A Review of Chinese Academy of Sciences (CASIA) Gait Database As a Human Gait Recognition Dataset

Rosa Andrie A[1,2], Achmad Basuki[1,3], Kohei Arai[1]
[1]Saga University, Japan
[2]State Polytechnics of Malang, Indonesia
[3]Electronic Engineering Polytechnic Institute of Surabaya, Indonesia
rosaandrie@gmail.com, basuki@eepis-its.edu, arai@is.saga-u.ac.jp

Abstract

Human Gait as the recognition object is the famous biometrics system recently. Many researchers had focused this subject to consider for a new recognition system. One of the important advantage in this recognition compare to other is it does not require observed subject's attention and cooperation. There are many human gait datasets created within the last 10 years. Some databases that are widely used are University Of South Florida (USF) Gait Dataset, Chinese Academy of Sciences (CASIA) Gait Dataset, and Southampton University (SOTON) Gait Dataset. This paper will analyze the CASIA Gait Dataset in order to see their characteristics. There are 2 pre-processing subsystems; model based and model free approach. We will use 2D Discrete Wavelet Transform (DWT). We select Haar wavelets to reduce and extract the feature.

Keywords: Gait Recognition, 2D Discrete Wavelet Transform (2D DWT), 2D Lifting Wavelet Transform (LWT), Haar Wavelet, CASIA Gait Dataset

1. Introduction

In recent years, there has been an increased attention on effectively identifying individuals for prevention of terrorist attacks. Many biometric technologies have emerged for identifying and verifying individuals by analyzing face, fingerprint, palm print, iris, gait or a combination of these traits [1].

Human Gait recognition giving some advantage compared to other recognition system. Gait recognition system does not require observed subject's attention and cooperation. It can also capture gait at a far distance without requiring physical information from subjects [3][4][5].

Human Gait Recognition as a recognition system divided in three main subject; preprocessing, feature reduction and extraction system, and classification.

There are 2 pre-processing subsystems to be used: model based and model free approach. Model-based approaches obtain a series of static or dynamic body parameters via modeling or tracking body components such as limbs, legs, arms and thighs. Gait signatures derived from these model parameters is employed for identification and recognition of an individual. It is evident that model-based approaches are view-invariant and scale-independent. These advantages are significant for practical applications, because it is unlikely that reference sequences and test sequences are taken from the same viewpoint. Model-free approaches focus on shapes of silhouettes or the whole motion of human bodies. Model-free approaches are insensitive to the quality of silhouettes. Its advantage is a low computational costs comparing to model-based approaches. However, they are usually not robust to viewpoints and scale [3].

There are some Human Gait Datasets widely used by researchers today. Many human gait datasets were created within the last 10 years. Some of widely used datasets are University of South Florida (USF) Gait Dataset, Chinese Academy of Sciences (CASIA) Gait Dataset, and Southampton University (SOTON) Gait Dataset. This paper reviews CASIA Gait Dataset to see their characteristics. CASIA Gait dataset has four class datasets: Dataset A, Dataset B (multi-view dataset), Dataset C (infrared dataset), and Dataset D (foot pressure measurement dataset). As a beginning, we use the B class dataset in 90 degrees point of view. This paper also reviews the database using 2D Discrete Wavelet Transform for the feature reduction and extraction.

2. Preprocessing

Some of the references use model-based gait as the preprocessing system. Model-based approaches are sensitive to the quality of gait sequences. Thus, to archive a high accuracy require high quality gait image sequences. Another disadvantage of the model-base approach is its large computation and relatively high time costs due to parameters calculations [2].

We create silhouettes by using background subtraction then apply some morphological operation to get the skeleton. Figure 1 is the block of background subsystem to get the silhouettes.

