Human Activity Classification Based on Gait Energy Image and Co-evolutionary Genetic Programming

Xiaotao Zou, Bir Bhanu
Center for Research in Intelligent Systems
University of California, Riverside, CA 92521, USA
{xzou,bhanu}@vislab.ucr.edu

Abstract

In this paper, we present a novel approach based on gait energy image (GEI) and co-evolutionary genetic programming (CGP) for human activity classification. Specifically, Hu's moment and normalized histogram bins are extracted from the original GEIs as input features. CGP is employed to reduce the feature dimensionality and learn the classifiers. The strategy of majority voting is applied to the CGP to improve the overall performance in consideration of the diversification of genetic programming. This learning-based approach improves the classification accuracy by approximately 7 percent in comparison to the traditional classifiers.

1. Introduction

Recognizing human activities has become an important research topic in computer vision and pattern recognition. Related research is placed in high priority for the sake of homeland security, anti-terrorism, automated surveillance, etc. Human activity is defined as "something that people do or cause to happen." The goal of human activity-related research is to detect, classify and recognize (ob)normal human activities from the video sequences.

With the increasing use of cheap video cameras in surveillance systems, the continuous and multi-perspective coverage of interested objects (e.g., humans) becomes less difficult. Therefore, it is no longer a dream to obtain the view of the moving human from any perspective continuously. Along with the more powerful computing power and smarter algorithms, the task of recognizing human activity becomes feasible.

Human activity generally fall under two categories: repetitive and non-repetitive. Repetitive human activity involves a regularly repeating sequence of motion events such as walking. The repeating sequence of human motions is called a "cycle." And such periodic motion of the walking human is usually referred as "gait." An efficient spatio-temporal representation of gait, gait energy image [1], is employed in our approach.

In this paper, the related work and our contribution are first discussed in section 2. Then, we present our learning-based approach in section 3 in detail. The experiment results of the proposed approach are presented and compared with other approaches in section 4. Finally, we provide our conclusions in section 5.

2. Related work and our contribution

Two major kinds of approaches have been proposed for human activity recognition: model-based and model-free. In model-based approaches, the structure of the human body and the motion patterns of local body parts are employed to reconstruct the motion model to interpret human activities. Guo et al. [2] represent the human body structure in the silhouette by a stick figure model. The human motion recorded as a sequence of the stick figure parameters are used as the input of a BP neural network for classification. In model-based approaches, the accuracy of the recovered human model strongly depends on the quality of the extracted human silhouettes. The estimated models may be unreliable in the presence of noise. Alternatively, model-free approaches make no attempt to recover the structural model of human motion. Davis [3] proposes a probabilistic reliable-inference framework to address the issue of rapid-and-reliable detection of human activities.

The contributions of our paper can be summarized as: first, we employ an efficient representation of human motion, i.e., gait energy image, for human activity classification; second, a learning algorithm based on co-evolutionary genetic programming is proposed for feature dimension reduction and classifier learning; third, the performance of our proposed approach is tested on the real data of walking people with or without briefcases, and it is found to be superior to the traditional classifiers.

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

Table 1. An example of Hu's moment for the GEI of a human with briefcase and without briefcase.

Hu's moment	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7
Carrying briefcase	.311988	.067428	1.3719e-3	1.4102e-4	3.3018e-9	1.7903e-6	6.1937e-8
Without briefcase	.375040	.084405	.012321	1.899e-3	8.1395e-6	4.8114e-4	4.2573e-6

3. Technical approach

3.1. Gait energy image

Gait energy image is an effective gait representation proposed by Han and Bhanu [1, 4]. The grayscale gait energy image (GEI) is defined as:

$$G(x,y) = \frac{1}{N}\sum_{k=1}^{N} B_k(x,y)$$

where $B_k(x,y)$ is the normalized and aligned binary silhouette image at time k in a gait cycle, N is the number of frames in a complete cycle, and x and y are the 2D image coordinates.

