A Frequency Domain Approach to Silhouette based Gait Recognition

Soumyadip Sengupta[a], Udit Halder[a], Rameswar Panda[b], Ananda S. Chowdhury[a]

[a]Department of Electronics and Tele-Communication Engineering, Jadavpur University, Kolkata, India.
[b]Department of Electronics and Tele-Communication Engineering, Silicon Institute of Technology, Bhubaneswar, India.

Email: senguptajuetce@gmail.com, udit.halder@ieee.org, rameswar.panda@silicon.ac.in, aschowdhury@etce.jdvu.ac.in

Abstract— In this paper, we propose a frequency domain based model-free gait recognition approach from silhouette inputs using Fourier Transform. Gait sequences are first converted into frequency domain using Fourier transform. Information content of the frequency components are analysed next to determine the number of effective frequencies which can help in the recognition process. These principal frequencies are treated separately to obtain scores based on the correlation coefficient between the gallery and the probe images. The individual scores are fused in the last stage to obtain the final score. The proposed approach is compared with other state-of-the-art model-free gait recognition algorithms. Experimental results on the USF HumanID database clearly indicate the supremacy of our technique.

I. INTRODUCTION

Gait refers to the style of walking and is hence unique for every individual [8] [9]. In recent years, gait recognition has gained much attention in the field of biometric authentication [2] [3] [4]. Other popular biometrics like fingerprint, iris or face recognition require cooperative subjects, view from some specific angles and physical contact or proximity [1]. In sharp contrast, gait can be acquired even from a distance and can serve as an important biometric in several applications (e.g. surveillance) where acquiring fingerprint, face or iris information can be quite difficult.

Existing gait recognition methods can be broadly categorized into two groups, namely, model-based and model-free. Model-based approaches [10] [11] match the gait sequences with some pre-defined models of human walking and then extract useful features for the purpose of recognition. These methods are insensitive to view points but complex models need to be generated for a recognition problem with huge population. Model-free approaches do not require such structural models of human motion. Since, the proposed method falls under the model-free category, we now mention only some well-known model-free gait recognition techniques. Kale *et al.* [5] proposed a Hidden Markov Model (HMM) based framework for gait recognition. Sarkar *et al.* [2] proposed the baseline algorithm where they have directly measured the similarity between gallery sequence and probe sequence to perform the recognition. Han and Bhanu proposed Gait Energy Image (GEI) [3] which is the average image over a gait cycle. GEI was found to be less sensitive to the silhouette noise in individual frames. Bashir *et al.* proposed Gait Entropy Image (GEnI) [4] for automatic feature selection between gallery sequence and probe sequence. Yu *et al.* proposed a gait recognition algorithm based on Key Fourier Descriptors (KFD) [7].

In this paper, we propose a Fourier Transform based approach for gait recognition from silhouette data. Among different types of frequency domain transforms, Fourier transform is chosen as it is simple and fairly efficient. The binary gait silhouette images are first converted to the frequency domain. Only certain frequencies with information content above a pre-defined threshold are selected next automatically. Each such frequency is then treated separately to yield a similarity score between the gallery and probe images using a correlation based approach. Gallery images or sequences of images are the generally the set of all known individuals and the probe is the image or sequence of images of unknown individual whose identity has to be verified with the individuals in the gallery. The individual scores are finally fused at the decision level to achieve the overall recognition. The main contribution lies in application of frequency domain approach for achieving accurate gait recognition.

II. FOURIER TRANSFORM BASED GAIT IMAGE REPRESENTATION

A. Binary Silhouette Extraction

From a given image sequence of the motion of a human being, a human silhouette is extracted from each frame using the method described by Sarkar *et al.* [2]. After applying size normalization and horizontal alignment to each extracted silhouette image, gait cycles are segmented by estimating gait frequency through maximum entropy.

