Where R is the intensity of red, G is the intensity of green and B is the intensity of the blue for a differenced pixel. ρ(R) is the variance found for the red pixel, ρ(G) is the variance of pixel green, and ρ(B) is the variance of pixel blue, and k is a constant based on the variances of the distributions. If $X<1$, pixelis considered as background. The pixel is assigned a value of zero. If $X>1$, means pixel is not in the background and is a part of the subject walkingin the frame.The pixel is assigned a value of one.

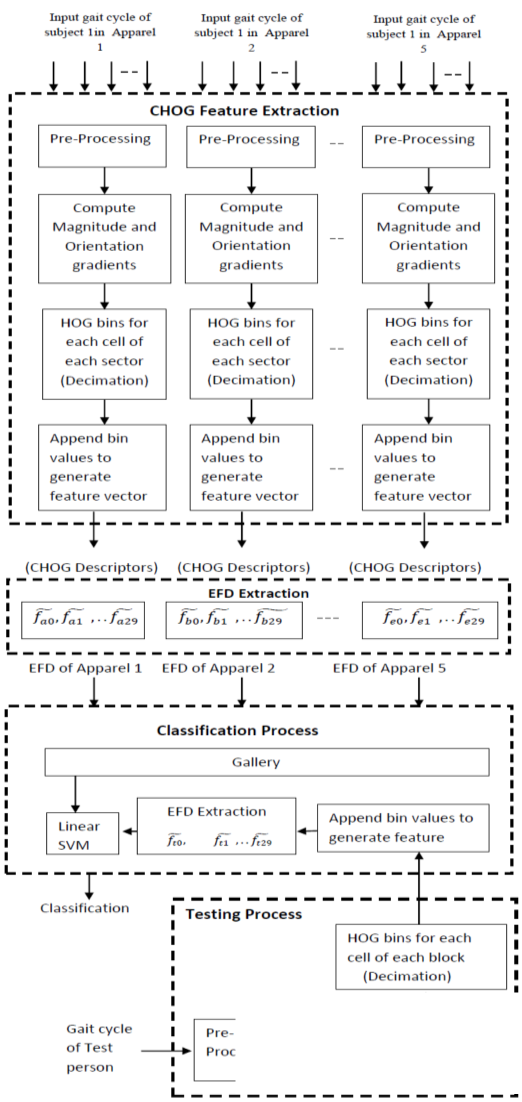


Figure 2 : Block diagram depicting CHOG and EFT feature extraction and classification using SVM Technique

The noise in the silhouettes is controlled by making ellipsoid longer and narrower by using covariance matrix.

$$C = \begin{bmatrix} 8.880 & 6.888 & 7.433 \\ 6.888 & 8.029 & 7.221 \\ 7.433 & 7.221 & 9.218 \end{bmatrix} \quad (4)$$

$$X = [R \;\; G \;\; B]^T * \begin{bmatrix} 8.880 & 6.888 & 7.433 \\ 6.888 & 8.029 & 7.221 \\ 7.433 & 7.221 & 9.218 \end{bmatrix} * [R \;\; G \;\; B]$$

$$X_s(t) = 1 \;\; \text{if } d(I_{st}, B_s) > \rho \quad (5)$$

0 otherwise, Where ρ is threshold value After testing all the pixel in the image as shown in equation5,the image is converted into binary image .this image is now referred to as a silhouette of the image. Thus a sequence of images is formed with a black background and white silhouette image.