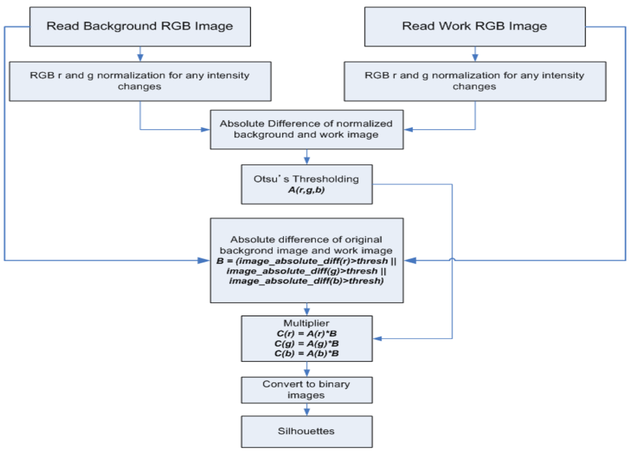

Figure 1. Background subtraction for silhouettes images

The model based pre-processing used in this paper is creating human model by skeleton per frame and per video sequence. First, we assign some crucial points of human joint movement. We will use 7 points as a reference. The points are waist, neck, head, both knee, and both ankle. Waist is in the middle. Waist – knee is about 25% length of the human body height. Knee – ankle is about 27% length of the human body height. Waist – neck is about 32% length of the human body height. Neck – top head is about 15% length of the human body height [6]. The skeleton created by using thinning and other morphological operation. From the skeleton created, we choose the key points, and reconstruct the skeleton by connecting the points using straight lines. Figure 2a is the example of the skeleton per frame, and figure 2b is the example of skeleton per frame sequence.

The model free preprocessing used in this paper by using the motion parameter per frame. We can get the motion by using background subtraction. The motion we get also per frame and per video sequence. Figure 3a is the example of the human motion per frame. Figure 3b is the example of human motion per video sequence.

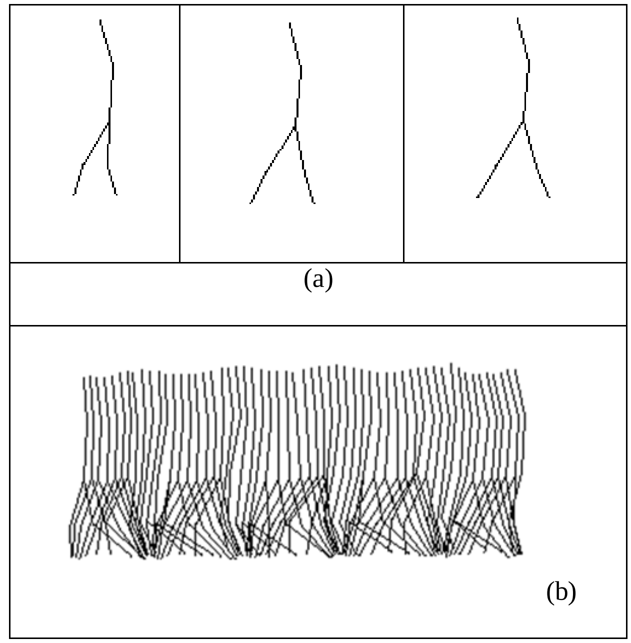

Figure 2. (a) Skeleton image per frame, (b) Skeleton image per video sequence

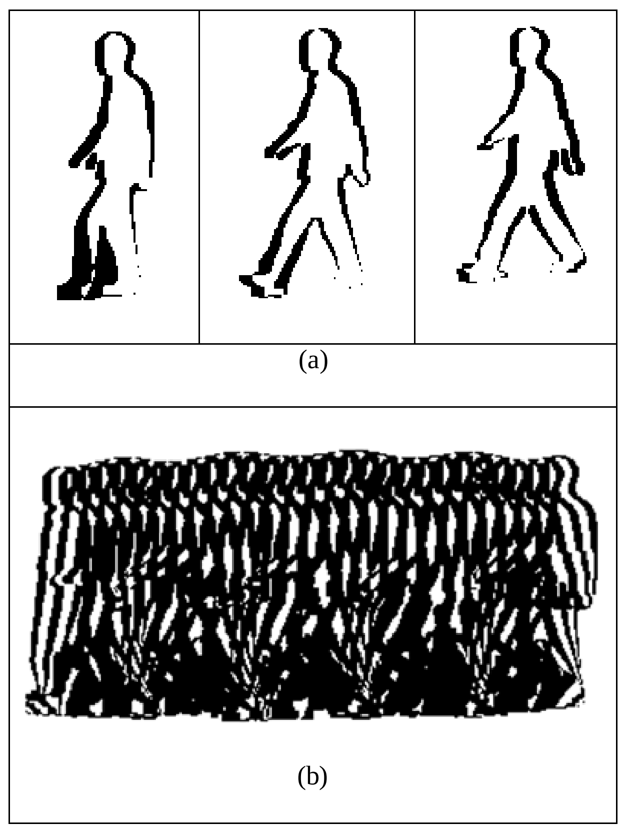

Figure 3. (a) Skeleton image per frame, (b) Skeleton image per video sequence

3. Discrete Wavelet Transform

Discrete wavelet transform (DWT) represents an image as a subset of wavelet functions using different locations and scales [8]. It makes some decomposition images. Any decomposition of an image into wavelet involves a pair of waveforms: the high frequencies corresponding to the detailed parts of an image and the low frequencies corresponding to the smooth parts of an image. DWT for an image as a 2-D signal can be derived from a 1-D DWT. According to the characteristic of the DW decomposition, an image can be decomposed to four sub-band images through a 1-level 2-D DWT, as shown in Fig. 4. These four sub-band images in Fig. 4 can be mapped to four sub-band elements representing LL (Approximation), HL (Vertical), LH (Horizontal), and HH (Diagonal) respectively.

