

Research Article

Efficient Human Action and Gait Analysis Using Multiresolution Motion Energy Histogram

Chih-Chang Yu,¹ Hsu-Yung Cheng,² Chien-Hung Cheng,² and Kuo-Chin Fan^{2,3}

¹Department of Computer Science and Information Engineering, Vanung University, Chung-Li 32061, Taiwan

²Department of Computer Science and Information Engineering, National Central University, Chung-Li 32001, Taiwan

³Department of Informatics, Fo-Guang University, I-Lan 26247, Taiwan

Correspondence should be addressed to Chih-Chang Yu, tacoyu@mail.vnu.edu.tw

Received 29 November 2009; Accepted 10 February 2010

Academic Editor: Jenq-Neng Hwang

Copyright © 2010 Chih-Chang Yu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Average Motion Energy (AME) image is a good way to describe human motions. However, it has to face the computation efficiency problem with the increasing number of database templates. In this paper, we propose a histogram-based approach to improve the computation efficiency. We convert the human action/gait recognition problem to a histogram matching problem. In order to speed up the recognition process, we adopt a multiresolution structure on the Motion Energy Histogram (MEH). To utilize the multiresolution structure more efficiently, we propose an automated uneven partitioning method which is achieved by utilizing the quadtree decomposition results of MEH. In that case, the computation time is only relevant to the number of partitioned histogram bins, which is much less than the AME method. Two applications, action recognition and gait classification, are conducted in the experiments to demonstrate the feasibility and validity of the proposed approach.

1. Introduction

Analyzing human's behavior or identity is a very interesting research topic because human is usually the most concerned object in many applications such as surveillance system or video understanding. Recently, this problem is usually solved by two kinds of approaches: video-based approaches or sensor-based approaches [1, 2]. The advantage of video-based approach is that the individuals do not have to put on additional devices and the hardware cost is also cheaper. For video-based approaches, there exists abundant considerable works made by previous researchers such as employing template matching [3, 4], Intensity-based features [5, 6], shape matching [7], and spatial-temporal features [8–13]. For spatial-temporal features, motion energy images (MEIs) are a very useful feature which incorporates temporal information into spatial images. The idea of MEI was firstly introduced by Bobick and Davis in [14]. The authors obtained the MEI by collecting a group of frames and extract scale invariant features for recognition. This idea was extended to the so called average motion energy (AME) by aligning and

normalizing the foreground silhouettes [15]. By doing so, the AME can depict human's motion in a two-dimensional space (the spatial domain) while preserving the temporal motion information. Unlike other approaches in [10], the process of generating AME is computationally inexpensive and can be employed in real-time applications. Besides, it had been proven that AME can provide reliable accuracy for action recognition by other researchers [16]. Recently, some researchers focus on more challenging problems such as gender classification [17, 18] and gait classification [19–24]. The AME idea was also employed in [22], in which they named AME as gait energy images (GEIs).

In [15], the authors directly employed the sum of absolute difference (SAD) for action recognition purpose and obtain adequate recognition results. However, the computation of SAD is inefficient when the size of image is large because the computational time is relevant to the image size. This problem is not severe when the amount of database images is small but can be expected with the increasing size of image database. To remedy this problem, we propose a histogram-based approach which can efficiently compute

the similarity among patterns. Firstly, the AME image is converted to the motion energy histogram (MEH). Then, we adopt a multiresolution structure to construct the multiresolution motion energy histogram (MRMEH). Last, we propose an uneven partitioning method to address the important part of MEH automatically and apply an efficient histogram matching algorithm by utilizing the characteristic of multiresolution histogram.

2. Histogram-Based Method

In this section, we will introduce the proposed histogram-based method for human action recognition. Section 2.1 describes the overview of the proposed framework. How to generate the AME and MEH is addressed in Sections 2.2 and 2.3, respectively. Then we will briefly describe the characteristic of MEH in Section 2.4. The construction and utilization of MRMEH are elaborated in Sections 2.5 and 2.6, followed by the time complexity analysis of the proposed method in Section 2.7.

2.1. System Overview. Figure 1 shows the overview of the system. Firstly, we extract the motion period from the input video. We can obtain the Average Motion Energy image based on the extracted motion period and convert it to the Motion Energy Histogram (MEH). Then we perform quadtree decomposition on the MEH to construct the Multiresolution Motion Energy Histogram (MRMEH). Finally we apply an efficient histogram matching algorithm on MRMEH to obtain the recognition result.

2.2. Average Motion Energy (AME). Before elaborating on the proposed histogram-based method, the average motion energy (AME) image proposed by Wang and Suter [15] is first introduced. Given a set of aligned human silhouettes $S(x, y, t)$, the AME can be defined as follows:

$$\text{AME}(x, y) = \frac{1}{T} \sum_{t=1}^T S(x, y, t), \quad (1)$$

where T is the number of frames. In their work, T is set as the motion period of the input action. There have been plenty of studies for periodicity detection of motions. For example, in [20], the authors adopted a simple strategy to extract the motion cycle. They observed the number of foreground pixels in the silhouette in each frame over time, $N_f(t)$. Afterwards the motion period can be extracted by seeking the local maximums of $N_f(t)$. This strategy works fair in the walking motion. However, for some other periodic motions such as waving hands or jumping, the size of the foreground silhouette does not change much. Figure 2 is an example which shows the variation of $N_f(t)$ of two motions, where Figure 2(a) represents the $N_f(t)$ of motion “walking” and Figure 2(b) represents the $N_f(t)$ of motion “waving one hand”, respectively. We can see that there are obvious local maximums and local minimums in $N_f(t)$ in the walking motion but not in the waving motion. Therefore, we adopt another robust method which measures object’s

self-similarity over time [25]. Obviously, if the periodic motion exists in the video sequence, we can find similar poses in different frames. In other words, high correlation value can be obtained when comparing the foreground blobs in different frames within a period of time. Thus we can get the motion period by subtracting the indices of these frames.

2.3. Motion Energy Histogram (MEH). From (1) we can see that the pixel value in $\text{AME}(x, y)$ represents the intensity at position (x, y) . From a histogram point of view, we can regard AME as a two-dimensional histogram whose bin value represents the frequency on position (x, y) during time interval t . Thus, we can reform the AME to the motion energy histogram (MEH) by using the following equation:

$$\text{MEH} = \frac{\text{AME}(x, y)}{\sum_{x, y \in \text{AME}} \text{AME}(x, y)}. \quad (2)$$

After transforming AME to MEH, the process of recognizing two different MEHs becomes a histogram matching process. Hence, various properties of histogram can be employed to improve the recognition performance. Through the observation, we find that the corresponding MEHs of two entirely different actions can be distinguished at very low-resolution. For example, Figure 3 illustrates the MEHs of two actions: walking and waving both hands. The left column is the original MEHs of these two actions and the right column is their corresponding MEH at low resolution. We can see that even under such a low resolution, these two actions can still be distinguished perceptually. In other words, if we can classify these actions at lower resolution level rather than directly compare them at the highest resolution level, the whole recognition process will become more efficient because the procedure will compare less histogram bins. However, the recognition rate may decrease under low resolution levels. In order to maintain the same recognition rate as the method proposed in [15], we adopt a multiresolution structure on the histogram to achieve this goal. Details will be introduced in the next subsection.

