

ACTIVITY RECOGNITION USING HISTOGRAM OF ORIENTED GRADIENT PATTERN HISTORY

Dipankar Das

Department of Information and Communication Engineering, University of Rajshahi,
Rajshahi-6205, Bangladesh

ABSTRACT

Human activity recognition is an important task in computer vision because it has many application areas such as, healthcare, security, entertainment, and tactical scenarios. This paper presents a methodology to automatically recognize human activity from input video stream using Histogram of Oriented Gradient Pattern History (HOGPH) features and SVM classifier. For this purpose, the proposed system extracts HOG features from a sequence of consecutive video frames and analyzes them to construct HOGPH feature vector. The HOGPH feature vectors are used to train a multi-class SVM classifier for different human activities. In test mode, we use the classifier with HOGPH feature vector to recognize human activity. We have experimented with video data of human activity in real environments for three different tasks (browsing, reading, and writing). The experimental result and its accuracy reveal that the proposed system is applicable to recognize human activity in real-life.

KEYWORDS

Human Activity, Activity Recognition, Histogram of Oriented Gradient, SVM.

1. INTRODUCTION

Recognizing human activity or task from real-time video data is one of the promising and challenging applications of computer vision. Recently, this research has attracted the attention of many researchers from different disciplines including Human-Computer-Interaction (HCI), and Human-Robot-Interaction (HRI). The main objective of human task or activity recognition is to provide useful information on a user's behaviour that allows computing system or robot to proactively assist users or to make a comfortable interaction with him/her [1]. Researchers in computer vision and machine learning have investigated gestures and activity recognition from static images and video in constrained environment or stationary settings[2][3][4]. However, there are very limited number of works to recognize human task or activity in unconstrained daily life settings. In this research, we recognize human task or activity in unconstrained real-world environments, such as Internet browsing in an office environment, reading and writing in an library environment etc., using Histogram of Oriented Gradient Pattern History (HOGPH) and Support Vector Machine (SVM) classifier.

Human activity is very complex and diverse characteristics because an activity can be performed in many different ways, depending on the different context and for a multitude of reasons. Although some state-of-the-art system obtained good performance on many activity recognition tasks, however, most of the researchers so far mainly focus on recognizing "which" activity is

being performed at a specific point in time. In contrast, only small number of researches investigated means to extract qualitative information from the sensor data that allows us to infer additional characteristics. It can be shown that such qualitative assessment are very difficult to perform automatically, and has so far only been demonstrated for constrained settings, such as sports[5][6]. For general activities or tasks, activity recognition research work is still far from reaching a similar understanding. However, in this research, we proposed an activity recognition system that tries to assess the qualitative characteristics of the human activity. For this purpose, the proposed activity recognition system first extracts HOG features of activity during a certain period of time to build up an activity pattern known as the histogram of oriented gradient pattern history (HOGPH). Then the HOGPH feature vector are used to classify the human activity using support vector machine (SVM) classifier.

To develop a good human activity recognition system, we need to develop a pattern recognition system that is robust to intra-class variability. In activity, such type of intra-class variability exists because same type of activity may be executed different ways by different person. Intra-class variability can also happen if an activity is performed by the same person at different time. Thus, for activity recognition, we need a classifier that adapts this type of intra-class variability. To address this issue, the proposed system uses the multi-class SVM classifier with an increase amount of training data for different activity classes using *liblinear* kernel. The proposed HOGPH features are highly discriminative and person independent that also make help the system to adapt intra-class variability.

2. RELATED WORKS

Since there are some successful researches in activity recognition, many researchers are motivated toward more challenging and application-specific scenarios. Some real-world application are benefited from task or activity recognition such as the office scenarios, the sports and entertainment sector[7][8], the industrial sector[9][10], and healthcare. The activity of daily living [11] attracted a significant amount of attention of the vision researchers [12] [13] [14]. The traditional medical methods can be enhanced by analyzing the daily human activities [6]. Another important aim of human activity analysis is to provide a healthy lifestyle. Thus many researchers are encouraged to research in the related human activities, e.g. food intake [15] and medication[16], brushing teeth or hand washing[17], or transportation routines[18].