3.2 Pre-processing On Silhouettes

When the subject under consideration walk towards the camera, the relative distance between the subject and camera changes. When the subject is nearer to the camera, the subject gait silhouette becomes bigger and becomes smaller when the subject is far. Thus, the preprocessing is performed to normalize binary gait silhouettes before computing the features. Depending on the size and the center of gravity of the silhouette, then normalized Silhouettes are computed as given in equation

$$s(i,j)=\begin{cases} 1 & \left| \begin{matrix} if\,(i,j) belongs\ to\ the\ foreground \\ otherwise \end{matrix} \right\} \\ 0 & \end{cases} \quad (6)$$

The center of the area (i_c ,j_c) of a silhouette is found using the equation 7

$$i_c = \frac{1}{P}\sum_{i,j} i\ .\ s(i,j) \quad j_c = \frac{1}{P}\sum_{i,j} j\ .\ s(i,j) \quad (7)$$

Where Pgives the number of foreground pixels and is given by P=Σ $S(i,j)$. Silhouettes are resized so that all silhouettes have the same height (the images are *120×120* in our approach), and then the center the silhouette image with respect to the horizontal center.

3.3 Acquiring Gait Cycle

Next step is to find the gait cycle length. Atperticular time t, an image drawn from the periodic image sequence with period P is denoted by the vector $\boldsymbol{x}$ (t), and satisfies$X$ (t + P) = x (t) $\forall t$. When a person is walking, the maximum width of his/her lower half silhouette presents obviously periodic change. The maximum width in a gait sequence shall be given as V=[v1,v2,...,vn]. Since the maximum value can be noisy and unstable, To acquire the best Gait cycle, the cycle length TG is identified by computing the autocorrelation coefficient R=[r1,r2,…,rn] of the signal V.

3.4 CHOG

In the gait image, local intensity gradient of each and every pixel is found by capturing the energy intensity and its flow direction using the new variant of HOG called as CHOG descriptor. The Gait image is divided into four non overlapping annual A1, A2, A3 and A4 from the center of the image. Further each annual is divided into four sub parts called annular cells by marking horizontal line and vertical passing through the centre of the image as shown in figure 3a. The local intensity gradient magnitude and their orientation of all the pixels are calculated using equations 8, 9, 10 and 11. The normalization of the same is done using equation 12.These gradient descriptors are quantized into 10 different angular bins based on their orientation angle of each pixel where each bin will cover 18^0 orientation as shown in figure 3b. Hence each cell will have 10 histogram feature descriptors as shown if figure 3c. The final total CHOG descriptors are the total concatenation of histogram feature descriptions of all the cells of the inner most annual followed by the outer annual cell and so on till all the annuals are covered as show in figure 3d.

The CHOG feature vector is computed using initial order high pass filter that has directive filter mask [-1,0,1] and $[-1,0,1]^T$along x and y direction respectively is used. The x - axislocal gradient and the y - axislocal gradient of each pixel is found as a measure of change in gradient value along horizontal direction and along vertical direction around every pixel as given in equation number 8 and 9.

$$Grad_x = I(x+1, y) - I(x-1, y) \quad (8)$$

$$\text{Grad}_y = I(x, y+1) - I(x, y-1) \quad (9)$$

Where $I(x,y)$ is local gradient value of pixel in the image at the location point (x,y).The magnitude intensity of gradient at the pixel I (x,y), is shown by

$$M = \sqrt{(\text{Grad}_x)^2 + (\text{Grad}_y)^2} \quad (10)$$

Its orientation angle is computed by

$$\text{M\_Angle} = \arctan\left(\frac{\text{Grad}_x}{\text{Grad}_y}\right) \quad (11)$$

To make the gradient value stable against illumination and shadow it should be normalized using L-2 normalization [13]. This eliminates local contrast and brightness variation effects. The L-2 normalization is given as

$$\text{Feature\_normalization} = \sqrt{\frac{\text{Feature}}{(\|\text{Feature}\|_2^2 + e^2)}} \quad (12)$$

The CHOG feature vector length is calculated as Fc = no. of annuls × no. of cells per annual × no. of bins (13)