2.4. Characteristic of Multiresolution Histogram. The basic idea of multiresolution was firstly introduced in [26] and was further extended to a general form by Yu et al. [27]. A specified partitioning method on histogram bins is used to downscale the resolution of a histogram. By performing this operation recursively we can obtain a pyramid structure of multiresolution histogram. For a given histogram X with b bins, the nonuniform partitioning process for the multiresolution structure is stated as follows. We firstly divide these b bins into m disjointed subsets, B' , $B' = \{B'_1, B'_2, \dots, B'_m\}$. A histogram X' with m bins is defined as the lower-resolution version of X and the bin values of X' are described as follows:

$$X'_k = \sum_{\forall i, i \in B'_k} X_i, \quad k = 1, \dots, m. \quad (3)$$

As Yu et al. show in [27], X and X' satisfy the inequality

$$d(X, Y) \geq d(X', Y'), \quad (4)$$

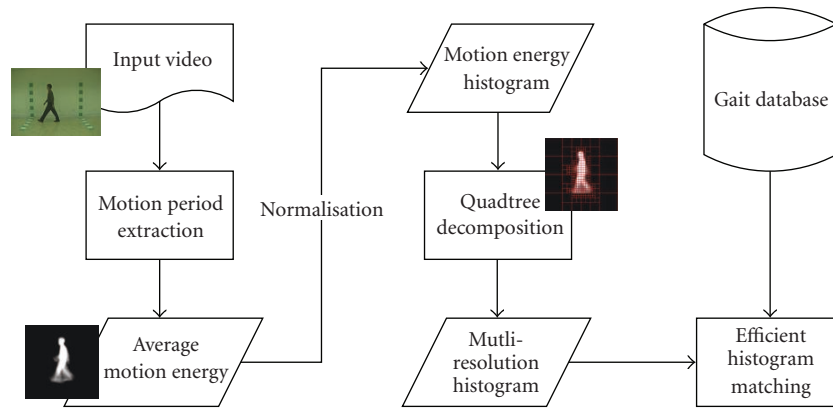


FIGURE 1: System Overview.

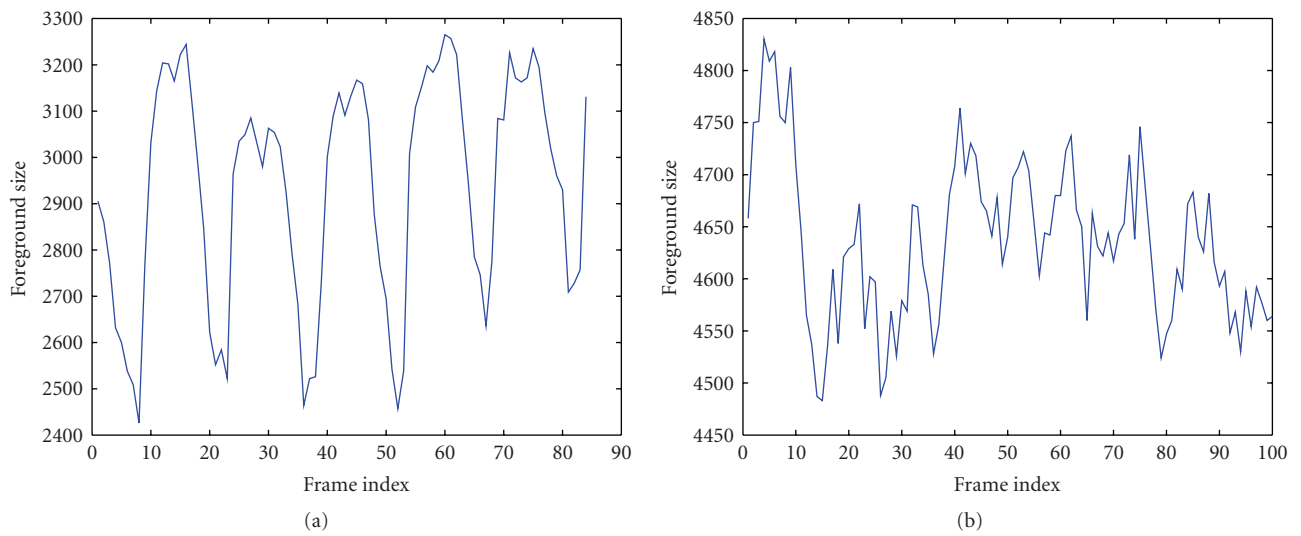


FIGURE 2: The size of foreground silhouette for two motions over time: (a) walking, and (b) waving one hand.

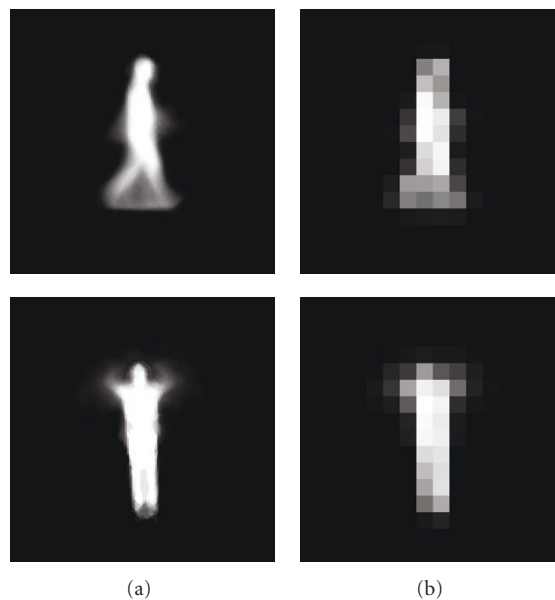


FIGURE 3: MEHs of two periodic motions (a) and their corresponding low-resolution image (b).