Currently different companies or agencies use activity recognition as a key component in their consumer products. For instant, to fundamentally change the game experience in game consoles the NintendoWii or the Microsoft Kinect rely on the recognition of gestures or even full body movements. Although they are mainly developed for the entertainment purpose, however, these systems have used in other application areas, such as for personal fitness training and rehabilitation, and also stimulated new activity recognition research [19]. These examples highlight the importance of human activity or task recognition in both academic area and industrial section. In spite of considerable improvement in inferring activities from on-body inertial sensors and in prototyping and deploying activity recognition systems [20], developing human activity recognition systems that meet application and user requirements remains a challenging task. In this research, we recognize human task or activity in unconstrained real-world environments, such as Internet browsing in an office environment, reading and writing in an library environment etc., using Histogram of Oriented Gradient Pattern History (HOGPH) and SVM classifier.

In human activity recognition system, Aggarwal *et al.*[21] propose state-space and template matching based techniques. Some researchers use context-based decisions making techniques to recognize human activity and action [22][23]. In [22], the authors propose an action recognition technique that includes entering a room, using a computer, opening a file cabinet, picking up the telephone, etc. In their system, they able to recognize actions based on prior knowledge about the layout of the room. However, the system is limited to actions like sitting and standing. Also, it is only able to recognize a picking action by knowledge of where the object is and tracking it after the person has come within a certain distance of it. Moreover, both of the above contextual based approaches require prior knowledge of the precise location of certain objects in the environment. Davis *et al.* use temporal plates for matching and recognition of human activity [24]. The system computes motion history images (MHI's) of the persons in the scene. Although template matching procedures have a lower computational cost, they are usually more sensitive to the variance in the duration of the movement.

3. PROPOSED APPROACH

The proposed human activity recognition system is illustrated in Figure 1 . In our approach, we recorded the video data of human activity, such as browsing, reading, and writing, in a real-world environment. The recorded video data is divided into two sections: training activity data, and testing activity data. The proposed system is also divided into two phases: training and testing. In training stage, we extract the histogram of oriented gradient features (HOG) [25] from each video frame.

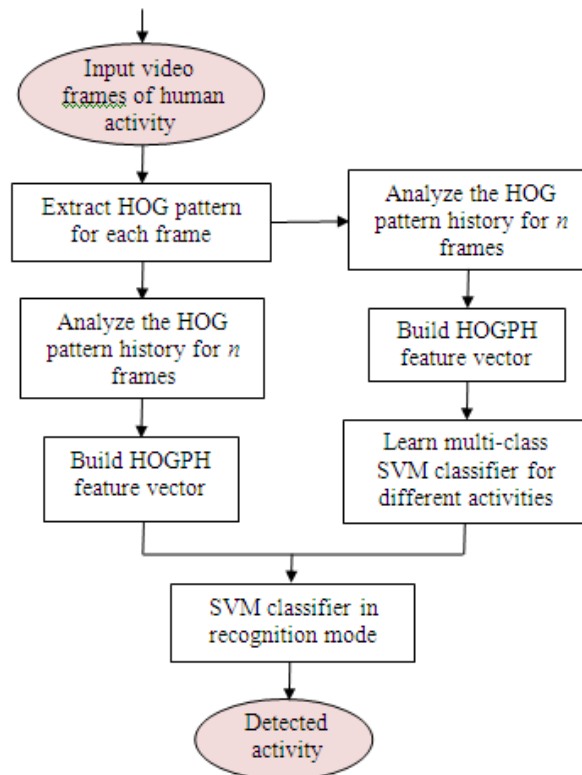


FIGURE 1: THE PROPOSED HUMAN ACTIVITY RECOGNITION APPROACH.

The HOG feature vectors from n consecutive video frames are analyzed to generate the histogram of oriented gradient pattern history (HOGPH). The generated HOGPH feature vectors are used to train the multi-class SVM classifier model for all activities. In testing stage, for each human activity we generate the HOGPH feature vector and feed into the SVM model in recognition mode. The model classifies the human activity into an appropriate class.

4. EXTRACTION OF HOG FEATURE

In this research, we use the histogram of oriented gradient (HOG) feature from each video frame to represent human activity. The method is originally developed by N. Dalal and B. Triggs in [25] for human detection purposes. Their method is based on evaluating well-normalized local histograms of image gradient orientations in a dense grid. The fundamental principle is that local object shape and appearance can often be represented rather well by the distribution of local intensity gradients, even without precise knowledge of the corresponding gradient or edge positions. In practical implementation, we divide the image window into small spatial regions known as cell. Then for each cell, we build up a local 1-D histogram of gradient directions over the pixels of the cell. The combined histogram entries form the gradient orientation representation. The histogram is also useful to contrast-normalize the local responses to make it invariance to illumination, and shadowing. This can be done by building up a measure of local histogram energy over somewhat larger spatial regions (block) and using the results to normalize all of the cells in the block. Here the normalized descriptor blocks are known as Histogram of Oriented Gradient (HOG) descriptors.

