

Activity and Individual Human Recognition in Infrared Imagery

Bir Bhanu and Ju Han

Center for Research in Intelligent Systems

University of California, Riverside, California 92521, USA

{bhanu, jhan}@cris.ucr.edu

ABSTRACT

In this chapter, we investigate repetitive human activity patterns and individual recognition in thermal infrared imagery, where human motion can be easily detected from the background regardless of the lighting conditions and colors of the human clothing and surfaces, and backgrounds. We employ an efficient spatio-temporal representation for human repetitive activity and individual recognition, which represents human motion sequence in a single image while preserving spatio-temporal characteristics. A statistical approach is used to extract features for activity and individual recognition. Experimental results show that the proposed approach achieves good performance for repetitive human activity and individual recognition.

Keywords

Thermal Infrared Imagery, Activity Recognition, Gait Energy Image, Individual Recognition, Principal Component Analysis, Multiple Discriminant Analysis.

1. INTRODUCTION

Repetitive human activity involves a regularly repeating sequence of motion events such as walking, running and jogging. Most existing human activity recognition approaches detect human motion in visible spectrum. However, it is very likely that some part of human body or clothing has similar colors as the background colors. In this case, human motion detection using currently available approaches usually fails on these parts of the human body. Moreover, the existence of shadows is a significant problem in the visible spectrum. In addition, sensors in the visible spectrum do not work under low lighting conditions such as night or indoor environments without any light.

To avoid the disadvantages of using sensors in the visible spectrum, we investigate the possibility of using the thermal infrared (**long wave infrared**) sensor for **human activity analysis** and **individual recognition**. Unlike a regular camera which records reflected visible light, a long wave (8~12 μ m) infrared camera records electromagnetic radiation emitted by objects in a scene as a thermal image whose pixel values represent temperature. In a thermal image that consists of humans in a scene, human silhouettes can be easily extracted from the background regardless of lighting conditions and colors of the human surfaces and backgrounds, because the temperatures

of the human body and background are different in most situations (Arlowe, 1992). Figure 1 shows an example of a walking human at different time of a day recorded using a thermal infrared sensor: noon (first row), late afternoon (second row) and night (third row). There are no obvious thermal shadows introduced in the thermal infrared images recorded at noon and late afternoon. Thermal images also have high contrast between the human and the background at night.

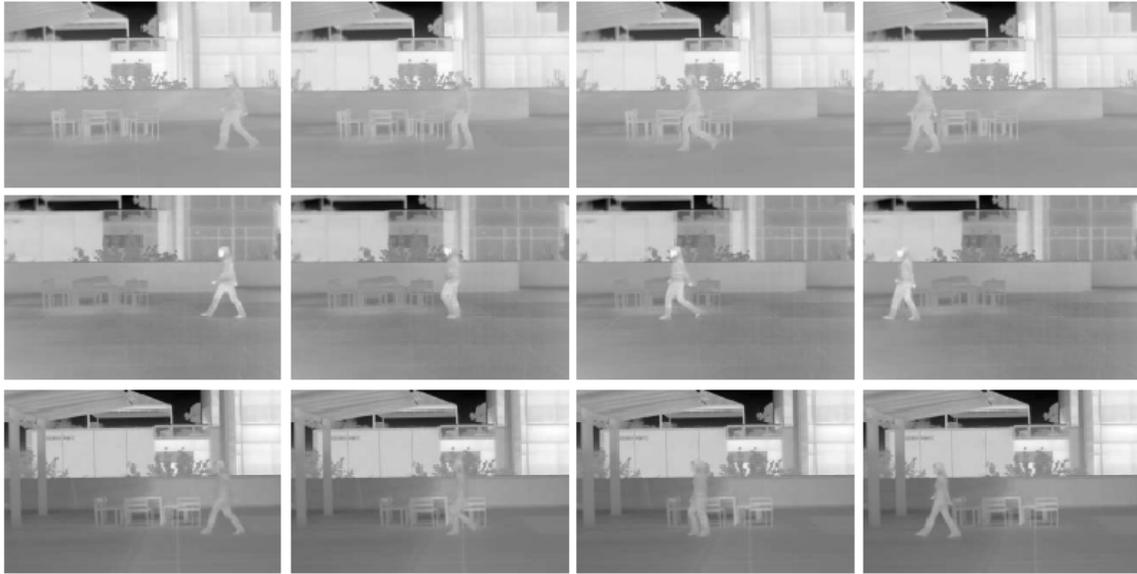


Figure 1. An example of human walking, at different times of a day, recorded using a thermal infrared sensor: noon (first row), late afternoon (second row) and night (third row).

In this chapter, we investigate repetitive human activities and individual recognition in thermal infrared imagery. First, human motion is detected and human silhouettes are extracted from the background. Then, we employ an efficient spatio-temporal representation, called **Gait Energy Image (GEI)**, for repetitive human activity and individual recognition. Unlike other representations which consider motion as a sequence of templates (poses), GEI represents human motion sequence in a single image while preserving essential spatio-temporal information. Finally, we use a statistical approach based on principal component analysis and multi-discriminant to extract features from GEI for activity and individual recognition.