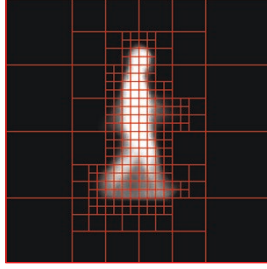


FIGURE 4: Quadtree decomposition results of a walking motion.

where d represents the similarity measurement function L_1 -norm and χ^2 -distance. By using the nonuniform partitioning method iteratively, we can form several multiresolution histograms and the inequality chain in (3) can be rewritten as

$$\begin{aligned} d(X, Y) &\equiv d^L(X_{b^L}, Y_{b^L}) \geq d^{L-1}(X_{b^{L-1}}, Y_{b^{L-1}}) \\ &\geq \dots \geq d^0(X_{b^0}, Y_{b^0}), \quad b^L > b^{L-1} > \dots > b^0, \end{aligned} \quad (5)$$

where b^i denotes the bin size of the current resolution level, $i = 0, \dots, L$ and $1 \leq$ for all $b^i \leq |B|$, where $|B|$ denotes the element number of the universal bin set B . We will introduce how to construct a multiresolution motion energy histogram in the next subsection.

2.5. Construct the Multiresolution Motion Energy Histogram (MRMEH). As we mentioned in Section 2.4, the multiresolution histogram can be constructed by summing certain bins at higher-resolution level. However, in order to make this multiresolution structure more effective, the selection of the disjointed subsets B' is a very important task. As addressed in [26], the uniform partitioning method is a simple and straightforward way. However, Yu et al. [27] proved that the nonuniform partitioning method has better performance than the uniform one. Hence, we decide to create B' via the nonuniform partitioning method. Here, we adopt a straightforward strategy to generate B' automatically. Since different actions have different characteristics, it is reasonable that for different types of actions, the partitioning method should also be different. In terms of image, the result of a quadtree represents a partition of space in two dimensions by decomposing the image into four equal quadrants, subquadrants, recursively. Each node in the quadtree either has exactly four children or no children (a leaf node). Each leaf node contains the data corresponding to a specific subregion of the image. Because of the characteristic of quadtree, we can address the importance of a given image automatically with a proper setting of decomposition criterion. Thus, we can realize that quadtree is a very efficient structure which can represent the characteristic of two-dimensional data. In our work, we set the decomposition rule as “nonzero value”. That is, if the current quadrant contains non-zero pixel values, it will be decomposed into four subquadrants. Since the quadtree decomposition method will divide the image into four subregions each time, the width and height of the image must be the power of 2.

Therefore, we set the bins of MEH in each dimension as 256. Figure 4 is the decomposition result on an MEH of motion “walking” with minimum region size of 64.

In this paper, we propose a method to construct the multiresolution motion energy histogram (MRMEH) by using the results of quadtree decomposition on MEH. The construction of MRMEH can be easily understood through Figure 5. The maximum resolution level is the depth of the tree. Nodes in different colors represent the histogram bin of certain resolution level(s) and the value of each node is the summation of all pixel values in the corresponding subregion. It is reasonable that level i should have 4^i bins at most if every node at level $i - 1$ is decomposed into four children. However, each resolution level will have less than 4^i bins in real world applications because not every node will be decomposed according to the decomposition criterion. More specifically, the bin set does not cover the universal bin set B at some resolution levels because not every node will be decomposed to that level. To solve this problem, if we found that a bin at level i is indecomposable, we will copy it to the higher resolution level to retain the information inside. By doing so, we can ensure that the histogram bin set at each resolution level is always equal to B . As shown in Figure 5, the number of histogram bins at level 2 and 3 will be 10 and 22, respectively. Figure 6 is the MRMEH of the action “running”. Note that the two-dimensional histograms are reformed to one-dimensional for better understanding. Figure 6(a) is the MEH and Figure 6(b) is the quadtree decomposition result. Figures 6(c) to 6(g) are the corresponding MRMEHs at resolution levels 1 to 5, respectively. In the next subsection, we will introduce how to utilize the MRMEH to recognize human actions.

2.6. Efficient Action/Gait Recognition Using MRMEH. As we mentioned in Section 2.4, for two histograms using the same partitioning method, the similarity between these two histograms at different resolution levels will obey the inequality described in (4). From this equation we can realize that further comparison is unnecessary at higher-resolution level if the similarity at lower-resolution level is above the threshold. Thus we can speed up the recognition process by comparing less histogram bins. It is noticeable that we must use same partition method for all compared histograms to let the similarity between different actions be comparable. Thus, we adopt a dynamic partition method. More specifically, we obtain the partition method from the query MEH and employ it to all MEHs in the database. Assume that there are k -defined actions in the database. The matching algorithm is described in Algorithm 1.

2.7. Time Complexity Analysis. In this section, we will analyze the time complexity between the MRMEH method and the AME-SAD method [15]. Here we evaluate the number of operations among different methods. Assume that the MEH is an n -by- n two-dimensional histogram. According to the Algorithm 1, the proposed method has to apply the quadtree decomposition before starting the matching procedure. Thus, we have to consider the time

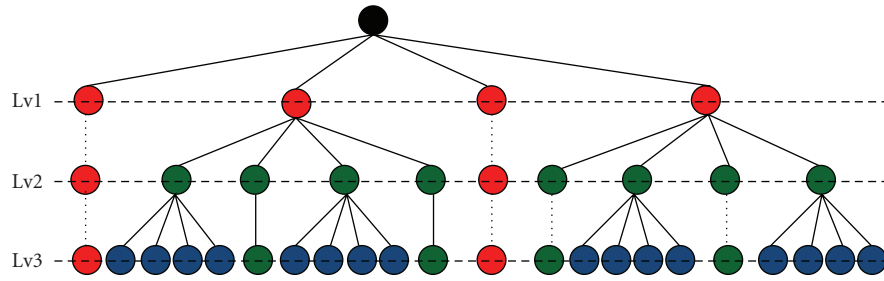
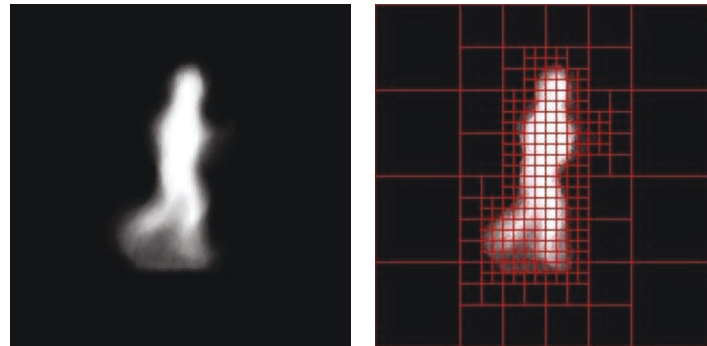
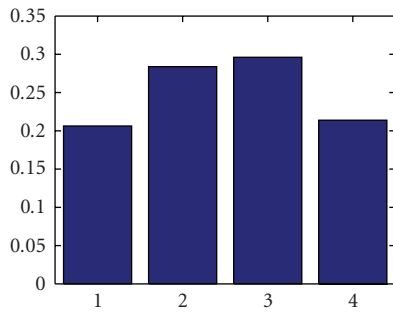


FIGURE 5: The idea of constructing an MRMEH.