5. CONSTRUCTION OF HOGPH FEATURE VECTOR

The human activity is determined by recognizing the pattern of task history in which the target person is involved. For instance, if the target person is involved in a “reading” task, then the pattern history such as “*downward the head*” indicates that his/her attention is toward the book. Such type of activity related pattern is determined by analysing HOG feature vector as follows.

Given a video sequence, we extract the histogram of orientation gradient (HOG) feature for each frame. The HOG features are combined for n consecutive frames to build a HOG feature pattern history, HOGPH according to the following equation,

$$HOGPH = HF_0 + \sum_{i=1}^n HF_{i-1} - HF_i \quad (1)$$

where HF_0 and HF_i are the HOG features of the first and i -th frames, respectively. The first frame captures the human appearance features involved in a task, and the rest of the HOG feature frames indicate the change of behaviour pattern in the activity during the observation history. Thus, the HOGPH feature captures the appearance of the activity and the activity related behaviour pattern history. Here, each bin in the histogram represents the number of edges that have the orientation within a given angular range. The angular range is set to 20 degrees and we use unsigned gradient. Thus, the bin size is equal to $180/20 = 9$. With this bin size, if we create the HOGPH feature vector for 10 consecutive video frames ($i = 1, \dots, 9$) then the HOGPH feature vector is of size 90. A multi-class support vector machine (SVM) is learned using HOGPH feature.

6. ACTIVITY CLASSIFICATION

In our activity recognition approach, a multi-class support vector machine (SVM) classifier is learned with HOGPH features vector. To represent the appearance of an activity, HOG descriptors are extracted from the first frame of an activity image. In order to describe the activity pattern history we analyze the subsequent HOG feature and calculate the difference of two consecutive frames. Here the activity pattern is represented by its local shape. The local shape is represented by orientations of an edge histogram within an object's subregion quantized into K -bin and each edge's contribution is weighted by its magnitude. Therefore, each bin in the histogram represents the number of edges that have orientations within a given angular range. The final pattern history of human activity is represented by the HOGPH feature vector. The multi-class SVM classifier [26] is learned using HOGPH feature vector. We use the LIBSVM package for our experiments in a multi-class mode with the *liblinear* and *rbf* exponential kernels.

In the verification step, the HOGPH feature vector is extracted from activity video frames and fed into the multi-class SVM classifier in recognition mode. Only the hypotheses for which a positive confidence measurement is returned are kept for each activity. Activity with the highest confidence level is recognized as the correct activity. The confidence level is measured using the probabilistic output of the SVM classifier.

7. EXPERIMENTAL RESULTS

We conducted experiments to investigate the performance of the vision-based system to human activity when s/he is involved in a task.

7.1. Experimental Data

We implemented the proposed system and conducted experiments to verify activity recognition performance. To collect experimental data, we asked 14-non-paid participants (11 males and 3 females) for three different tasks: browsing, reading, and writing. They were graduate students at Saitama University, Japan, with an average age of 27 years old. They did not receive any information for their activity or task. All the tasks were recorded using a video camera. The average recorded lengths for each person for the activities of browsing, reading, and writing were 8, 9, and 9 minutes, respectively. Figure 2 show some snapshot of activities where the humans were involved in browsing, reading and writing tasks, respectively.

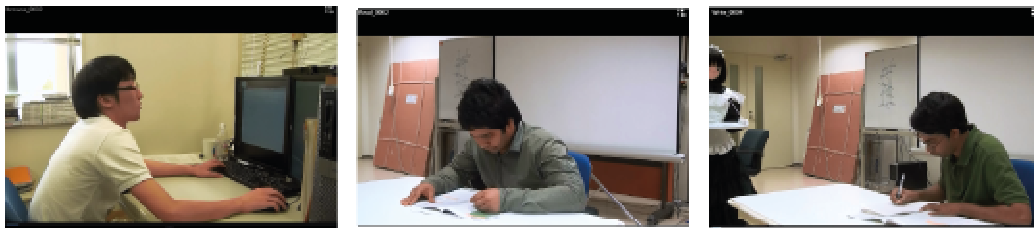


Figure2: Human involved in activities browsing, reading, and writing, respectively.