The example GEIs of the persons carrying a briefcase and without one are shown in Fig. 1. All six GEIs in the upper row are from the same sequence of the person carrying a briefcase. Similarly, GEIs in the bottom row are from a walking person without the briefcase. A pixel with higher intensity value in GEI means that human walking occurs more frequently at this position than at those positions corresponding to darker pixels.

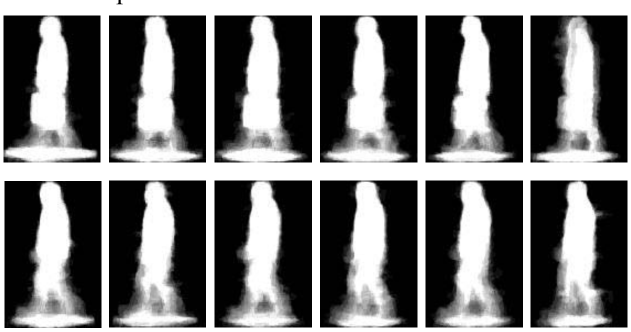


Figure 1. GEIs of the walking person carrying a briefcase (above); GEIs of the people without briefcase (below).

As illustrated in the Fig. 1, gait energy image (GEI) reflects major shapes of human silhouettes and their changes over the motion cycle. Inspired by the successful application of GEI in human identification [1, 4], we apply the idea of GEI in the recognition of repetitive human activity.

3.2 Feature extraction

The dimensionality of the original GEI, e.g., 88 columns * 128 rows = 11264, constrains it as the input feature to classifiers. Additionally, the huge redundancy of gait energy image (over half of GEI pixels are dark) also makes it an implausible and ineffective method of representation. Thus, we consider extracting features from the original GEI to obtain a compact and effective representation.

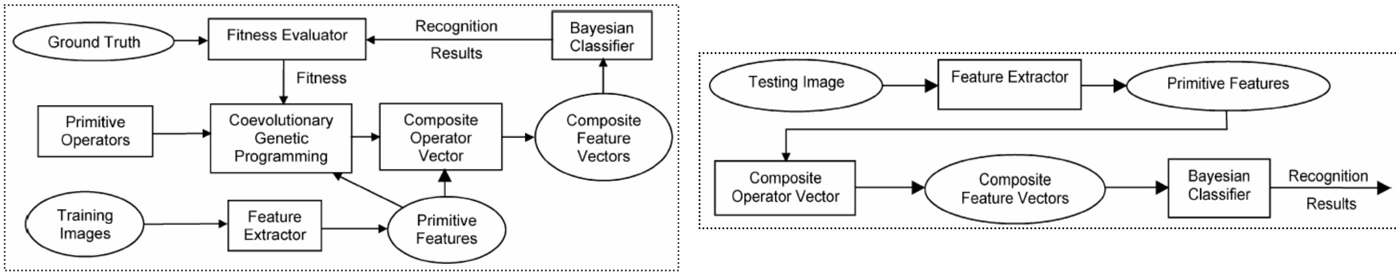
First, we calculate the histogram of grayscale values from the GEI and normalize it to 17 bins. The 17 bins of histograms are used as one set of features to represent GEI.

Furthermore, we calculate the Hu's moment ($\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6, \phi_7$) [5] of the GEIs, which is a set of seven moments derived from the second/third central moments and invariant to translation, rotation and scale change. Examples of Hu's moments of the GEIs for the people with or without briefcases are shown in Table 1. Due to the good property of invariance, we mainly use Hu's moment as the set of features to represent GEI. The normalized histogram bins are also used as classification features for comparison.

3.3. Co-evolutionary genetic programming

Genetic programming (GP) is an evolutionary computational paradigm [6] that is an extension of genetic algorithm and works with a population of individuals. An individual in a population can be any complicated data structure such as trees and graphs, etc.