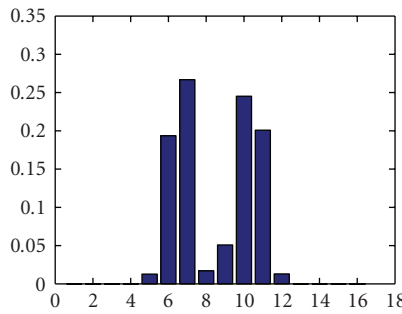


(a)

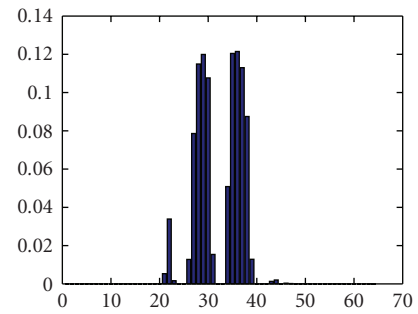
(b)



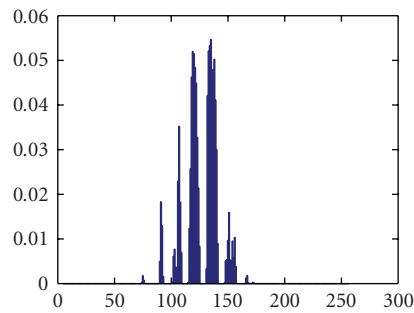
(c)



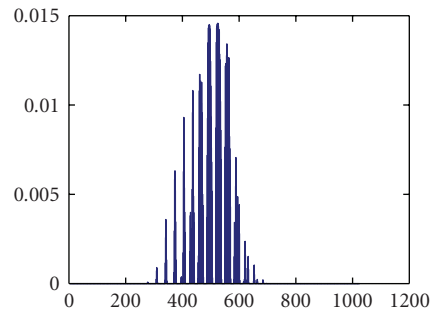
(d)



(e)



(f)



(g)

FIGURE 6: A five-level MRMEH for the motion “running”. (a) Original MEH. (b) Quadtree decomposition result. (c)–(g) Corresponding histogram from levels 1 to 5, respectively.

cost of quadtree decomposition. The number of operations of quadtree decomposition is bounded between n^2 and $n^2 \log_2 n$. Since the number of bins at each resolution level is dynamically determined, it is very difficult to estimate how many bins in total will be compared for each query.

Here, we assume that this branch-and-bound algorithm has to compare m bins to decide whether eliminating a candidate MEH or not. Then, for each H_q , the number of operations will be kn^2 to $kn^2 \log_2 n$ for constructing the corresponding MRMEH and km for matching. This is even worse than the

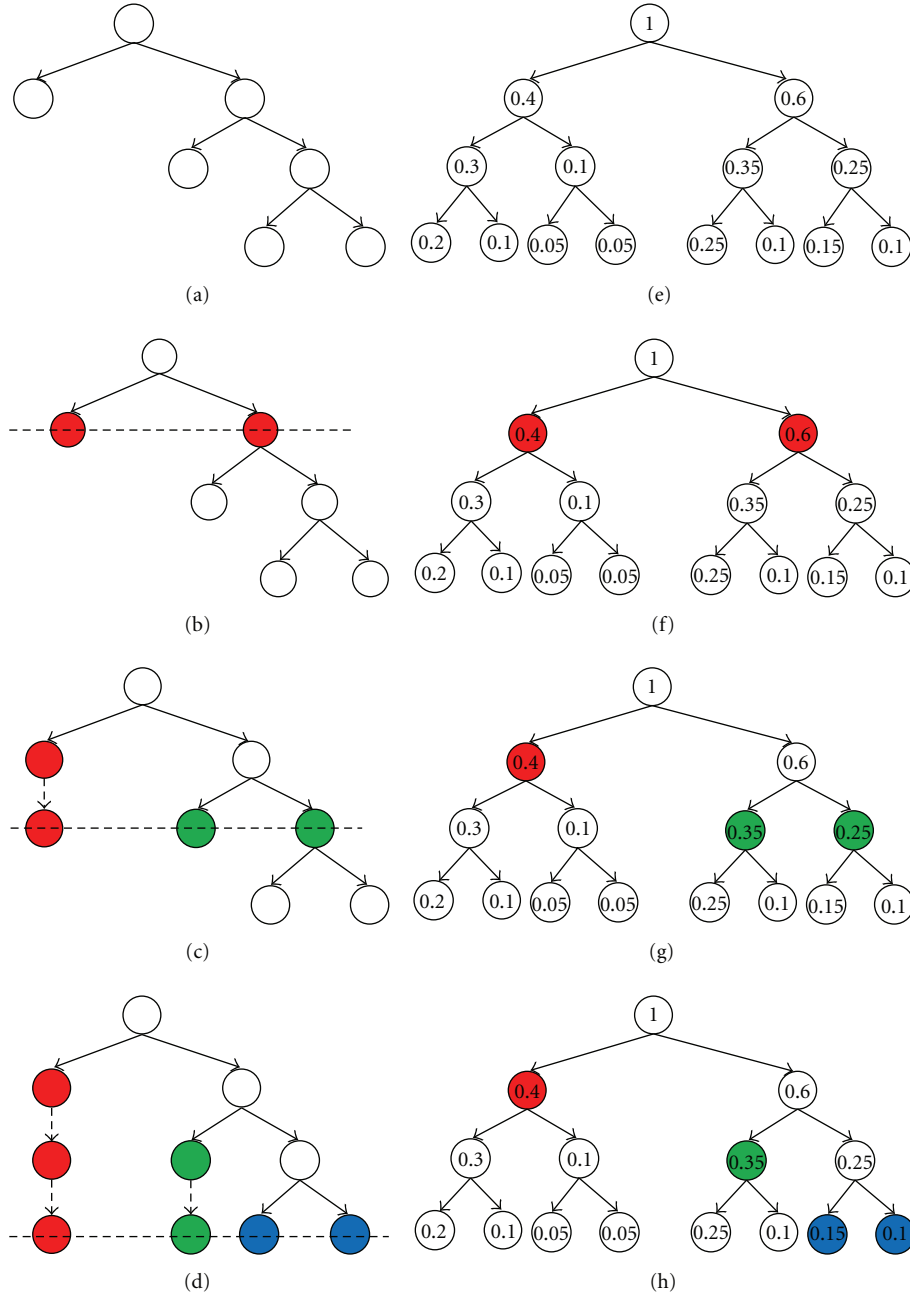


FIGURE 7: An example of matching two histograms effectively. (a) Tree of the query histogram. (e) The preconstructed tree in the database. (b)–(d) Compared bins at different level. (f)–(h) Corresponding bin values in the database.