7.2. Training and Testing the SVM Classifier

Each of the recorded tasks was divided into test and train sample videos. As described in Section 5, one sample features a vector consisting of ten frames of the video stream. Table 1 shows the number of training and testing sample used in the experiments and their total length for three different activities. The multi-class SVM classifier was learned using the training samples for three activities. In the recognition mode, each test sample value was tested against three tasks and the task with the highest probability was selected as the detected task. Total 8740 test samples for three activities were used to test the system. The SVM classifier was used with two different kernels: *liblinear* and *rbf*. The performance of the classifier was compared for different kernels on the proposed activity test sample dataset.

Table1: Sample Training and Test Data for Experiments

Activity	Training Data		Test Data	
	Length (in Min)	No. of Samples	Length (in Min)	No. of Samples
Browsing	19.30	2482	22.93	3104
Reading	16.00	2058	19.10	2904
Writing	15.00	1929	19.91	2732
Total	50.30	6469	61.94	8740

7.3. Activity Recognition Performance

We measured the activity recognition performance on the test video data. From the test video data, we used 8740 test activity samples. First, we measured the performance of the SVM classifier with *rbf* exponential kernel. In this case, among 8740 test samples 6064 activity samples were correctly recognized with 69.38% accuracy. Table 2 shows the activity specific recognition performance and cross category confusion of activity recognition system using the *rbf* exponential kernel.

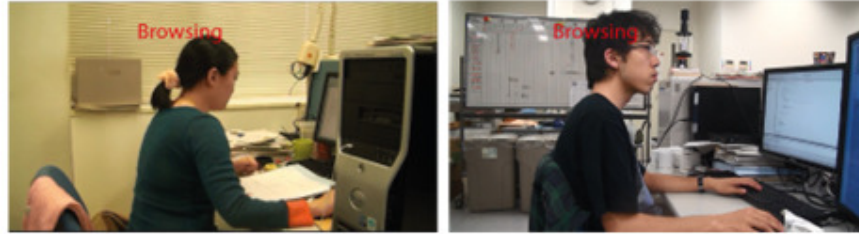
Table 2: Activity Recognition Performance Using SVM Classifier and rbf Exponential Kernel

Activity	Browsing	Reading	Writing
Browsing	2439	302	363
Reading	318	1935	651
Writing	278	764	1690

Second, the performance of the proposed system was also measured using SVM classifier and *liblinear* kernel. Here, among 8740 test samples 8277 activity samples were correctly recognized with 94.70% recognition accuracy. Table 3 shows the activity specific recognition performance and cross category confusion of activity recognition system using the *liblinear* kernel. From the experimental result, it was shown that the proposed system performed significant improvement for human activity recognition task with *liblinear* kernel than *rbf* kernel. The *liblinear* kernel was also decreased the cross-category confusion rate. Figure 3 shows some snapshots of the recognized human activities for “browsing”, “reading”, and “writing”, respectively.

Table 3: Activity Recognition Performance Using SVM Classifier and liblinear Kernel

Activity	Browsing	Reading	Writing
Browsing	3101	3	0
Reading	28	2448	428
Writing	0	4	2728



(c) Browsing activity



(b) Reading activity



(a) Writing activity

FIGURE 3: SNAPSHOT OF RECOGNIZED ACTIVITIES

8. CONCLUSION

In this paper, we propose an automatic human activity recognition system that can accurately recognize three different activities (browsing, reading, and writing) from real-life video data. The proposed system uses the HOG pattern history (HOGPH) feature vectors that represent human activity and its behavior accurately and precisely. We analyze and combine the HOG features of human activity for few consecutive video frames. The first frame represents appearance of the activity and the remaining frames indicate the change of behavior pattern during activity. The HOGPH feature vectors are used to learn and classify human activity using SVM classifier. In this research, the performance of the classifier is compared using two different kernels: *rbf* and *liblinear*. The experimental result shows that the proposed system produces higher accuracy with

liblinear kernel than *rbf* kernel. In future, we want to extend and compare the proposed system for human activity recognition with more versatile databases.