In section 2, related work on object detection, activity and individual recognition is presented. Gait energy image based representation of human motion is discussed in detail in Section 3. Statistical approach for recognition of activities and individuals is presented in Section 4. Experimental results are presented in Section 5. Finally, conclusions of the chapter are provided in Section 6.

2. RELATED WORK

2.1 Object Detection

Over the years many different approaches have been proposed to detect moving objects in color imagery, including running Gaussian average (Wren et al., 1996), mixture of Gaussians (Stauffer & Grimson, 1999), kernel density estimation (Elgammal, Harwood & Davis, 2000), and Eigen-background (Oliver Rosario & Pentland, 2000), etc. These approaches vary in computational

complexity and their performance depends on the complexity of background variations. However, images from different kind of sensors generally have different pixel characteristics due to the phenomenological differences between the image formation processes of the sensors. Object detection in **thermal infrared imagery** has been widely used in automated object recognition in surveillance environments, especially for human detection. Andreone et. al (Andreone et. al, 2002) propose an approach for detecting vehicles in thermal infrared imagery. Initially the attention is focused on portions of the image that contain hot objects only. The result is further investigated exploiting specific vehicle thermal characteristics. Arlowe (Arlowe, 1992) develop an automatic detection systems based on the thermal contrast and motion of human intruders. The conditions and energy transfer mechanisms that lead to difficult thermal detection are discussed in his work. The heat flux balance equation can be used in an iterative computer program to predict the surface temperatures of both the background and the target for human intruder detection. Ginesu et. al (Ginesu et. al, 2004) propose a novel method to detect foreign bodies, which are not detectable using conventional methods, by inspecting food samples using thermographic images. Pavlidis et. Al (Pavlidis, Levine & Baukol, 2000) propose a method for detecting suspects engaged in illegal and potentially harmful activities in or around critical military or civilian installations. Thermal image analysis is performed to detect facial patterns of anxiety, alertness, and/or fearfulness at a distance. Bhanu and Han (Bhanu & Han, 2002) investigated the use of kinematic models for analyzing human motion in infrared video. Nadimi and Bhanu (Nadimi & Bhanu, 2004) investigate phenomenology based models for object detection by fusing color and infrared video. Han and Bhanu (Han & Bhanu, 2007) develop registration techniques for human detection in color and infrared videos. Yoshitomi et. al (Yoshitomi et al., 1997) develop a face identification approach based on thermal image analysis. The front-view face is first normalized in terms of location and size, the temperature distribution is then measured as well as the locally averaged temperature and the shape factors of face. These features are used for supervised classification by neural network.

2.2 Human Activity Recognition

In recent years, various approaches have been proposed for **human activity recognition**. These approaches generally fall under two major categories: model-based approaches and model-free approaches.

When people observe human motion patterns, they not only observe the global motion properties, but also interpret the structure of the human body and detect the motion patterns of local body parts. The structure of the human body is generally interpreted based on their prior knowledge. Model-based activity recognition approaches focus on recovering a structural model of human motion, and the motion patterns are then generated from the model parameters for activity recognition. Guo & Tsuji (Guo & Tsuji, 1994) represent the human body structure in the silhouette by a stick figure model. The human motion characterized by a sequence of the stick figure parameters are used as input of a neural network for classification. Fujiyoshi and Lipton (Fujiyoshi & Lipson, 1998) analyze the human motion by producing a star skeleton determined by extreme point estimation from the silhouette boundaries extracted. These cues are used to recognize human activities such as walking or running. Sappa et. al (Sappa et al., 2000) develop a technique for human motion recognition based on the study of feature points' trajectories. Peaks and valleys of points' trajectories are first detected to classify human activity using prior knowledge of human body kinematics structure together with the corresponding motion model. In model-based approaches, the accuracy of human model reconstruction strongly depends on the

quality of the extracted human silhouette. In the presence of noise or low-resolution data, the estimated parameters may not be reliable.

Model-free approaches make no attempt to recover a structural model of human motion. Polana and Nelson (Polana & Nelson, 1994) analyze human repetitive motion activity based on bottom up processing, which does not require the prior identification of specific parts. Motion activity is recognized by matching against a spatiotemporal template of motion features. Rajagopalan and Chellappa (Rajagopalan & Chellappa, 2000) develop a higher order spectral analysis-based approach for detecting people by recognizing repetitive motion activity. The stride length is determined in every frame, and the bi-spectrum which is the Fourier transform of the triple correlation is used for recognition. Sarkar and Vega (Sarkar & Vega, 2001) discriminate between motion types based on the change in the relational statistics among the detected image features. They use the distribution of the statistics of the relations among the features for recognition. Davis (Davis, 2004) proposes a probabilistic reliable-inference framework to address the issue of rapid and-reliable detection of human activities using posterior class ratios to verify the saliency of an input before committing to any activity classification.