AME-SAD method, which takes kn^2 operations. However, this matching process can be reduced to km by storing the decomposition information in advance. Hence, the total number of operations will be bounded from $n^2 + km$ to $n^2 \log_2 n + km$ and the recognition time will be relevant to m , not n with the increasing k . In our experiments, m will be much smaller than n^2 and several quantitative results are made to verify this approach.

First of all, we divide the recognition process into two parts: the offline and online procedures. In the offline processing, we will construct a complete quadtree for each

action. More specifically, all leaf nodes will be in the deepest level of the tree. The value in each nonleaf node is the summation of its four children's value. In the online process, after obtaining the quadtree decomposition result of H_q , we can directly find out the value in the corresponding node from the preconstructed tree. By doing so the computation cost will only be related to the number of bins we compared with H_q^{MR} , which is m supposedly. For simplicity, here we take a binary tree, for example, to illustrate this idea. As shown in Figure 7, the right column is the preconstructed tree in the database. Given a quadtree structure in the left



FIGURE 8: Subjects used in the action recognition experiment.

column of Figure 7, the corresponding histogram at levels 1 to level 3 is shown from the second row to the fourth row in the right column, respectively. Thus it is very easy to find out the corresponding node to construct the histogram at a specific resolution level.

3. Experiments

In this section, some experiments were conducted to demonstrate the feasibility, robustness, compatibility, and validity of the proposed approach on two applications: human action recognition and gait classification. All statistics are evaluated

using Matlab 2007b and the Intel Core 2 quad CPU 2.4 G with 2.0 G RAM.

3.1. Action Recognition. To evaluate the performance of the proposed method, we adopt the Weizmann database used in [10]. Because nonperiodic motions are hard to extract the motion period, we only use the periodic motions in their database, which includes 7 periodic actions from 9 different subjects and each subject performs 7 periodic motions. For simplicity, we numbered these 7 actions from A1 to A7, which are walking (A1), running (A2), jumping-forward (A3), jumping jack (A4), waving one hand (A5), waving

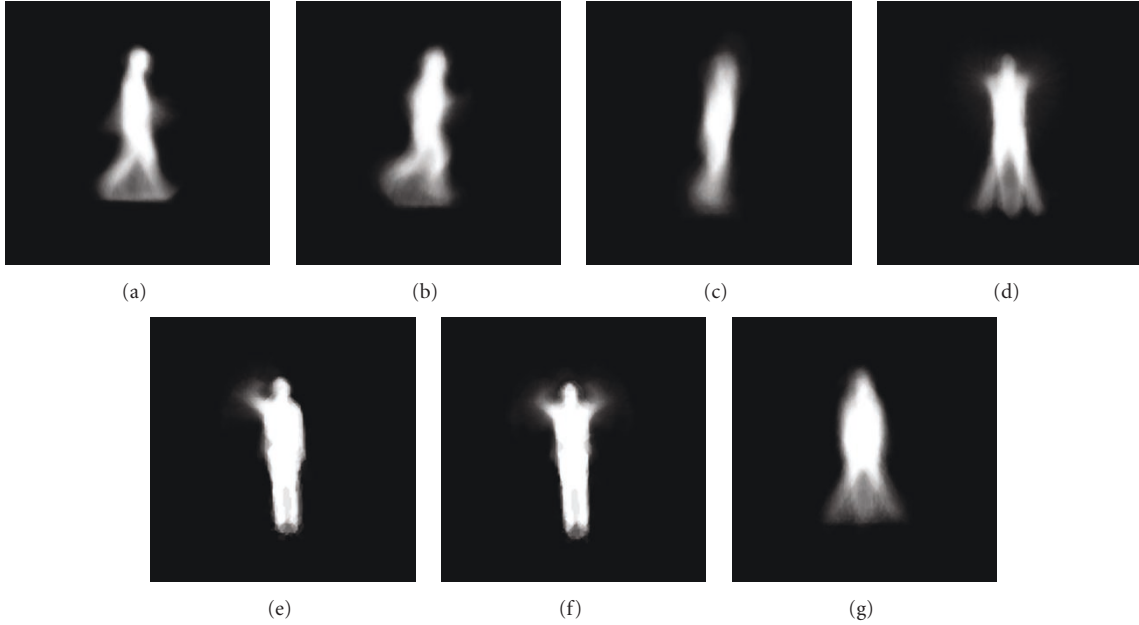


FIGURE 9: MEH of 7 actions. (a) Walking, (b) Running, (c) Jumping, (d) Jumping Jack, (e) Waving one hand, (f) Waving two hands, (g) galloping-sideways.

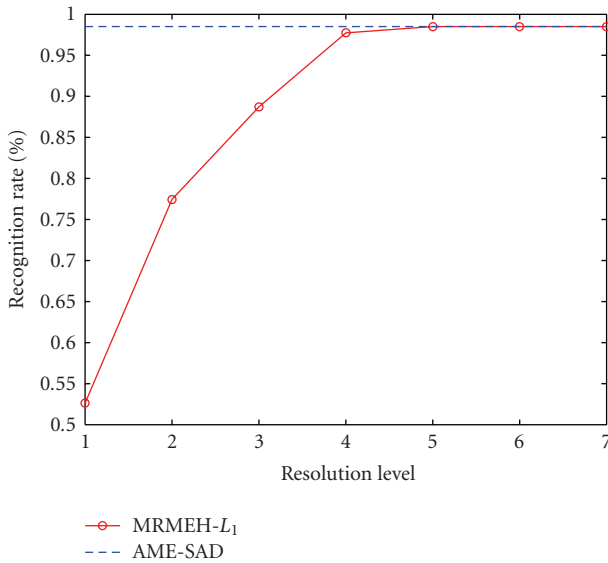


FIGURE 10: Comparison between MRMEH- L_1 and AMD-SAD in terms of recognition rate.

two hands (A6), and galloping-sideways (A7). The resolution of the videos is 180-by-144 and 25 fps. Moreover, we add additional 10 subjects with our own videos in a different environment to enlarge the database. Hence the number of testing subjects is 19 in the whole experiments. The resolution of the additional videos is 360-by-240 and 30 fps. Sample images of all subjects are shown in Figure 8.