3.5 Elliptical Fourier Transform

The angular curve in the human gait can be accurately traced using EFT feature. The CHOG feature coefficients in spatial domain are transformed into Elliptical Fourier coefficient features. The gait image's structural feature is represented by sum of harmonics using Fourier series expansion of its coordinates. Four terms are required to quantify each harmonics namely a_{xk} ,b_{xk} , a_{yk} , and b_{yk}. These coefficients contain information about location, size and angular orientation of the image outline. The CHOG's annuls are compared with polar coordinate system in which annuls gives distances to the pole and bin orientation is indulged angle. Further, the annuals in the CHOG can be matched with polar co-ordinate system in which the different annuls indicate various distance to the pole and orientation bin is indulged angle. The Cartesian coordinates are computed from polar coordinates using the equations 14, 15 and 16.

$$x = \rho\cos(\theta), \quad y = \rho\sin(\theta) \quad (14)$$

Where (ρ, θ) are the coordinates in the polar system and (x,y) are the coordinates in the Cartesian system and

$$\rho = (\rho_0 + \text{Feat}(q \times r, p)) \quad (15)$$

$$\theta = 2\pi\frac{2\pi}{cell\_num} \times q + r \times \frac{2\pi}{cell\_num \times bin\_num} \quad (16)$$

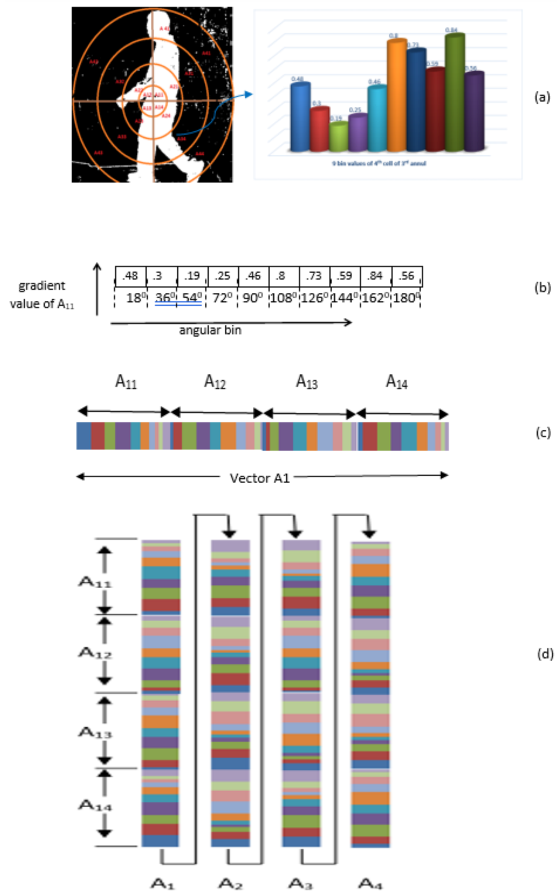


Figure 3. CHOG feature extraction process. For given gait image with 4 annuls and 4 cells in each annuli (a) , sample feature extraction of first annuli's first cell A_{11} with 10 bin values (b) , feature extraction of all 4 cells of annuli 1 i.e., A_{11}, A_{12}, A_{13}& A_{14} (c) , CHOG descriptor of all 4 annuli (d)

Where p, q and r represent the point of the angular orientation r of annular cell k in annulus p, and ρ_0 is a constant. The parameter ρ_0 is used because of very small accumulated gradient. The EFT is applied on a complex plane where all the pixels in the image are represented by complex number in which the first coordinate refers to the real part, and the second coordinate refers to the imaginary part as shown in equation 17.

$$C(n) = x(n) + I\, y(n) \qquad (17)$$

Where x (n), y (n) is the coordinate point of the image. To perform EFT, Equation is expanded. In which EFT coefficients can be defined using equation 18.

$$C(k) = c_{xk} + ic_{yk} \qquad (18)$$

Based on the relationship between the coefficients in the form of the exponential function and the trigonometric function, the coefficients c_{xk}andc_{yk}in the trigonometric function can be defined as the discrete coefficients a_{xk} ,b_{xk} , a_{yk} , and b_{yk} with k elements.

$$a_{xk} = \frac{2}{m}\sum_{l=1}^{m} x(l)cos(kwl\tau)\ ,\ b_{xk} = \frac{2}{m}\sum_{l=1}^{m} x(l)sin(kwl\tau)\quad ,a_{yk} = \frac{2}{m}\sum_{l=1}^{m} y(l)cos(kwl\tau)$$

$b_{yk} = \frac{2}{m}\sum_{l=1}^{m} y(l) sin(kwl\tau)$
(19)

where $x(l)$ and $y(l)$ are the values of the sampling points of the functions x(j) and y(j), m is the total number of sampling points, and w is the fundamental frequency, which is equal to T=2π, where T is the period of the function.