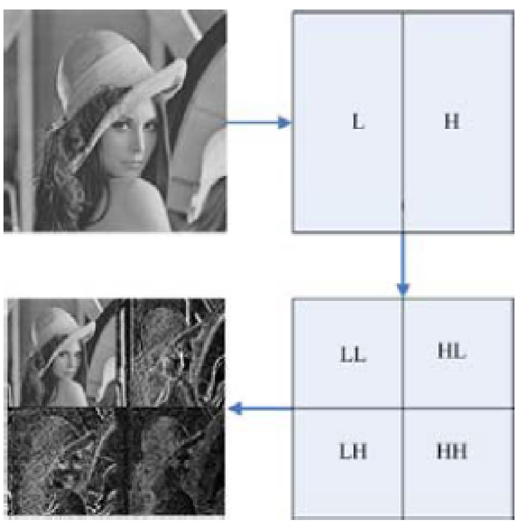

Figure 4. 1-Level Decomposition 2D DWT

The discrete Wavelet Transform will decompose a given signal into other signal known as the approximation and detail coefficients. A given function f(t) can be expressed through the following representation:

$$f(t) = \sum_{j=1}^{L} \sum_{K=-\infty}^{\infty} d(j,K)\varphi(2^{-j}t - K) + \sum_{K=-\infty}^{\infty} a(L,K)\theta(2^{-L}t - K) \qquad (1)$$

Where: $\varphi(t)$ is the mother wavelet and $\theta(t)$ is the scaling function. $a(L,K)$ is called the approximation coefficient at scale L and d(j,K) is called the detail coefficients at scale j. The approximation and detail coefficients can be expressed as

$$a(L,K) = \frac{1}{\sqrt{2^L}} \int_{-\infty}^{\infty} f(t)\theta(2^{-L}t - K)dt \qquad (2)$$

$$d(j,K) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{\infty} f(t)\varphi(2^{-j}t - K)dt \qquad (3)$$

Based on the choice of the mother wavelet φ(t) and scaling function θ(t), different families of wavelets can be constructed [7]. We will use the Haar (DB1) Wavelet and in level-1 decomposition.

To draw easier the characteristics of the Wavelet coefficients, we use the energy of each coefficient, then creates the 2D scatter graph for every combination of the coefficients.

These are the formula of the energy for every coefficients in a frame sequence (video) :

$$E_{a(L,K)} = \sum_{i=1}^{video} \sqrt[2]{\sum_{j=1}^{frame} |a(L,K)|^2} \qquad (4)$$

$$E_{d(j,K)} = \sum_{i=1}^{video} \sqrt[2]{\sum_{j=1}^{frame} |d(j,K)|^2} \qquad (5)$$

Then we do an energy normalization using the formula below,

$$E_{Total} = E_{a(L,K)} + E_{d(j,K)} \qquad (6)$$

$$\%E_{a(L,K)} = \frac{100 \times E_{a(L,K)}}{E_{Total}} \qquad (7)$$

$$\%E_{d(j,K)} = \frac{100 \times E_{d(j,K)}}{E_{Total}}\ \% \qquad (8)$$

Which percentage of total energy is:

$$100\%\ Energy = \left(\%E_{a(L,K)} + \%E_{d(j,K)}\right) \qquad (9)$$

Then we normalize the energy data using simple normalization,

By using all those formula, we work on some different procedures for preprocessing data:

1. Apply the wavelet transform to the single skeleton frame, averaging all the energy from one video (using the fig.2a preprocessing data). It is a high-cost computation.
2. Apply the wavelet transform to the skeleton frame sequence, (using the fig.2b preprocessing data). It is a high-cost computation.
3. Apply the wavelet transform to the single motion frame, averaging all the energy from one video

(using the fig.3a preprocessing data). It is a low-cost computation.

4. Apply the wavelet transform to the motion frame sequence, averaging all the energy from one video (using the fig.3.b preprocessing data). It is a low-cost computation.