For those videos taken by ourselves, we employ a robust background subtraction approach [28] to extract the silhouette. For the database used in [10], we directly use the

masks provided by the authors. All silhouettes were aligned by the center of the mass. Since the width and height of input image must be the power of 2 for quadtree decomposition, we adopted a 256-by-256 image in our experiments. Same subject may have different heights due to different distances to the camera in these videos; we normalize the silhouette according to the scale of height to provide fair comparisons. In our experiments, we normalize the silhouette to 150 pixels high. Figure 9 shows the generated MEH of all 7 actions from all subjects.

3.2. Recognition Accuracy Analysis. In the recognition process, we adopt the leave-one-out cross-validation rule to obtain fair recognition rate. The histogram similarity function used in this paper is L_1 -norm. Although the χ^2 -distance is also compatible of the multiresolution structure, we find that L_1 -norm similarity measurement already provides high recognition rate in the experiments. Table 1 tabulates the recognition rate at different levels. Numbers in bold represent the recognition rate for each action at each resolution level. Figure 10 is the comparison between the AME-SAD method employed in [15] and the proposed MRMEH- L_1 method. The AME-SAD method always compares the whole images to the database images so that the recognition rate is fixed. We can find that the recognition rate is equal to the AME-SAD method at resolution level 5. Therefore, we only have to construct the quadtree to level 5 (whose minimum block size is 8-by-8) and save more computation time. Tables 2, 3, 4, 5, and 6 show the confusion matrix at each resolution level.

3.3. Recognition Efficiency Analysis. In the second experiment, several quantitative results are conducted to evaluate

TABLE 6: Confusion matrix at resolution level 1.

	A1	A2	A3	A4	A5	A6	A7
A1	0.53	0.05	0.05	0	0.05	0	0.32
A2	0	0.79	0.21	0	0	0	0
A3	0.11	0.11	0.68	0.05	0.05	0	0
A4	0.26	0	0	0.27	0.05	0.11	0.37
A5	0.05	0	0	0.11	0.52	0.16	0.16
A6	0.16	0	0	0	0.05	0.42	0.37
A7	0.16	0	0.05	0.05	0.11	0.21	0.42

TABLE 7: Average number of bins at each resolution level.

Lv	A1	A2	A3	A4	A5	A6	A7
1	4	4	4	4	4	4	4
2	16	16	16	16	16	16	16
3	40	40	39.8	40	40	39.9	39.9
4	84.7	86.5	76.8	97.9	81.6	87	88.6
5	204.4	219.5	168.2	268.3	184.7	207.1	213

TABLE 8: Comparisons between AME-SAD and two kinds of partitioning methods in action recognition application.

Number of operations	AME-SAD	Quadtree- L_1	Uniform- L_1
MRMEH construction (no. of pixels retrieved)	—	191451.43	327680
Action recognition (no. of bins compared)	458752	1438.71	5114.69

than AME-SAD method which use the whole image (256-by-256) for comparison. Furthermore, if we can eliminate as many dissimilar MEH as possible at lower-resolution level, the computation efficiency can be further improved. In our experiments, the average processed resolution level is 2.83 when the maximum resolution level is 5. Table 8 depicts the average number of pixels (bins) need to be processed during the recognition process for each query. In order to construct the MRMEH, we have to retrieve each pixel in the AME several times. The average number of pixels needed to be retrieved and compared for all actions is shown in the second and third rows of Table 8, respectively. It is noticeable that the recognition rate of the proposed MRMEH- L_1 (no matter using quadtree-based or uniform partitioning method) and AME-SAD will be the same (see Algorithm 1). Hence the only difference among these methods is the computation time. Although the proposed algorithm has to spend additional time on organizing the MRMEH using quadtree decomposition or uniform partitioning method, the multiresolution structure approach still has better efficiency than the AME-SAD method. With the increasing of the database templates, the proposed method will be much more effective than using the AME-SAD approach.

3.4. Real-Time Action Recognition. Here we demonstrate a system which is able to quickly classify different periodic motions in a single video sequence. As described in [25], we

also adopt the self-similarity matrix S_s in this application. If the motion cycle is successfully extracted at frame t , then we empty the matrix S_s , obtain the corresponding MEH, and perform the action recognition. If we cannot extract the motion period at frame t , we keep computing the self-similarity information at frame t and added it into the matrix S . Since it takes several frames to accomplish a motion cycle, in our work the system tries to extract the motion period in S_s every 15 frames. Because it is meaningless to compute the periodicity between two dissimilar actions, we choose to empty the matrix S when a motion cycle is detected in order to avoid computing the self-similarity between two dissimilar actions. Figure 11 shows the recognition results on a subject who is performing several motions sequentially. These motions are waving one hand, side walking, waving two hands, and running. Since the recognition time of the proposed method is very short, our method can response human's motions immediately and correctly.

3.5. Gait Classification. In this experiment, we will demonstrate the advantage of our method on another application: the human gait classification. Unlike the action recognition problem, the human gait classification problem involves many subjects (usually more than 100) with similar actions so that the classification task becomes more challenging. In order to quickly identify each individual in a large database, an efficient matching algorithm is required. We adopt the CASIA Gait Database (Dataset B) [19] which consists of 124 subjects' gaits from 11 view angles (see Figure 12). Each subject performs the walking motion 10 times which include six natural walks, two walking sequences with a bag, and two sequences with a coat. For simplicity, we only consider the side-view and natural walk sequences in our experiment. We select 120 subjects whose silhouette images are acceptable for gait analysis. Figure 13 shows some examples of the AME images used in our experiments.

We adopt the leave-one-out cross-validation rule and the cumulative match score (CMS) to evaluate the performance of our method. The resolution level of MRMEH is from 1 to 7. In our experiment, the classification rate is 96.39% using the AME-SAD and 1-NN rule. Figure 14 shows the CMS among the top 10 matches at different resolution levels where the y -axis is the recognition rate. In our work, the recognition rate of MRMEH at resolution level 7 reaches the same accuracy as that of AME-SAD. Note that for the action recognition problem described in Section 3.2, the recognition rate of MRMEH at resolution level 5 already reaches the same accuracy as that of AME-SAD. Because the gait classification problem is more sophisticated than the action recognition problem, we have to use a finer resolution level in order to identify different people. Table 9 tabulates comparison of the average number of bins at different resolution levels between the nonuniform (quadtree-based) partitioning method and the uniform partition method. We can find that the bin number of quadtree-based method is only 9.63% of the uniform partitioning method at resolution level 7. Since both methods have the same recognition rate, obviously the quadtree-based partitioning method is