2.3 Individual Human Recognition

Various techniques have been proposed for human recognition by gait. Like the classification of human activities these techniques can be divided as model-based and model-free approaches. Little and Boyd (Little & Boyd, 1998) describe the shape of the human motion with scale-independent features from moments of the dense optical flow, and recognize individuals by phase vectors estimated from the feature sequences. Sundaresan et al. (Sundaresan, RoyChowdhury & Chellappa, 2003) proposed a hidden Markov models (HMMs) based framework for individual recognition by gait. Huang et al. (Huang, Harris & Nixon, 1999) extend the template matching method to gait recognition by combining transformation based on canonical analysis and eigenspace transformation for feature selection. Sarkar et al. (Sarkar et al., 2005) directly measure the similarity between the gallery sequence and the probe sequence by computing the correlation of corresponding time-normalized frame pairs. Collins et al. (Collins, Gross & Shi, 2002) first extract key frames from a sequence and then the similarity between two sequences is computed using the normalized correlation.

3. REPRESENTATION OF HUMAN MOTION

Repetitive human activity is a cyclic motion where human motion repeats at a stable frequency (Cutler & Davis, 2000). Assuming that the order of poses in a specific repetitive activity is the same among different people, it is possible to compose a spatio-temporal template in a single image instead of an ordered image sequence as usual. The fundamental assumptions made here are: (a) the order of poses in different cycles is the same, i.e., limbs move forward and backward in a similar way among normal people; (b) differences exist in the phase of poses in a motion cycle, the extend of limbs, and the shape of the torso, etc. In this chapter, we propose to use a silhouette based **human motion representation**, Gait Energy Image (GEI), for both activity recognition and individual recognition. Silhouette-based representation may be dependent on certain environmental conditions, such as camera views and walking surfaces. Model-based human motion representation may partially overcome this limitation with increased computational complexity. An alternative way to solve this problem is to synthesize normalized

silhouette (e.g., synthesize side-view silhouette from other views (Kale, Chowdhury & Chellappa, 2003)), or generate synthetic silhouettes that simulate different environmental conditions and summarize invariant gait features from them (Han & Bhanu, 2006).

3.1 Human Silhouette Extraction

In this chapter, we suppose that the data have been divided into short image sequences each of which only contains one individual and one activity. It is reasonable to assume the background variation is very small in such a short period (typically less than one minute). Here, we use a simple running Gaussian average (Wren et al., 1996) method for silhouette extraction.

The raw silhouettes are further processed by size normalization (proportionally resizing each silhouette image so that all silhouettes have the same height) and horizontal alignment (centering the upper half silhouette part with respect to its horizontal centroid). In a so-obtained silhouette sequence, the time series signal of lower half silhouette part size from each frame indicates the motion frequency and phase information. The obtained time series signal consists of few cycles and lots of noise, which lead to sidelobe effect in the Fourier spectrum. To avoid this problem, we estimate the motion frequency and phase by maximum entropy spectrum estimation (Little & Boyd, 1998) from the obtained time series signal as shown in Figure 2.

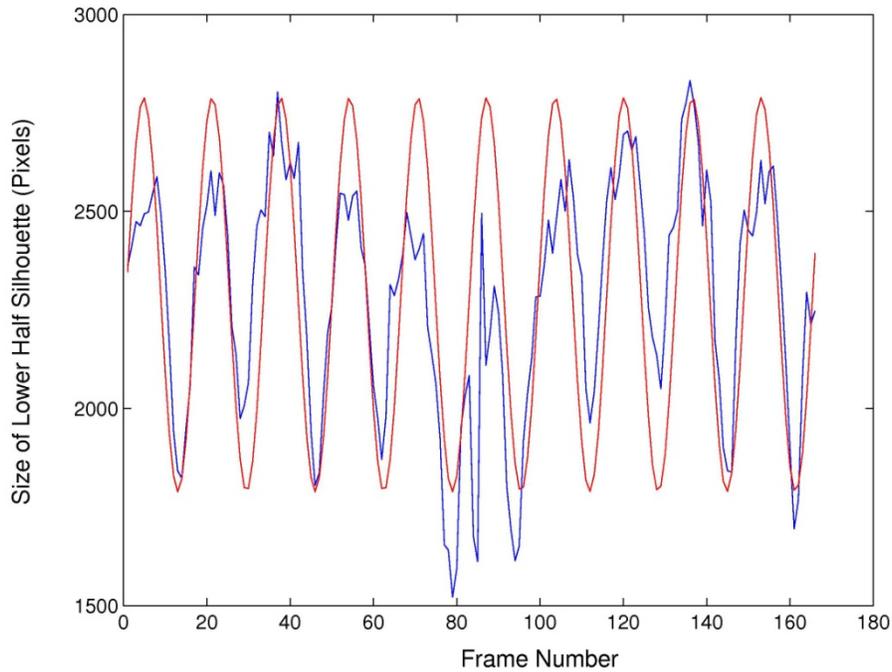


Figure 2. The time series signal of lower half silhouette size for different frames of walking humans.