Now consider the time variation of intensity each pixel of the binary silhouette over the entire image sequence. As each pixel is binary in nature, this time variation is just a periodic variation of binary signal, where period of the waveform is equal to N. In earlier works, researchers have produced GEI images out of this time variation. GEI image is nothing but the time average of this waveform. But certainly in using a simple time-average we lose a lot of vital information. Also if we are able to transform this binary information into a real number domain, we may have more information at hand to work with. This influences our idea of transforming the images from binary space to a real valued space. This is where Fourier Transform plays the trick. Fourier Transform is effective in capturing the time variation of a signal and to reflect that variation in the frequency domain.

B. Fourier Transformation over the binary silhouette image

Any individual pixel within the video sequence is represented by $I(x, y, t, i)$. Here I denotes the silhouette image, x and y denote the pixel co-ordinates, t denotes the frame number within a gait cycle and i denotes the gait cycle number. Note that $t = 0, ..., N-1$ and $i = 1, ..., C$. N denotes the number of frames within one gait cycle and C denotes total number of gait cycles. As a binary silhouette is periodic in nature with period N, we perform a N point Fast Fourier Transform (FFT) to capture the temporal variation of individual pixels.

$$I(x, y, k, i) = \sum_{t=0}^{N-1} I(x, y, t, i) exp(-\frac{2\pi kt}{N}) \qquad (1)$$

The amplitude spectrum ($\widehat{I}(x, y, k, i) = \|I(x, y, k, i)\|$) of these signals is used to construct the gait image. Here gait image is the combined representation of the human movement in the given sequence at a particular frequency. Thus we have one Gait Image at each frequency.
The amplitude spectrum is symmetric about a central point, except the DC component. So, only $\lceil N/2 \rceil$ frequencies are sufficient to characterize the original video sequence. But even processing these $\lceil N/2 \rceil$ Gait images for recognition is computationally intensive. So, we plan to use an optimal number of frequency images for obtaining good recognition.

In Figure 1, sample binary silhouettes of a gallery sequence and their GEI image is shown. In Figure 2, gait images corresponding to first 8 frequencies for a sample gallery sequence are presented. The first image, *i.e.*, the d.c. component is actually the GEI image. The other gait images capture the movement characteristics of the person as shown in the figure. So frequency decomposition additionally generate higher order GEI images which capture the motion characteristics better and thus when used together improves the overall performance. We represent each gait image as $G(k, i)$ at gait cycle k and frequency i.

C. Gait Image at Each Frequency

At present we have $C * \lceil N/2 \rceil$ gait images, *i.e.*, C gait images for each frequency. This number is also large for a single person. So, we aim to combine the gait cycles using suitable weights $(w_{i,k})$ to produce single gait image per frequency. The weights are chosen based on the entropy of individual gait images and are given by:

$$w_{i,k} = - \sum_{l=0}^{M-1} p_l \ln(p_l). \qquad (2)$$

where M is the number of gray levels and p_l is the probability associated with gray level l. Maximum entropy is achieved in the case of a uniform probability distribution. If $M = 2^n$, then p_l is a constant and is given by:

$$p_l = \frac{1}{M} = 2^{-n}. \qquad (3)$$

So, the net gait image for each frequency is given by:

$$G(k) = \sum_{i=1}^{C} w_{i,k} G(k, i). \qquad (4)$$

D. Selection of Frequency Threshold

At this stage, we have $C * \lceil N/2 \rceil$ gait images for each frequency. In a typical gallery sequence we have $N = 38$ and $C = 5$ and hence we have total 95 gait images which is too high for computation. So a reduction in the number of gait images is necessary.

As each images are of silhouette, they are essentially grey level images and hence their amplitude spectra at each frequency is also a grey level image. Hence Entropy of these gait images is calculated at each frequency. Figure 4 shows a curve illustrating a decrease in the information content with increase in frequency. So, we apply a cut-off λ, and choose any of those frequencies whose information content is above this threshold. This process is performed for all the image sequences in the gallery set, and the average number of frequency components (F) is determined. We use this F to be the number of gait images for recognition as well as the threshold for all probe sets.