The curve c(j) can be described as$c(j) = \frac{a_{x0}}{2} + \sum_{k=1}^{\infty}(a_k\, cos(kwj) + b_{xk}\, sin(kwj)) + i(\frac{a_{y0}}{2} + \sum_{k=1}^{\infty}(a_{yk}\, cos(kwj) + b_{yk}\, sin(kwj)))$
(20)

In matrix form it can be written as $\begin{bmatrix} x(j) \\ y(j) \end{bmatrix} = \frac{1}{2}\begin{bmatrix} a_{x0} \\ a_{y0} \end{bmatrix} + \sum_{k=1}^{\infty}\begin{bmatrix} a_{xk} & b_{xk} \\ a_{yk} & b_{yk} \end{bmatrix}\begin{bmatrix} cos(k\omega j) \\ sin(k\omega j) \end{bmatrix}$
(21)

For each item in equation (20), if k has a fixed value, the sum of the trigonometric functions defines an ellipse in the complex plane. Assuming a change in j, the point will move along the ellipse at a speed that is proportional to the associated frequency k, where k is the number of circles that pass through this point. Each spindle of the ellipse is calculated using ak and bk .If ρ is the rotation angle, the coordinate values are as follows
$\begin{bmatrix} x'(j) \\ y'(j) \end{bmatrix} = \frac{1}{2}\begin{bmatrix} a_{x0} \\ a_{y0} \end{bmatrix} + \begin{bmatrix} cos(\rho) & sin(\rho) \\ -sin(\rho) & cos(\rho) \end{bmatrix}\sum_{k=1}^{\infty}\begin{bmatrix} a_{xk} & b_{xk} \\ b_{yk} & b_{yk} \end{bmatrix}\begin{bmatrix} cos(k\omega j) \\ sin(k\omega j) \end{bmatrix}$
(22)

4. RESULTS AND DISCUSSIONS

4.1 Experiment 1

In the figure 4, the pixel point 'P' is the position at which CHOG is to be computed. Point 'P' is at angle β from the origin. Its orientation angle is γ and the total angle α = β + γ. Now if the view angle of point 'P' changes to 'P^{1}' then the view angle $\alpha^1 = \beta^1 + \gamma^1$. It can be show that $\gamma^1 = \alpha 1 - \beta^1 = (\alpha + \rho) - (\beta + \rho) = \alpha + \beta = \gamma$, this proves the point of view angle invariance. The CHOG results show that view angle invariance can be achieved on every binning angle. ie if the point 'P' changes with a view angle of18^0 or multiple of it, then view angle invariance can be achieved.
Figure 5 indicates that, when the original gait image with 0^0 view angle is tested versus the same subject image with 18^0 view angle, the related CHOG histogram remains same with its each histogram shifting by one place with respect to histogram of its original image, thus retaining the same gradient energy with its orientation value and thereby achieving view angle invariance. This suggests that CHOG feature contributes to 100% recognition rate, but at only view angles multiple of binning angle. For any random view angle the CHOG feature showcases poor recognition rate.