3.2 Gait Energy Image

Given the preprocessed binary human silhouette images $B_t(x, y)$ at time t in a sequence, the grey-level gait energy image (GEI) is defined as follows

$$G(x, y) = \frac{1}{N} \sum_{t=1}^N B_t(x, y) \quad (1)$$

where N is the number of frames in the complete cycle(s) of a silhouette sequence, t is the frame number in the sequence (moment of time), x and y are values in the 2D image coordinate. Figure 3 shows the sample silhouette images in a motion cycle of different activities of the same person, and the right most image is the corresponding GEI. As expected, GEI reflects major shapes of silhouettes and their changes over the motion cycle. We refer to it as gait energy image because: (a) each silhouette image is the space-normalized energy image of human walking at this moment; (b) GEI is the time-normalized accumulative energy image of human walking in the complete cycle(s); (c) a pixel with higher intensity value in GEI means that human walking occurs more frequently at this position (i.e., with higher energy).

3.2.1 GEI Properties

In comparison with the activity representation by binary silhouette sequence, GEI representation saves both storage space and computation time for recognition and is less sensitive to silhouette noise in individual frames. Consider a noisy silhouette image $B_t(x, y)$ that is formed by the addition of noise $\eta_t(x, y)$ to an original silhouette image $f_t(x, y)$, that is, $B_t(x, y) = f_t(x, y) + \eta_t(x, y)$, where we assume that at every pair of coordinates (x, y) the noise at different moments t is uncorrelated and identically distributed. Under these constraints, we further assume that $\eta_t(x, y)$ satisfies the following distribution:

$$\eta_t(x, y) = \begin{cases} \eta_{1t}(x, y), & \text{if } f_t(x, y) = 1 \\ \eta_{2t}(x, y), & \text{if } f_t(x, y) = 0 \end{cases} \quad (2)$$

where

$$\begin{aligned} \eta_{1t}(x, y): P\{\eta_t(x, y) = -1\} &= p, P\{\eta_t(x, y) = 0\} = 1 - p \\ \eta_{2t}(x, y): P\{\eta_t(x, y) = 1\} &= p, P\{\eta_t(x, y) = 0\} = 1 - p. \end{aligned}$$

we have

$$E\{\eta_t(x, y)\} = \begin{cases} -p, & \text{if } f_t(x, y) = 1 \\ p, & \text{if } f_t(x, y) = 0 \end{cases} \quad (3)$$

and

$$\sigma_{\eta_t(x, y)}^2 = \sigma_{\eta_{1t}(x, y)}^2 = \sigma_{\eta_{2t}(x, y)}^2 = p(1 - p) \quad (4)$$

Given a walking cycle with N frames where $f_t(x, y) = 1$ at a pixel (x, y) only in M frames, we have

$$G(x, y) = \frac{1}{N} \sum_{t=1}^N B_t(x, y) = \frac{1}{N} \sum_{t=1}^N f_t(x, y) + \frac{1}{N} \sum_{t=1}^N \eta_t(x, y) = \frac{M}{N} + \bar{\eta}(x, y).$$

Therefore, the noise in GEI is

$$\begin{aligned} E\{\bar{\eta}(x, y)\} &= \frac{1}{N} \left[\sum_{t=1}^M E\{\eta_{1t}(x, y)\} + \sum_{t=M+1}^N E\{\eta_{2t}(x, y)\} \right] \\ &= \frac{1}{N} [M(-p) + (N - M)p] = \frac{N - 2M}{N} p \end{aligned}$$

and

$$\begin{aligned} \sigma_{\bar{\eta}(x, y)}^2 &= E\{[\bar{\eta}(x, y) - E\{\bar{\eta}(x, y)\}]^2\} \\ &= \frac{1}{N^2} E \left\{ \left[\sum_{t=1}^M [\eta_{1t}(x, y) - E\{\eta_{1t}(x, y)\}] + \sum_{t=M+1}^N [\eta_{2t}(x, y) - E\{\eta_{2t}(x, y)\}] \right]^2 \right\} \\ &= \frac{1}{N^2} [M\sigma_{\eta_{1t}(x, y)}^2 + (N - M)\sigma_{\eta_{2t}(x, y)}^2] = \frac{1}{N^2} \sigma_{\eta_t(x, y)}^2 \end{aligned}$$

Therefore, the mean of the noise in GEI varies between $-p$ and p depending on M while its variability ($\sigma_{\bar{\eta}(x,y)}^2$) decreases. If $M = N$ at (x, y) (all $f_t(x, y) = 1$), $E\{\bar{\eta}(x, y)\}$ becomes $-p$; if $M = 0$ at (x, y) (all $f_t(x, y) = 0$), $E\{\bar{\eta}(x, y)\}$ becomes p . At the location (x, y) , the mean of the noise in GEI is the same as that in the individual silhouette image, but the noise variance reduces so that the probability of outliers is reduced. If M varies between 0 and N at (x, y) , $E\{\bar{\eta}(x, y)\}$ also varies between p and $-p$. Therefore, both the mean and the variance of the noise in GEI are reduced compared to the individual silhouette image at these locations. At the extreme, the noise in GEI has zero mean and reduced variance where $M = N=2$. As a result, GEI is less sensitive to silhouette noise in individual frames.

3.3 Comparison to MEI and MHI

Bobick and Davis (Bobick & Davis, 2001) propose **motion-energy image (MEI)** and **motion-history image (MHI)** for human movement type representation and recognition.