REFERENCES

- [1] G. D. Abowd, A. K. Dey, R. Orr and J. Brotherton (1998), "Context-awareness in wearable and ubiquitous computing," *Virtual Reality*, vol. 3, no. 3, pp. 200-211.
- [2] S. Mitra and T. Acharya (2007), "Gesture Recognition: A Survey," *IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 37, no. 3, pp. 311-324.
- [3] P. Turaga, R. Chellappa, V. S. Subrahmanian and O. Udrea (2008), "Machine Recognition of Human Activities: A Survey," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1473-1488.
- [4] J. K. Aggarwal and M. S. Ryoo (2011), "Human activity analysis: A review," *Comput. Surveys*, vol. 43, no. 3, pp. 16:1-16:43.
- [5] E. Veloso, A. Bulling, H. Gellersen, W. Ugulino and H. Fuks (2013), "Qualitative Activity Recognition of Weight Lifting Exercises," in 4th Augmented Human International Conference (AugmentedHuman 2013).
- [6] B. Tessorodorf, F. Gravenhorst, B. Arnrich and G. Troster (2011), "An IMU-based Sensor Network to Continuously Monitor Rowing Technique on the Water," in ISSNIP.
- [7] K. Kunze, M. Barry, E. Heinz, P. Lukowicz, D. Majoe and J. Gutknecht (2006), "Towards Recognizing Tai Chi—An Initial Experiment Using Wearable Sensors," in FAWC.
- [8] D. Minnen, T. Starner, I. Essa and C. Isbell (2006), "Discovering Characteristic Actions from On-Body Sensor Data," in 10th IEEE International Symposium on Wearable Computers (ISWC).
- [9] I. Maurtua, P. T. Kirisci, T. Stiefmeier, M. L. Sbodio and H. Witt (2007), "A Wearable Computing Prototype for Supporting Training Activities in Automotive Production," in 4th International Forum on Applied Wearable Computing.
- [10] T. Stiefmeier, D. Roggen, G. Ogris, P. Lukowicz and G. Troster (2008), "Wearable Activity Tracking in Car Manufacturing," *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 42-50.
- [11] S. Katz, T. Downs, H. Cash and R. Grotz (1970), "Progress in development of the index of ADL," *The Gerontologist*, vol. 10, no. 1, p. 20.
- [12] L. Bao and S. S. Intille (2004), "Activity Recognition from User-Annotated Acceleration Data," in *Pervasive*.
- [13] N. Ravi, N. Dandekar, P. Mysore and M. L. Littman (2005), "Activity recognition from accelerometer data," in 17th International Conference on Innovative Applications of Artificial Intelligence.
- [14] B. Logan, J. Healey, M. Philipose, E. Tapia and S. Intille (2007), "A long-term evaluation of sensing modalities for activity recognition," in *UbiComp*.
- [15] G. Pirkel, K. Stockinger, K. Kunze and P. Lukowicz (2008), "Adapting magnetic resonant coupling based relative positioning technology for wearable activity recognition," in *ISWC*.
- [16] R. d. Oliveira, M. Cherubini and N. Oliver (2010), "MoviPill: Improving Medication Compliance for Elders Using a Mobile Persuasive Social Game," in *UbiComp*.
- [17] J. Lester, T. Choudhury and G. Borriello (2006), "A Practical Approach to Recognizing Physical Activities," in *International Conference on Pervasive Computing*.
- [18] J. Krumm and E. Horvitz (2006), "Predestination: Inferring Destinations from Partial Trajectories," in *UbiComp*.
- [19] J. Sung, C. Ponce, B. Selman and A. Saxena (2011), "Human Activity Detection from RGBD Images," in *AAAI Workshop on Plan, Activity, and Intent Recognition*.
- [20] D. Ashbrook and T. Starner (2010), "MAGIC: a motion gesture design tool," in *CHI*.
- [21] J. K. Aggarwal and Q. Cai (1999), "Human motion analysis: A review," *Computer Vision and Image Understanding*, pp. 428-440.
- [22] D. Ayers and M. Shah (1998), "Recognizing human action in a static room," in *Computer Vision and Pattern Recognition*.
- [23] S. S. Intille, J. Davis and A. Bobick (1997), "Real time closed world tracking," in *IEEE International Conference on Computer Vision and Pattern Recognition*.
- [24] J. Davis and A. Bobick (1997), "The representation and recognition of action using temporal plates," in *Computer Vision and Pattern Recognition*.

- [25] N. Dalal and B. Triggs (2005), "Histograms of oriented gradients for human detection," in CVPR (1).
[26] I. Tsochantaridis, T. Joachims, T. Hofmann and Y. Altun (2005), "Large margin methods for structured and interdependent output variables," Journal of Machine Learning Research, vol. 6, pp. 1453-1484.

Authors

Dipankar Das received his B.Sc. and M.Sc. degree in Computer Science and Technology from the University of Rajshahi, Rajshahi, Bangladesh in 1996 and 1997, respectively. He also received his PhD degree in Computer Vision from Saitama University Japan in 2010. He was a Postdoctoral fellow in Robot Vision from October 2011 to March 2014 at the same university. He is currently working as an associate professor of the Department of Information and Communication Engineering, University of Rajshahi. His research interests include Object Recognition and Human Computer Interaction.