III. CORRELATION BASED RECOGNITION

Here we use correlation among two data sets as a measure of their similarity. Correlation based recognition [7] is simpler and faster as compared to PCA or HMM based recognition techniques. Pearson correlation coefficient is employed in this paper to restrict the net score between -1 and +1. The similarity co-efficient is given by [6]:

$$\varrho = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum(X_i - \overline{X})^2}\sqrt{\sum(Y_i - \overline{Y})^2}}. \qquad (5)$$

where, X and Y represents the two gait sequence, formed by concatenating the rows of gait image. $\overline{X}$, $\overline{Y}$ represent the average values.

Gait images at each frequency is used separately to perform recognition. After a score is generated, the identification experiment is performed on the basis of rank r. Each gait image of each probe sequence is compared with respective frequency based gait images of all the gallery sets and corresponding scores are generated. These scores are sorted in descending order. If the probe is present within the top r scores, it is considered as being recognised. As the above process is carried out for all the frequencies, a binary vector B is generated, whose elements B_i ($i = 1, ..., F$) signify whether a probe is correctly recognised at a given frequency i. Overall recognition is achieved by fusing the B_i's using OR voting rule. OR voting rule is like a simple OR operation, where we assume correct recognition is performed if any of the B_i's is 1.

IV. EXPERIMENTAL RESULTS

We used the binary silhouettes from USF HumanID gait challenge dataset (version 2.1) [2] for evaluating the performance of our proposed algorithm. The dataset contains persons walking in elliptical paths in front of camera and includes up to six covariates for each person. Twelve experiments (experiments A to L) are designed for human recognition problem. The specifications of gallery set, each probe set and corresponding covariates are given in Table I. It is to be noted that, higher the number of covariates, more difficult becomes the task of recognition. For our experiments, we

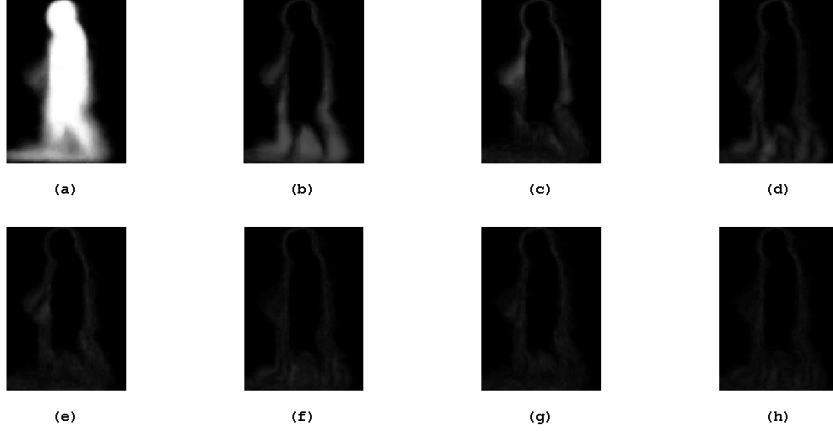Fig. 1. Sample binary silhouettes of a gallery sequence and their GEI image (extreme right)

Fig. 2. Gait images corresponding to first 8 frequencies for a sample gallery sequence: (a) GEI, (b)-(h) other higher frequency components

TABLE I. COMPARISON OF IDENTIFICATION RATES OF BASELINE ALGORITHM, GEI AND OURS. LEGENDS FOR COVARIATES: 1:VIEW; 2:SHOE; 3:SURFACE; 4:BRIEFCASE; 5:TIME; 6:CLOTHES. GALLERY SET: G, A, R, NB, M/N