MEI is a binary image which represents where motion has occurred in an image sequence:

$$E_{\tau}(x, y, t) = \bigcup_{i=0}^{\tau-1} D(x, y, t - i) \quad (5)$$

where $D(x, y, t)$ is a binary sequence indicating regions of motion, τ is the duration of time, t is the moment of time, x and y are values of 2D image coordinates.

MHI is a grey-level image which represents how motion in the image is moving:

$$H_{\tau}(x, y, t) = \begin{cases} \tau, & \text{if } D(x, y, t) = 1; \\ \max\{0, H_{\tau}(x, y, t - 1) - 1\}, & \text{otherwise.} \end{cases} \quad (6)$$

Both MEI and MHI are vector-images where the vector value at each pixel is a function of the motion properties at this location in an image sequence. As compared to MEI and MHI, GEI targets the representation of repetitive activity, which has been successfully used in human identification by gait (Han & Bhanu, 2006).

4. ACTIVITY AND INDIVIDUAL RECOGNITION

4.1 Repetitive Activity Recognition

In this section, we describe the proposed repetitive activity recognition approach using gait energy image. In the training procedure, GEI templates are generated from the original silhouette sequences. A component and discriminant analysis is then performed on the training templates for feature extraction. Human activity recognition is based on the extracted features.

Given a series of training GEI templates for each activity, the problem of their excessive dimensionality occurs. To reduce their dimensionality, there are two classical approaches of finding effective linear transformations by combining features - **Principal Component Analysis (PCA)** and **Multiple Discriminant Analysis (MDA)**. As described in (Duda, Hart & Stork, 2000), PCA seeks a projection that best represents the data in a least square sense, while MDA seeks a projection that best separates the data in a least-square sense. We combine PCA and MDA to achieve the best data representation and the best class separability simultaneously (Huang, Harris & Nixon, 1999).

Given n d -dimensional training templates $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, PCA minimizes the criterion function

$$J_{d'} = \sum_{k=1}^n \left\| \left(\mathbf{m} + \sum_{i=1}^{d'} a_{ki} \mathbf{e}_i \right) - \mathbf{x}_k \right\|^2 \quad (7)$$

where $d' < d$, $\mathbf{m} = \frac{1}{N} \sum_{k=1}^n \mathbf{x}_k$, and $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{d'}\}$ are a set of unit vectors. $J_{d'}$ is minimized when $\mathbf{e}_1, \mathbf{e}_2, \dots$, and $\mathbf{e}_{d'}$ are the d' eigenvectors of the scatter matrix S having the largest eigenvalues, where

$$S = \sum_{k=1}^n (\mathbf{x}_k - \mathbf{m})(\mathbf{x}_k - \mathbf{m})^T \quad (8)$$

The d' -dimensional principal component vector \mathbf{y}_k is obtained from the d -dimensional GEI template \mathbf{x}_k by multiplying the transformation matrix $[\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{d'}]$:

$$\mathbf{y}_k = [a_1, a_2, \dots, a_{d'}]^T = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{d'}]^T \mathbf{x}_k, \quad k = 1, \dots, n \quad (9)$$

where n is the number of the expanded GEI templates from all people in the training dataset.

Although PCA finds components that are useful for representing data, there is no reason to assume that these components must be useful for discriminating between data in different classes because PCA does not consider the class label of training templates. Multiple discriminant analysis (MDA) seeks a projection that is efficient for discrimination. Suppose that the n d_0 -dimensional transformed training templates $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$, belong to c classes. MDA seeks a transformation matrix W that maximizes the ratio of the between-class scatter S_B to the within-class scatter S_W :

$$J(W) = \frac{|S_B|}{|S_W|} = \frac{|W^T S_B W|}{|W^T S_W W|} \quad (10)$$

The within-class scatter S_B is defined as

$$S_W = \sum_{i=1}^c S_i \quad (11)$$

where

$$S_i = \sum_{\mathbf{y} \in D_i} (\mathbf{y} - \mathbf{m}_i)(\mathbf{y} - \mathbf{m}_i)^T \quad (12)$$

and

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{\mathbf{y} \in D_i} \mathbf{y} \quad (13)$$

where D_i is the training template set that belongs to the i th class and n_i is the number of templates in D_i . The within-class scatter S_B is defined as

$$S_B = \sum_{\mathbf{y} \in D_i} n_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \quad (14)$$

where

$$\mathbf{m} = \frac{1}{n} \sum_{\mathbf{y} \in D} \mathbf{y} \quad (15)$$

and D is the whole training template set. $J(W)$ is maximized when the columns of W are the generalized eigenvectors that correspond to the largest eigenvalues in

$$S_B \mathbf{w}_i = \lambda_i S_W \mathbf{w}_i \quad (16)$$

There are no more than $c-1$ nonzero eigenvalues, and the corresponding eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_{c-1}$ form transformation matrix. The $(c-1)$ -dimensional multiple discriminant vector \mathbf{z}_k is obtained

from the d' -dimensional principal component vector \mathbf{y}_k by multiplying the transformation matrix $[\mathbf{v}_1, \dots, \mathbf{v}_{c-1}]$:

$$\mathbf{z}_k = [\mathbf{v}_1, \dots, \mathbf{v}_{c-1}]^T \mathbf{y}_k, \quad k = 1, \dots, n \quad (12)$$

The obtained multiple discriminant vectors compose the feature database for activity recognition. Assuming that the obtained feature vectors in each class are Gaussian distributed with the same covariance matrix, Bayesian classifier becomes minimum Euclidean distance classifier that is used for activity recognition.