3. Results

We show the experiment results in the following figure. They show the state for every combination and every procedure did. One point in every figure represents data of one person in one video. One color consists of 6 points in one scatter, which means that one person have 6 video dataset. All the procedures are using Haar Wavelet at 1-level decomposition. All the possible combination is following:

1. eAeH (energy from Approximation and Horizontal Detail),
2. eAeV (energy from Approximation and Vertical Detail),
3. eAeD (energy from Approximation and Diagonal Detail),
4. eHeV (energy from Horizontal and Vertical Detail),
5. eHeD (energy from Horizontal and Diagonal Detail),
6. and eVeD (energy from Vertical and Diagonal Detail).

Figure 5 shows all possible energy combination per frame with 2 persons. The energy combination eHeV shows the better cluster condition. It can separate data. In the other energy combination, there are many intersection data. It will be quite difficult to classify, even if we just using 2 persons data.

Figure 6 shows all possible energy combination per sequence with 2 persons. It will be easier to do the classification, except on the eHeD combination energy. It will be difficult if we add some data as shown in figure 7. In figure 7, we add the data 2 more persons, so the total data is 4 persons. This figure shows that there is no possible energy combination used for classification.

Figure 8 shows that the result is better than the result in figure 6 for 2 persons. All the combination energy looks possible to make the classification, but the best for all is the energy combination of eHeD. Figure 9, by using 4 person data, the best energy combination is eHeD. Figure 10, by using 4 person data, we can see that the best energy combination is eHeV. The result of eHeV (in figure 10) is even better than the eHeD (in figure 9).