4.2 Experiment 2

The four coefficient terms of the elliptic Fourier spectrum harmonics a_{xk} ,b_{xk} , a_{yk} , and b_{yk}have two terms for the *x* series and two for the *y*. It contains information about the location, size, and rotational orientation of the gait image outline and is calculated over a series of harmonic amplitudes (n). In this work arbitrary criterion is used to decide on how many Fourier harmonics are necessary to quantify the image shape present in the sample gait image. The gait image outline was represented with EFT with variable amplitude harmonics as shown in the figure 6. The best structural representation of the gait for entire image sequence cycle was achieved for 40 harmonics. The experiment of EFT with n=40 would therefore require 160 coefficient terms to represent each of the gait image feature. The reconstruction or matching of the gait cycle was best achieved with harmonic amplitude equal to 40. The EFT with harmonics less than 40 will not give accurate representation of the gait image and it becomes redundant in gait cycle matching for harmonic amplitude greater than 40.

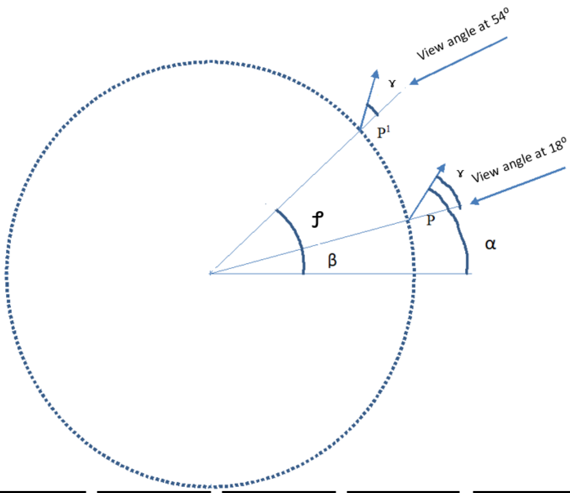


Figure 4 : The gradient magnitude and its orientation of a pixel at point 'P' versus new view angle P¹

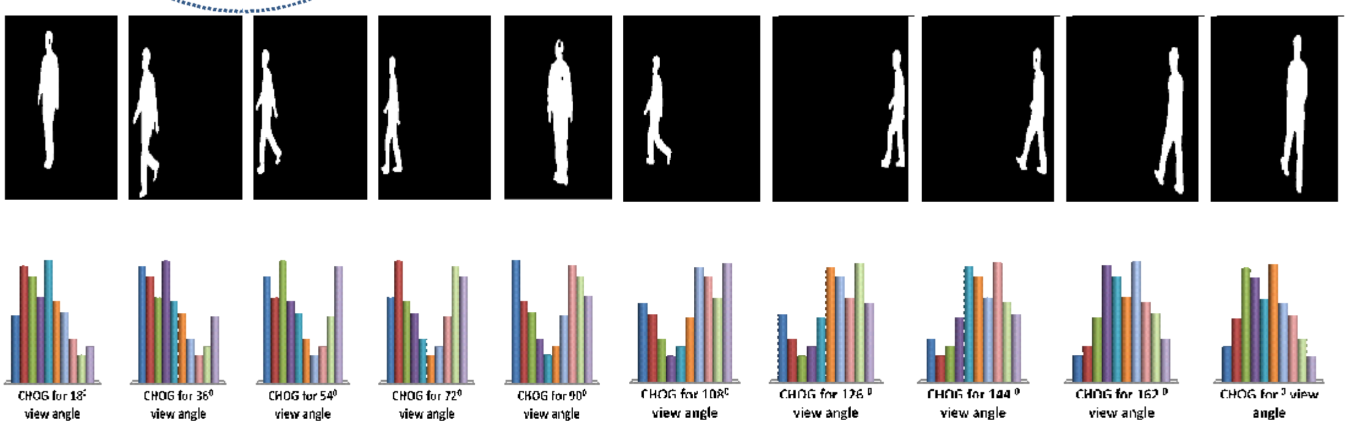


Figure 5. Sample circular histogram of oriented gradients of the subject gait with different view angles