4.2 Individual Recognition

The same framework as described in Section 4.1 is used for individual recognition during each of the activities (walking and running). In this case the training data consists of all the individuals for a given activity under a given **contextual conditions** (noon, late afternoon, night) and the testing is performed on the data acquired under different contextual conditions.

5. EXPERIMENTAL RESULTS

We have recorded real thermal image data of human activities by a FLIR SC2000 long-wave infrared camera in an outdoor environment. The image size is 240*320. The field-of-view of the camera is fixed during a human walking. Repetitive activities of five people are recorded at different time: noon (four people), late afternoon (three people) and night (two people). Each person was asked to slow walk, fast walk and run forward and backward along the fronto-parallel direction at each time. The corresponding backgrounds are shown in Figure 4. Each background is normalized by the temperature range individually.

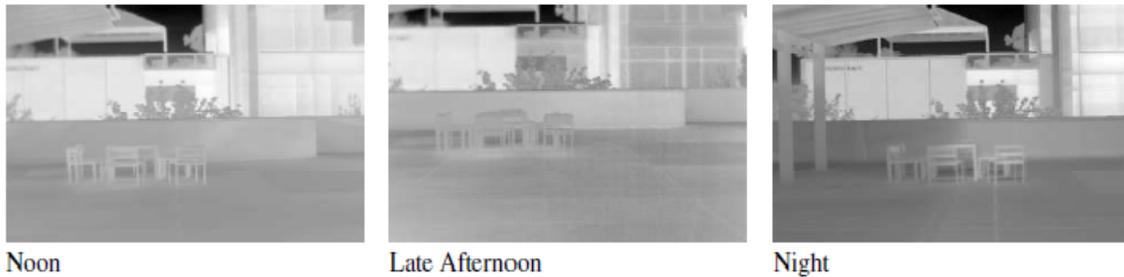


Figure 4. Thermal images of the background at different times of a day, recorded using a thermal infrared sensor: noon, late afternoon and night. Each image is normalized by its temperature range individually.

For activity recognition three data sets recorded at noon are used for training, and other data sets are used for testing. Figure 5 shows GEI Examples of the 9 data sets (54 human motion sequences) used in our experiments. An observation from this figure is that the silhouette extraction performance at late afternoon is better than that at noon and night. This means that the temperature contrast between the human object and the background is larger at late afternoon. The motion of trees in the background also contributes to the silhouette extraction performance in some frames.

The goal of our activity recognition here is to discriminate human walking or running regardless of their speed (slow or fast walking). In the recorded data, the speed in some fast walking sequences is equivalent or faster than that in some training sequences. Therefore, the speed is not appropriate for recognition of activity (walking or running). We employ the approach

of combining PCA and MDA for **feature extraction**. The four-dimensional vectors obtained by PCA are used as the input of the MDA, and the final feature vectors are of one dimension. The recognition performance on training data and testing data are all 100%. This demonstrates that the proposed approach achieves good performance for human repetitive activity recognition in the limited dataset.