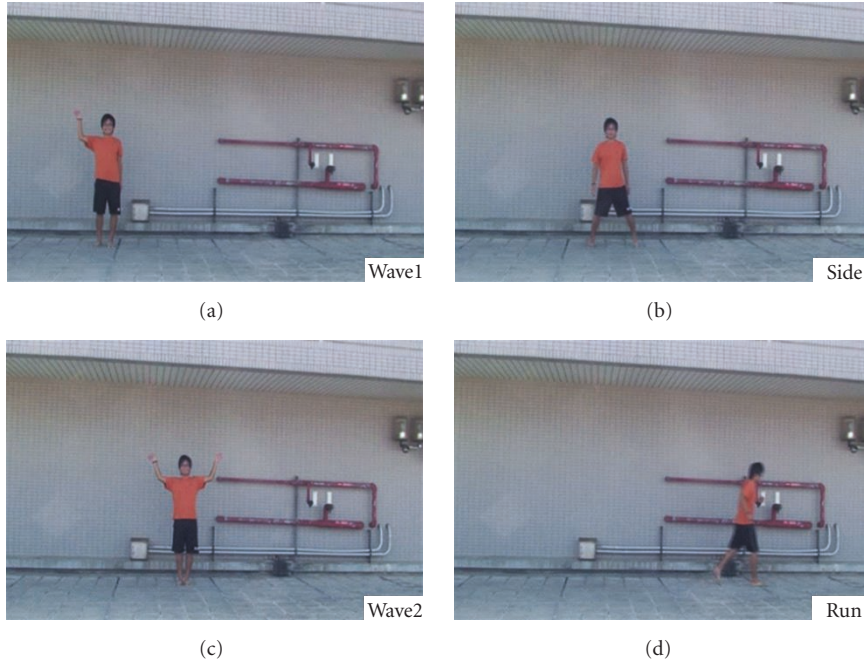


FIGURE 11: Action recognition in a single video sequence.

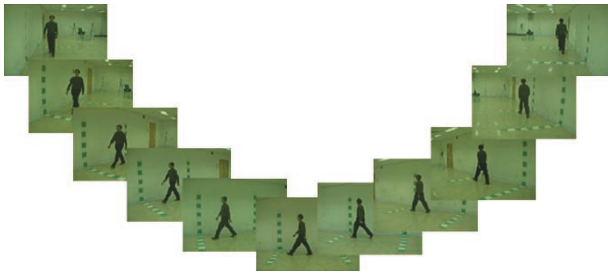


FIGURE 12: Human gaits captured from 11 different view angles.

the better choice in order to obtain better computation efficiency. Table 10 tabulates the average number of comparisons for each query between the AME-SAD and the MRMEH method. We can find that the MRMEH approach (either quadtree-based or uniform partitioning method) performs better than the AME-SAD method. Moreover, the quadtree-based partitioning method has the least number of operations (only 2% of the AME-SAD method) among the three methods.

4. Conclusion

In this paper, we propose a histogram-based approach on human action recognition. We transform the average motion energy (AME) image to the two-dimensional motion energy histogram (MEH). By doing so, various characteristics of histogram can be adopted to improve the performance on similarity measurement. We discover that the MEH generated within a motion period provides rich information to distinguish different type of motions. To improve the

TABLE 9: Number of bins at different resolution levels using the quadtree-based and uniform partitioning method.

Resolution level	Quadtree-based partitioning	Uniform partitioning
1	4	4
2	16	16
3	39.94	64
4	86.98	256
5	216.67	1024
6	581.73	4096
7	1578.30	16384

TABLE 10: Comparisons between AME-SAD and two kinds of partitioning methods in gait classification application.

Number of operations	AME-SAD	Quadtree- L_1	Uniform- L_1
MRMEH construction (no. of pixels retrieved)	—	205753.33	65536
Action recognition (no. of bins compared)	7864320	142225.45	1693850.2

recognition efficiency, we adopt a multiresolution structure on the MEH. The multiresolution motion energy histogram (MRMEH) is a very useful structure which is able to remarkably speed up the recognition process. In order to construct an adequate MRMEH for different actions or people, we proposed an automated partitioning method using the characteristic of quadtree decomposition. The important part in each MEH can be automatically addressed through the decomposition process. Experiments show that the

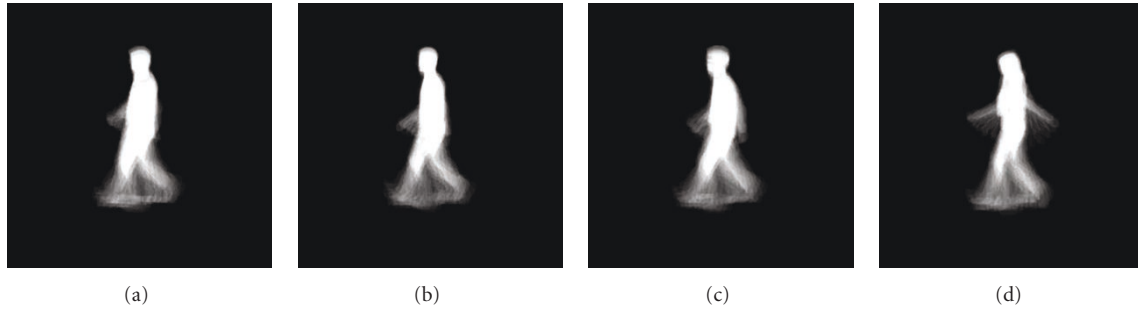


FIGURE 13: The extracted AME images in our experiments.

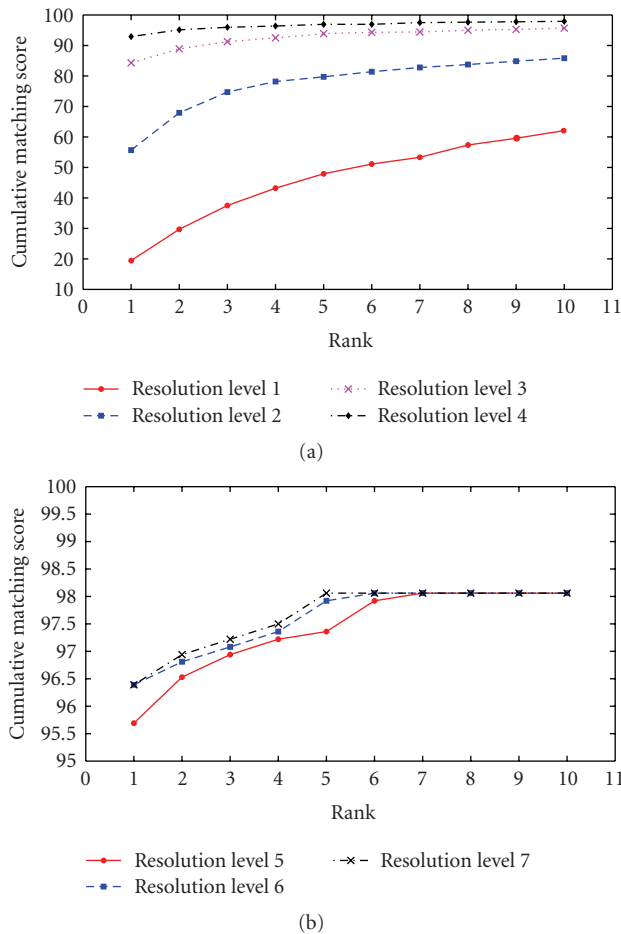


FIGURE 14: The CMS at different resolution levels.