| Exp.	Probe [No. of Subjects]	Covariates	Rank 1 Performance(%)				Rank 5 Performance(%)		
			Baseline	GEI	HMM	Ours	Baseline	GEI	Ours
A	G, A, L, NB, M/N [122]	1	73	90	89	97	88	94	100
B	G, B, R, NB, M/N [54]	2	78	91	88	93	93	94	95
C	G, B, L, NB, M/N [54]	1, 2	48	81	68	95	71	93	95
D	C, A, R, NB, M/N [121]	3	32	56	35	42	66	78	68
E	C, B, R, NB, M/N [60]	2, 3	22	64	28	50	55	81	69
F	C, A, L, NB, M/N [121]	1, 3	17	25	15	36	42	56	65
G	C, B, L, NB, M/N [60]	1, 2, 3	17	36	21	43	38	53	67
H	G, A, R, BF, M/N [120]	4	61	64	85	99	85	90	100
I	G, B, R, BF, M/N [60]	2, 4	57	60	80	89	78	83	98
J	G, A, L, BF, M/N [120]	1, 4	36	60	58	85	62	82	97
K	G, A/B, R, NB, N [33]	2, 5, 6	3	6	17	39	12	27	70
L	C, A/B, R, NB, N [33]	2, 3, 5, 6	3	15	15	34	15	21	69
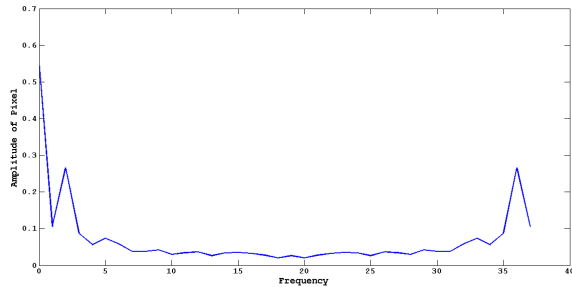
Fig. 3. Amplitude Spectrum of a sample pixel

have experimentally set, the threshold, $\lambda = 0.6$ and was kept constant throughout the experiment and the value of F was found to be 8.

The experimental results are given in Table I. Identification

TABLE II. COMPARISON AMONG RESULTS OF INDIVIDUAL FREQUENCY COMPONENTS AND FUSED RESULT: RANK 5 PERFORMANCE

Exp.	Results of Individual Frequency Components								Fused Result
	0th	1st	2nd	3rd	4th	5th	6th	7th	
K	39	36	58	36	48	55	58	52	70
L	24	24	48	27	45	42	39	36	69

rates P_I (in %) at ranks 1 and 5 are presented for the proposed algorithm and for other algorithms ([2], [3], [5]). Rank 5 results of HMM based algorithm over this particular dataset(Version 2) is not available in [5]. We have shown results for rank 1 and rank 5 as suggested in ([2]. The identification rates for a specific rank r are computed as follows. For a specific subject in probe set, we generate a score for each of the gallery subjects. The gallery subjects are sorted in a descending manner based on their respective scores. If the subject chosen in the probe set is found among the top r sorted gallery subjects; we say that the algorithm has correctly
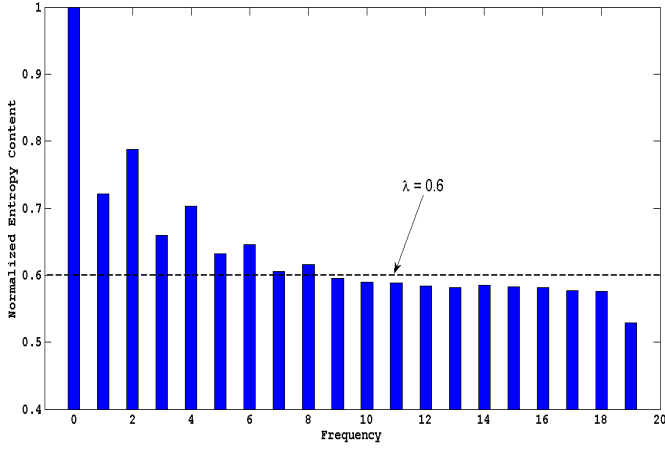
Fig. 4. Variation of information content of Gait images at different frequency

V. CONCLUSION

In this paper, a novel frequency based approach using Fourier transform is proposed to recognize individuals from their walking styles. Overall recognition is achieved by fusing the recognition scores of the individual frequency components. The experimental results on standard datasets indicate the efficiency of our algorithm. One direction of future research will be to achieve more meaningful image representation for improving the accuracy of recognition. We will also work towards a better fusion strategy to achieve high recognition rates.