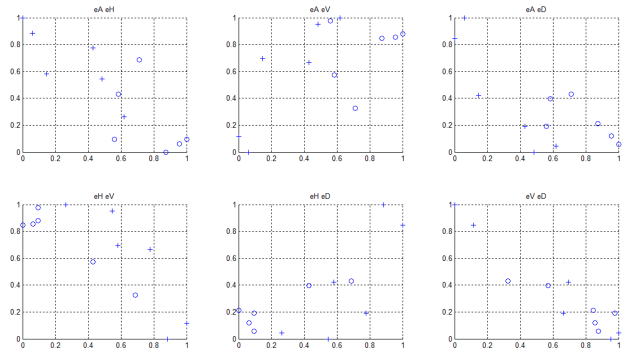

Figure 5. Result from single skeleton frame of 2 persons

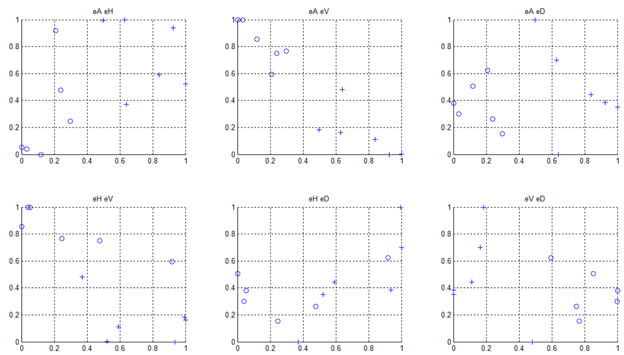

Figure 6. Result from skeleton frame sequence of 2 persons

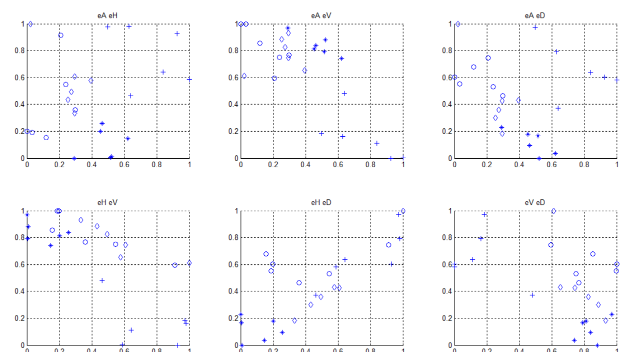

Figure 7. Result from skeleton frame sequence of 4 persons

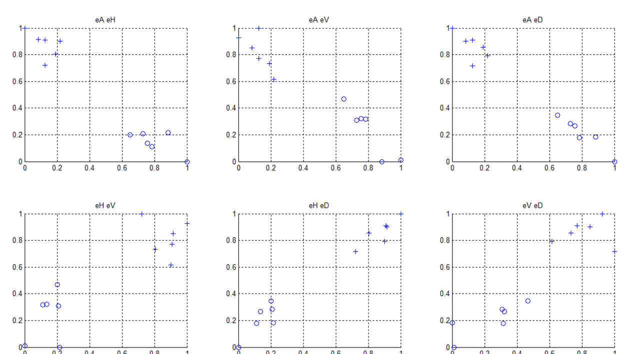

Figure 8. Result from single motion frame of 2 persons

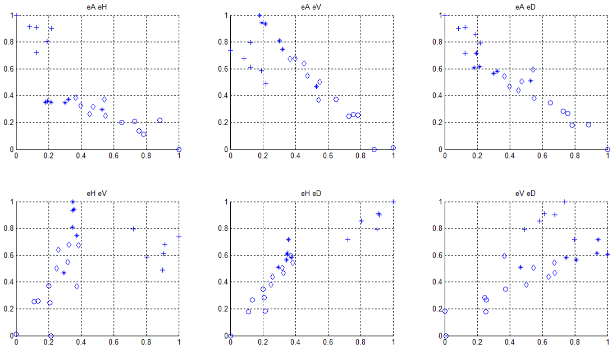

Figure 9. Result from single motion frame of 4 persons

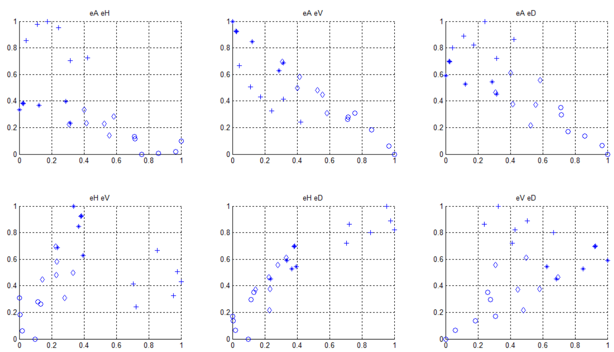

Figure 10. Result from motion frame sequence of 4 persons

4. Conclusion and Discussion

Using the experiment results, using 2D DWT, Haar Wavelet and 1 level decomposition, we conclude as following:

(1) The best preprocessing data is a motion in frame sequence.

(2) Best combination parameters for classification are Horizontal Detail and Vertical Detail.

(3) It needs more number of energy and decomposition to create human gait classification.

This research is the beginning process to identify person based on human gait. In this research we choose the small number of energy to create feature vector. It is not enough to express the state, and needs more number of decomposition. It is the easiest way to show effective energy on wavelets decomposition for human gait classification. In the next research, we will analyze the number of wavelets decomposition and create the classification module based on human gait. It is also recommended to use different feature extraction, such as analyzing of spatio temporal information.

Acknowledge

Portions of the research in this paper use the CASIA Gait Database collected by Institute of Automation, Chinese Academy of Sciences.

References

[1] X. Qinghan, “Technology review – Biometrics Technology, Application, Challenge, and Computational Intelligence Solutions”, IEEE Computational Intelligence Magazine, vol. 2, pp. 5-25, 2007.

[2] Jin Wang, Mary She, Saeid Nahavandi, Abbas Kouzani, “A Review of Vision-based Gait Recognition Methods for Human Identification”, IEEE Computer Society, 2010 International Conference on Digital Image Computing: Techniques and Applications, pp. 320 - 327, 2010

[3] N. V. Boulgouris, D. Hatzinakos, and K. N. Plataniotis, “Gait recognition: a challenging signal processing technology for biometric identification”, IEEE Signal Processing Magazine, vol. 22, pp. 78-90, 2005.

[4] M. S. Nixon and J. N. Carter, "Automatic Recognition by Gait", Proceedings of the IEEE, vol. 94, pp. 2013-2024, 2006.

[5] Y. Jang-Hee, H. Doosung, M. Ki-Young, and M. S. Nixon, “Automated Human Recognition by Gait using Neural Network”, in First Workshops on Image Processing Theory, Tools and Applications, 2008, pp. 1-6.

[6] Wilfrid Taylor Dempster, George R. L. Gaughran, “Properties of Body Segments Based on Size and Weight”, American Journal of Anatomy, Volume 120, Issue 1, pages 33–54, January 1967

[7] Gilbert Strang and Truong Nguen, “Wavelets and Filter Banks”. Wellesley-Cambridge Press, MA, 1997, pp. 174-220, 365-382

[8] I. Daubechies, *Ten lectures on wavelets*, Philadelphis, PA : SIAM, 1992.

[9] CASIA Gait Database, http://www.cbsr.ia.ac.cn/english/index.asp