number of operations is highly reduced using the proposed nonuniform partitioning method so that the computation efficiency is greatly improved. Moreover, the recognition accuracy remains the same as the AME-SAD method via the multiresolution histogram matching algorithm. Because of the computationally inexpensive approach, real-time system is practical by using the MRMEH with high recognition accuracy. Through bunch of quantitative experiments we verified our thoughts as well as demonstrated the powerful capabilities of MRMEH in action recognition and gait classification applications.

Acknowledgment

This work is supported by the National Science Council (no. 98-2218-E-238-002-).

References

- [1] L. Middleton, A. A. Buss, A. Bazin, and M. S. Nixon, "A floor sensor system for gait recognition," in *Proceedings of the 4th IEEE Workshop on Automatic Identification Advanced Technologies (AUTO ID '05)*, pp. 171–180, New York, NY, USA, October 2005.
- [2] D. Gafurov and E. Snekkenes, "Gait recognition using wearable motion recording sensors," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, Article ID 415817, 16 pages, 2009.
- [3] A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 257–267, 2001.
- [4] T. H. W. Lam, R. S. T. Lee, and D. Zhang, "Human gait recognition by the fusion of motion and static spatio-temporal templates," *Pattern Recognition*, vol. 40, no. 9, pp. 2563–2573, 2007.
- [5] C. Schüldt, I. Laptev, and B. Caputo, "Recognizing human actions: a local SVM approach," in *Proceedings of the International Conference on Pattern Recognition*, vol. 3, pp. 32–36, 2004.
- [6] A. Fathi and G. Mori, "Action recognition by learning mid-level motion features," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, 2008.
- [7] S. Carlsson and J. Sullivan, "Action recognition by shape matching to key frames," in *Proceedings of the Workshop on Models versus Exemplars in Computer Vision (CVPR '01)*, Kauai, Hawaii, USA, December 2001.
- [8] A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 257–267, 2001.
- [9] A. Veeraraghavan, A. R. Chowdhury, and R. Chellappa, "Role of shape and kinematics in human movement analysis," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 730–737, 2004.
- [10] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri, "Action as space-time shapes," in *Proceedings of the 10th IEEE*

- International Conference on Computer Vision (ICCV '05)*, pp. 1395–1402, Beijing, China, 2005.
- [11] J. Han and B. Bhanu, “Human activity recognition in thermal infrared imagery,” in *Proceedings of the 6th IEEE Workshop on Object Tracking and Classification Beyond and in the Visible Spectrum (OTCBVS '05)*, San Diego, Calif, USA, June 2005.
- [12] H. Lakany, “Extracting a diagnostic gait signature,” *Pattern Recognition*, vol. 41, no. 5, pp. 1644–1654, 2008.
- [13] J. Davis, “Hierarchical motion history images for recognizing human motion,” in *Proceedings of the IEEE Workshop on Detection and Recognition of Events in Video (EVENT '01)*, Vancouver, BC, Canada, July 2001.
- [14] A. Bobick and J. Davis, “The representation and recognition of action using temporal templates,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 257–267, 2001.
- [15] L. Wang and D. Suter, “Informative shape representations for human action recognition,” in *Proceedings of the International Conference on Pattern Recognition*, vol. 2, pp. 1266–1269, 2006.
- [16] X. Zou and B. Bhanu, “Human activity classification based on gait energy image and convolutionary genetic programming,” in *Proceedings of the 18th International Conference on Pattern Recognition*, vol. 3, pp. 556–559, 2006.
- [17] C. Shan, S. Gong, and P. W. McOwan, “Fusing gait and face cues for human gender recognition,” *Neurocomputing*, vol. 71, no. 10–12, pp. 1931–1938, 2008.
- [18] X. Li, S. J. Maybank, S. Yan, D. Tao, and D. Xu, “Gait components and their application to gender recognition,” *IEEE Transactions on Systems, Man and Cybernetics Part C*, vol. 38, no. 2, pp. 145–155, 2008.
- [19] S. Yu, D. Tan, and T. Tan, “A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition,” in *Proceedings of the International Conference on Pattern Recognition*, vol. 4, pp. 441–444, 2006.
- [20] 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 Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 162–177, 2005.
- [21] L. Lee and W. E. L. Grimson, “Gait analysis for recognition and classification,” in *Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition (FGR '02)*, pp. 155–162, Washington, DC, USA, May 2002.
- [22] J. Han and B. Bhanu, “Individual recognition using gait energy image,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 316–322, 2006.
- [23] M.-H. Cheng, M.-F. Ho, and C.-L. Huang, “Gait analysis for human identification through manifold learning and HMM,” *Pattern Recognition*, vol. 41, no. 8, pp. 2541–2553, 2008.
- [24] X. Huang and N. V. Boulgouris, “Human gait recognition based on multiview gait sequences,” *EURASIP Journal on Advances in Signal Processing*, vol. 2008, Article ID 629102, 8 pages, 2008.
- [25] R. Cutler and L. S. Davis, “Robust real-time periodic motion detection, analysis, and applications,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 781–796, 2000.
- [26] B.-C. Song, M. J. Kim, and J. B. Ra, “A fast multiresolution feature matching algorithm for exhaustive search in large image databases,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 5, pp. 673–678, 2001.
- [27] C.-C. Yu, F.-D. Jou, C.-C. Lee, K.-C. Fan, and T. C. Chuang, “Efficient multi-resolution histogram matching for fast image/video retrieval,” *Pattern Recognition Letters*, vol. 29, no. 13, pp. 1858–1867, 2008.
- [28] T. Horprasert, D. Harwood, and L. S. Davis, “A statistical approach for real-time robust background subtraction and shadow detection,” in *Proceedings of the IEEE Frame-Rate Workshop (ICCV '99)*, 1999.