4.3 Experiment 3

The experiment is carried out to compute the effectiveness of CHOG and EFD feature extraction technique on gait recognition for multiple view angles and multiple cloth approach. It underlines the significance of these techniques through their effectiveness on gait invariant approach. The experiment is performed on standard data set CASIA dataset B and CMU MoB. 36 subjects with 11 different clothes, 3 different objects and 10 different view angles are considered in which each gait cycle has 16 frames of 128 x 256 resolution. For each subject 4 gait cycles were used for training and 2 other gait cycles were used for testing. The input gait image sequences of the subject with single cloth, with 11 different clothes and in multiple view angles were considered. The graph 1 shows that the recognition rates have increased to 83 % for cloth invariance gait recognition versus 47% recognition rate achieved by abbas G[10]. The Rank 1 recognition rate for view invariant gait stands at 97% has outperformed the results of 87% achieved by Wei Zeng and Cong Wang [11]. The efficient results shown on graph 1 were majorly achieved from Elliptical Fourier Transform with 40 ellipses (n=40). The Elliptical Fourier Transform on normal image merely gave 72% recognition rate, but when EFT was applied on CHOG features of the image the recognition rate surged to 97% suggesting that CHOG features strongly hold theeffective distribution of local intensity gradient that helps us trace the small inter frame gradient movement.

International Journal of Future Generation Communication and Networking
Vol. 13, No. 3, (2020), pp. 243- 252

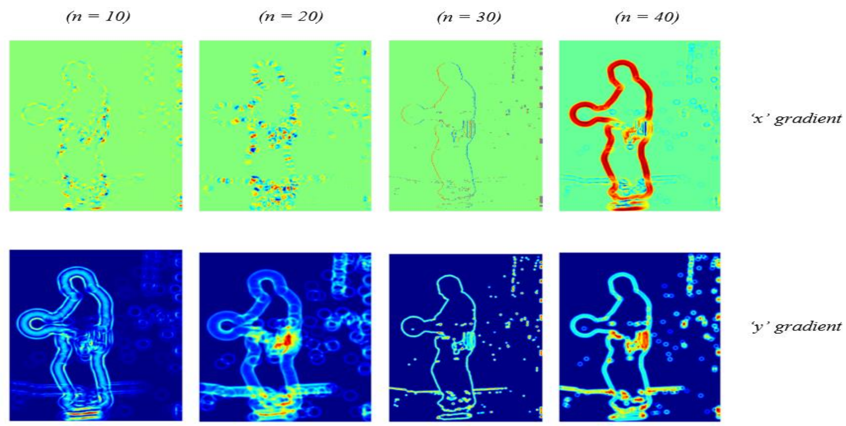
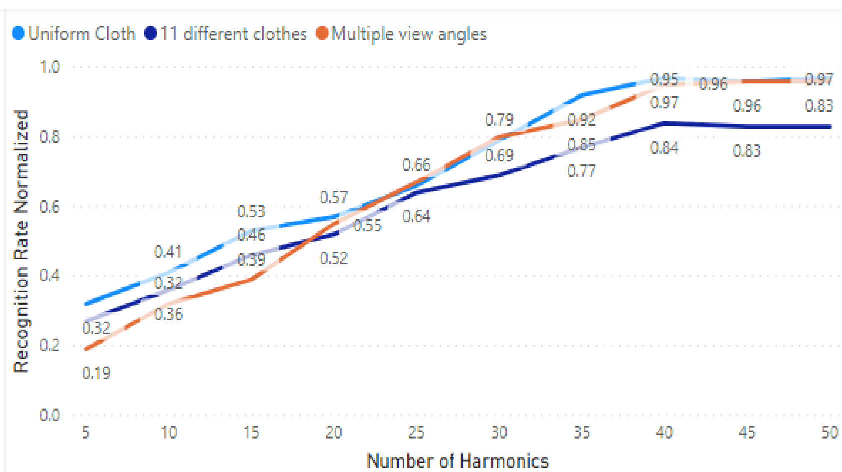


Figure 6. Shows the elliptic Fourier gradient in x and y direction for given number of harmonic amplitude(n)



Graph 1. Performance of CHOG EFT on multi-view angles and multi-cloth gait recognition

ISSN: 2233-7857 IJFGCN