Co-evolutionary genetic programming (CGP) [7] is an extension of GP in which several populations are maintained and employed to evolve solutions cooperatively. A population maintained by CGP is called a sub-population and it is responsible for evolving a part of a solution. A complete solution is obtained by combining the partial solutions from all the sub-populations. For the task of object recognition, individuals in sub-populations are composite operators, which are the elements of a composite operator vector. A composite operator is represented by a binary tree whose internal nodes are the pre-specified domain-independent primitive operators (linear operators such as summation and non-linear ones such as multiplication) and leaf nodes are original features (e.g., Hu's moments).



(a) (b)

Figure 2. Co-evolutionary Genetic Programming: (a) training phase; (b) testing phase.

The CGP algorithm shown in Fig. 2 includes two different phases: training and testing. It first learns the composite operator vectors and Bayesian classifier over the training set; then applies the learned composite operators and Bayesian classifier to the testing set.

3.4. Majority voting

Multi-agent methodology can be used to boost the overall performance of the evolutionary process. The basic prerequisite for the agents' fusion is their diversification. In our approach, agents correspond to Bayes classifiers accompanying the CGP. Therefore, the diversification is naturally provided by the random nature of the genetic search. We run multiple genetic searches that start from different initial populations. With the same parameters, a number of independent GP synthesis processes provide a statistical significance to the results. Each run starts with the same GP parameters, but with different, randomly created, initial population of feature synthesis programs. In classification, we employ the majority voting strategy as the decision rule: the Bayes classifier following each GP run is considered to be a voting agent where votes can be cast for each of the testing images. In the end we take the majority vote for each testing image (GEI).

4. Experimental results

4.1. Data

The experiments are conducted using the data from the computer vision/image analysis research lab at the University of South Florida. This data set consists of silhouette sequences of humans walking with or without a briefcase on an elliptical path. Therefore, there are two classes of activity for our task: with a briefcase and without a briefcase. In each sequence, approximately 200 to 250 frames are available. The data set contains silhouettes of 470 people without a briefcase and those of 466 people carrying a briefcase. The size of all silhouettes is normalized to 88*128 pixels. The number of silhouettes in one cycle to form a GEI varies from 15 to 19. And depending on the length of sequences, the number of GEIs in a sequence could be between 5 and 14, and, in most cases, it occurs at 10 and 11. From the data set, we choose a small set and a large set for our experiment (described in Table 2). In constructing the small dataset, only one GEI in each sequence is selected. Therefore, the total number of GEIs in the small set is equal to the number of sequences for each class. While in constructing the large dataset, all GEIs in each sequence are selected. So, the number of GEIs for each class increases dramatically in comparison to that of the small dataset. Furthermore, both datasets are split into training and testing sets accordingly. The selection of training and testing sets are done randomly. The total GEIs in each sequence are divided into training and testing sets approximately equally.

4.2. Experiments

After obtaining the small and large datasets of GEIs, we calculate the desired histogram and Hu's moment features for classification. The original feature size of the Hu's moments is 7. The size of GEI histogram bin vectors is 17. In the following, we first discuss the execution of our experiment on the small dataset. And the experiment on the large dataset is similarly performed and the classification results are presented later in this section.

First, the original Hu's moment features are sent to two statistical classifiers: Bayesian classifier and Fisher discriminant. As for the first one, the classification accuracy in training and testing are

Table 2. Description of experimental data.

Class	Number of sequences	Small set			Large set		
		Total GEIs	# of GEIs for training	# of GEIs for testing	Total GEIs	# of GEIs for training	# of GEIs for testing
With briefcase	942	942	500	442	10068	3765	6303
Without briefcase	930	930	500	430	9693	3714	5979

74.1% and 69.5% respectively. The latter one achieves the accuracy of 72.3% in training, and 70.59% in testing.