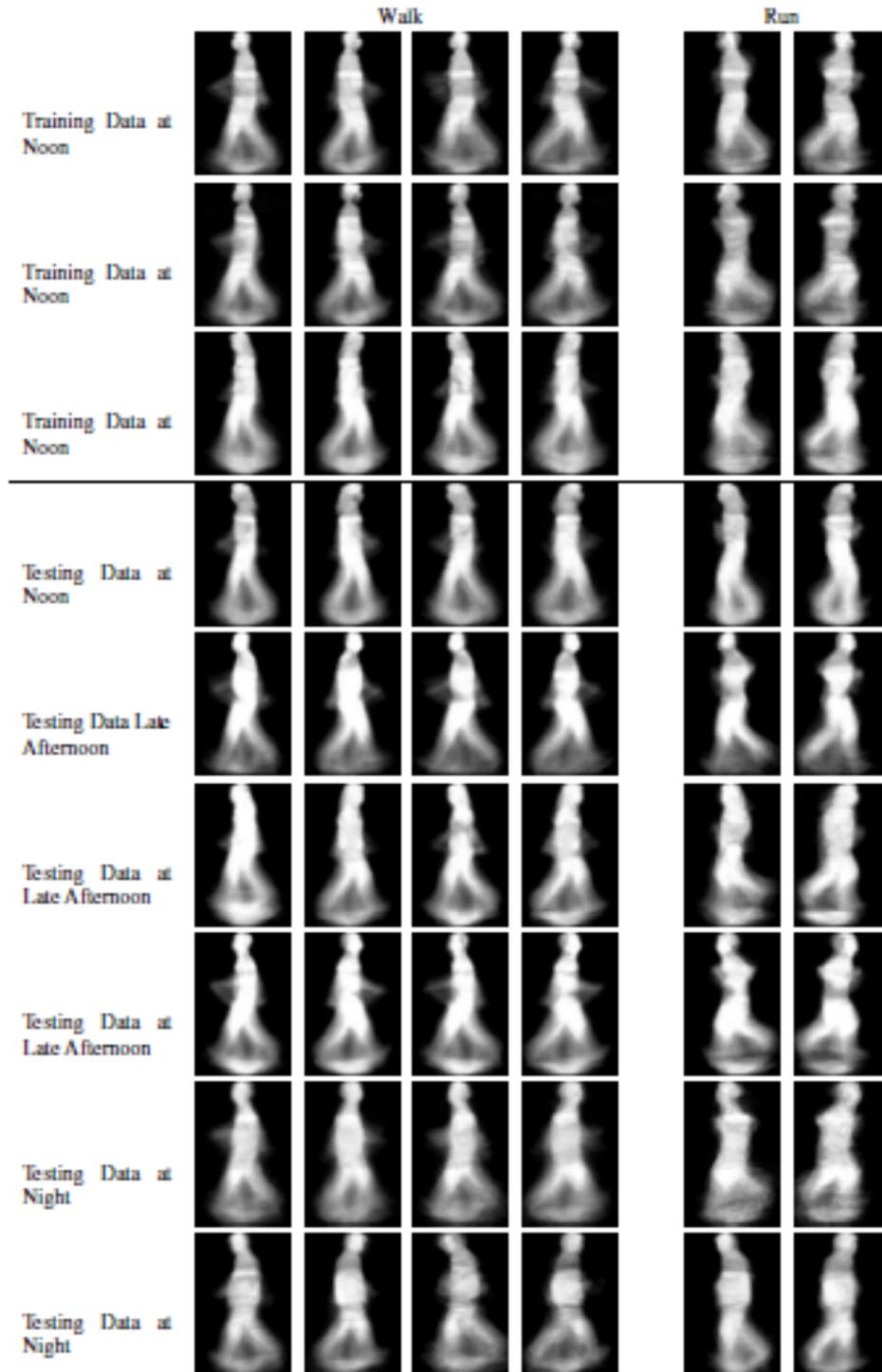


Figure 5. GEI Examples of the 54 human motion sequences used in our experiments. Training and testing data shown in the figure correspond to activity recognition experiments.

We do the recognition of only four (first 4 rows in Figure 5) individuals. Individuals are recognized for the same activity and the training/testing data are appropriately chosen for the same/different contextual conditions. The recognition of these individuals is quite simple as evidenced by different GEI images. Elsewhere we have done an extensive evaluation and comparison of individual recognition of 122 individuals in color video imagery on DARPA HumanID Database and the results are briefly outlined in Table 1 (Han & Bhanu, 2006). Gait features generated from original GEI templates in a gait sequence may be dependent on certain environmental conditions. In [13], synthetic GEI samples were generated to simulate certain changes in environmental conditions, and the gait features learned from original and synthetic GEI samples were combined to improve the individual recognition performance. These results demonstrate that our approach is effective in recognizing individuals in complex environmental conditions with color imagery, and, therefore, expected to be extensible to thermal imagery. Furthermore, the proposed GEI-base individual recognition framework is extensible to address certain invariance problems by synthesizing invariant GEI templates. For example, side-view GEI templates synthesized from various views (Kale, Chowdhury & Chellappa, 2003) may lead to view-insensitive gait features, and improve the individual recognition performance in certain circumstances.

Table 1. Comparison of Recognition Performance Using Different Approaches on USF HumanID gait database (Sarkar et al., 2005).

dataset	A	B	C	D	E	F	G
USF (Sarkar et al., 2005)	79%	66%	56%	29%	24%	30%	10%
CMU (Collins, Gross & Shi, 2002)	87%	81%	66%	21%	19%	27%	23%
UMD (Sundaresan, RoyChowdhury & Chellappa, 2003)	99%	89%	78%	36%	29%	24%	18%
GEI (Han & Bhanu, 2006)	100%	90%	85%	47%	57%	32%	31%

6. CONCLUSIONS

In this chapter, we use a spatio-temporal gait representation, called the Gait Energy Image (GEI), for human repetitive activity recognition. Human motion is detected in thermal infrared imagery, which provides good contrast between human objects and backgrounds regardless of lighting conditions and colors of the human surfaces and backgrounds. Unlike other motion representations which consider gait as a sequence of templates (poses), GEI represents human motion sequence in a single image while preserving temporal information. A statistical approach is used to extract features from GEI for activity and individual recognition. The proposed GEI-based recognition method could be further extended by introducing synthetic GEI templates to address certain invariance problems (e.g., pose invariance). Experimental results on 5 people and 54 sequences of infrared data and extensive color video data show that the proposed approach achieves good performance for human repetitive activity and individual recognition. The performance of the approach will be further explored on a larger thermal infrared dataset which is under construction.