REFERENCES

[1] A. K. Jain, A. Ross, S. Prabhakar, "An Introduction to Biometric Recognition", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4-20, 2004.

[2] S. Sarkar, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother, and K.W. Bowyer, "The Humanid Gait Challenge Problem: Data Sets, Performance, and Analysis", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 162-177, Feb. 2005.

[3] J. Han and B. Bhanu, "Individual recognition using Gait Energy Image", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 316322, Feb. 2006.

[4] K. Bashir, T. Xiang and S. Gong, "Gait recognition without subject cooperation", *Pattern Recognition Letters*, Vol. 31, pp. 2052-2060, 2010.

[5] A. Kale, A. Sundaresan, A. N. Rajagopalan, N. P. Cuntoor, A. K. Roy-Chowdhury, V. Krger, and Rama Chellappa, "Identification of Humans Using Gait", *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1163-1173, 2004.

[6] R. Sagawa, Y. Makihara, T. Echigo, Y. Yagi, "Matching Gait Image Sequences in the Frequency Domain for Tracking People at a Distance", *7th Asian Conference on Computer Vision*, pp. 141-150, Hyderabad, India, 2006.

[7] S. Yu, L. Wang, W. Hu, and T. Tan, "Gait Analysis for Human Identification in Frequency Domain", *Proc. of the 3rd International Conference on Image and Graphics (ICIG'04)*, pp. 282-285, 2004.

[8] M. P. Murray, A. B. Drought, and R. C. Kory, "Walking patterns of normal men", *J. Bone and Joint Surgery*, vol. 46-A, no. 2, pp. 335360, 1964.

[9] M. P. Murray, "Gait as a total pattern of movement", *American Journal of Physical Medicine*, vol. 46, pp. 290332, June 1967.

[10] D. K. Wagg and M. S. Nixon, "On automated model-based extraction and analysis of gait", *Proc. of 6th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 1116. Seoul, Korea, May 2004.

[11] S. D. Choudhury and T. Tjahjadi, "Silhouette-based gait recognition using Procrustes shape analysis and elliptic Fourier descriptors", *Pattern Recognition*, vol. 45, pp. 3414-3426, 2012.

identified that subject as mentioned in [2]. We perform this procedure for every probe subjects and P_I is calculated by the number of subjects those are correctly identified by the algorithm (represented in percentage). From Table I, it is evident that our algorithm outperforms its competitors in most of the experiments. In most difficult experiments (K-L), our method beats the other algorithms comprehensively. However, for two experiments (D, E) the proposed method falls marginally behind the competitors by around 10% with respect to GEI. Experiment D and E mainly deal with the change in ground condition, i.e., change in grass and concrete, which may cause the poor performance of our method. Performance of our algorithm is also shown by means of Cumulative Match Characteristics (CMC) curves [2] in Figure 5. The higher the slopes of the CMC curves, the better the performance of the algorithm. The identification rates of each frequency component along with the fused result are shown in Table II. This clearly demonstrates the superiority of the fusion strategy over the individual frequency components.
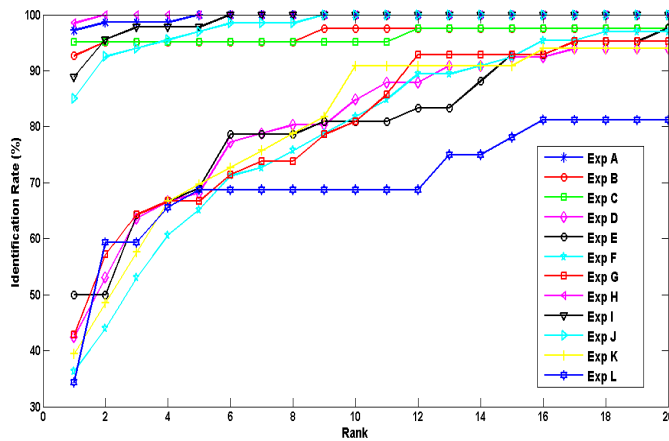

Fig. 5. Performance on 12 challenge experiments of USF HumanID dataset: CMC curves