Then, the Hu's moment features are sent to the Co-evolutionary genetic programming (CGP) to learn the composite feature vectors and the Bayesian classifier. In CGP, the dimensionality of feature vectors is reduced from 7 to 2. The fitness values (i.e., accuracy) of the Bayesian classifier in training and testing phases are 73.7% and 72.83%, respectively. In consideration of the random nature of the genetic programming, we further apply the "majority voting" strategy on the Bayesian classifiers to improve the classifier performance. In the experiment, we set the number of runs of CGP to be five. And those candidate classes with at least three votes are chosen as the final results. Thus, the training and testing fitness values on the small dataset are 77.69% and 77.3%.

For comparison, we send the GEI histogram features to the CGP and learn the Bayesian classifier for classification. The histogram features are reduced from 17D to 2D by CGP. The accuracy of Bayesian classifier is 71.15%. With the help of majority voting, the classification accuracy is improved to 75.77%. Also, we perform experiments with different feature reduction schemes, e.g., 17D to 5D, 17D to 1D, etc. Among them, the performance of '17D to 2D' is the best. So, we neglect the experiments of other feature reduction schemes.

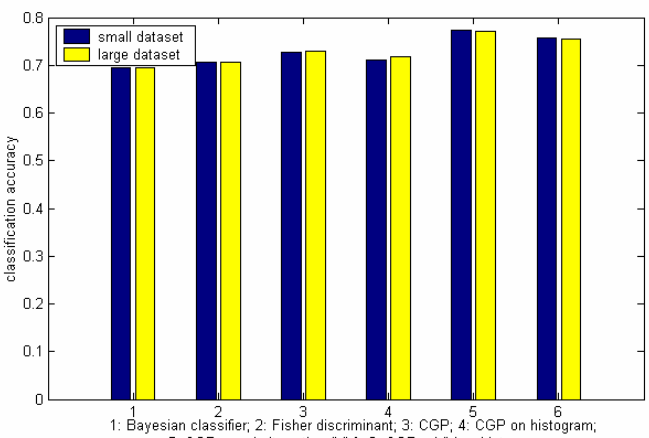


Figure 3. Performance comparison of different classifiers. By default, the input feature is Hu's moment.

4.3. Classifier performance comparison

The complete performance comparison of different classifiers on both large and small datasets is presented in Fig. 3. We can find from it that the CGP improves the classification accuracy by 2- 3 percent. With the help of majority voting, it can further improve by 4- 5 percent. Also, the Hu's moment is found to perform better than the histogram bin features.

The CGP takes a relatively large amount of time to train the composite feature operators and classifier. But it can be performed offline. And the online testing of CGP is really fast and comparable to the Bayesian classifier and Fisher discriminant.

5. Conclusions

In this paper, we present a novel approach based on gait energy image (GEI) and co-evolutionary genetic programming (CGP) for human activity classification. Hu's moment and histogram bins are extracted from the original GEI as input features. CGP is employed to reduce the feature dimensionality and learn the classifier. And "majority voting" is applied to the CGP to improve the overall performance due to the diversification of genetic programming. This learning-based approach improves the classification accuracy by approximately 7 percent in comparison to the Bayesian classifier and Fisher discriminant. CGP, with input from Hu's moment, in conjunction with majority voting performs the best result for classification.

6. References

[1] J. Han, and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 28, No. 2: 316- 322, February 2006.

[2] Y. Guo, G. Xu, and S. Tsuji, "Understanding human motion patterns," *Proc. IEEE Intl. Conf. on Pattern Recognition,* Vol. 2:325- 329, 1994.

[3] J. Davis, "Sequential reliable-inference for rapid detection of human actions," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Vol. 1: 111- 118, June 2004.

[4] J. Han, and B. Bhanu, "Statistical feature fusion for gait-based human recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Vol. 2: 842- 847, June 2004.

[5] M. K Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Info. Theory*, No. 8: 179- 187, 1962.

[6] J. Koza, *Genetic Programming II: Automatic Discovery of Reusable Programs*, MIT Press, 1994.

[7] Y. Lin, B. Bhanu, "Evolutionary feature synthesis for object recognition," *IEEE Trans. Systems, Man, and Cybernetics, Part C,* Vol. 35, No. 2: 156- 171, May 2005.