REFERENCES

- Andreone, L., Antonello, P., Bertozzi, M., Broggi, A., Fascioli, A., & Ranzato, D. (2002) Vehicle detection and localization in infrared images. *Proc. IEEE International Conference on Intelligent Transportation Systems*, 141-146.
- Arlowe, H. (1992) Thermal detection contrast of human targets. *Proc. IEEE International Carnahan Conference on Security Technology*, 27-33.
- Bhanu, B., & Han, J. (2002) Kinematic-based motion analysis in infrared sequences. *Proc. IEEE Workshop on Applications of Computer Vision*, 208-212.
- Bobick, A., & Davis, J. (2001) The recognition of human movement using temporal templates. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(3):257-267.
- Collins, R., Gross, R., & Shi, J. (2002) Silhouette-based human identification from body shape and gait. *Proc. IEEE Intl. Conf. on Automatic Face and Gesture Recognition*, 351-356.
- Cutler, R., & Davis, L. (2000) Robust real-time periodic motion detection, analysis and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):781-796.
- Davis, J. (2004) Sequential reliable-inference for rapid detection of human actions. *Proc. Conference on Computer Vision and Pattern Recognition*, 111-118.
- Duda, R., Hart, P., & Stork, D. (2000) *Pattern Classification*. John Willy & Sons.
- Elgammal, A., Harwood, D., & Davis, L. (2000) Non-parametric model for background subtraction. *Proc. European Conf. on Computer Vision*, II:751-767.
- Fujiyoshi, H., & Lipson, A. (1998) Real-time human motion analysis by image skeletonization. *Proc. 4th IEEE Workshop on Applications of Computer Vision*, 15-21.
- Ginesu, G., Giusto, D., Margner, V., & Meinschmidt, P. (2004) Detection of foreign bodies in food by thermal image processing. *IEEE Trans. Industrial Electronics*, 51(2):480-490.
- Guo, Y., & Tsuji, S. (1994) Understanding human motion patterns. *Proc. International Conference on Pattern Recognition*, 2:325-329.
- Han, J., & Bhanu, B. (2006) Individual recognition using gait energy image. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(2):316-322.
- Han, J., & Bhanu, B. (2007) Fusion of color and infrared video for moving human detection. *Pattern Recognition*, 40(6):1771-1784.
- Huang, P., Harris, C., & Nixon, M. (1999) Recognizing humans by gait via parameteric canonical space. *Artificial Intelligence in Engineering*, 13:359-366.
- Kale, A., Chowdhury, A., & Chellappa, R. (2003) Towards a view invariant gait recognition algorithm. *Proc. of IEEE Conference on Advanced Video and Signal Based Surveillance*, 143-150.
- Little, J., & Boyd, J. (1998) Recognizing people by their gait: the shape of motion. *Videre: Journal of Computer Vision Research*, 1(2):1-32.

- Nadimi, S., & Bhanu, B. (2004) Physical models for moving shadow and object detection in video. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(8):1079-1087.
- Oliver, N., Rosario, B., & Pentland, A.P. (2000) A bayesian computer vision system for modeling human interactions. *IEEE Trans. on Patt. Anal. and Machine Intell.*, 22(8):831-843.
- Pavlidis, I., Levine, J., & Baukol, P. (2000) Thermal imaging for anxiety detection. *Proc. IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, 104-109.
- Polana, R., & Nelson, R. (1994) Low level recognition of human motion (or how to get your man without finding his body parts). *Proc. IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, 77-82.
- Rajagopalan, A., & Chellappa, R. (2000) Higher-order spectral analysis of human motion. *Proc. International Conference on Image Processing*, 3:230-233.
- Sappa, A., Aifanti, N., Malassiotis, S., & Strintzis, M. (2000) Unsupervised motion classification by means of efficient feature selection and tracking. *Proc. International Symposium on 3D Data Processing, Visualization and Transmission*, 912-917.
- Sarkar, S., Phillips, P., Liu, Z., Vega, I., Grother, P., & Bowyer, K. (2005) The HumanID gait challenge problem: Datasets, performance and analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(2):162-177.
- Sarkar, S., & Vega, I. (2001) Discrimination of motion based on traces in the space of probability functions over feature relations. *Proc. IEEE International Computer Society Conference Computer Vision and Pattern Recognition*, 1:976-983.
- Stauffer C., & Grimson, W. (1999) Adaptive background mixture models for real-time tracking. *Proc. Computer Vision and Pattern Recognition*, 246-252.
- Sundaresan, A., RoyChowdhury, A., & Chellappa R. (2003) A hidden Markov model based framework for recognition of humans from gait sequences. *Proc. ICIP*, 2:93-96.
- Wren, C., Azarbayejani, A., Darrell, T., & Pentland, A. (1996) Pfunder: real-time tracking of the human body. *Proc. IEEE Conference on Automatic Face and Gesture Recognition*, 51-56.
- Yoshitomi, Y., Miyaura, T., Tomita, S., & Kimura, S. (1997) Face identification using thermal image processing. *Proc. IEEE International Workshop on Robot and Human Communication*, pages 374-379.