J Electr Eng Technol.2015; 10(4): 1910-1914
http://dx.doi.org/10.5370/JEET.2015.10.4.1910

ISSN(Print) 1975-0102
ISSN(Online) 2093-7423

Human Posture Recognition: Methodology and Implementation

Kyaw Kyaw Htike* and Othman O. Khalifa†

Abstract – Human posture recognition is an attractive and challenging topic in computer vision due to its promising applications in the areas of personal health care, environmental awareness, human-computer-interaction and surveillance systems. Human posture recognition in video sequences consists of two stages: the first stage is training and evaluation and the second is deployment. In the first stage, the system is trained and evaluated using datasets of human postures to 'teach' the system to classify human postures for any future inputs. When the training and evaluation process is deemed satisfactory as measured by *recognition rates*, the trained system is then deployed to recognize human postures in any input video sequence. Different classifiers were used in the training such as Multilayer Perceptron Feedforward Neural networks, Self-Organizing Maps, Fuzzy C Means and K Means. Results show that supervised learning classifiers tend to perform better than unsupervised classifiers for the case of human posture recognition.

Keywords: Posture recognition, Human activities, Intelligent classifiers.

1. Introduction

This Human Posture Recognition is a key component of many application-oriented computer vision systems, for instance in automated visual surveillance, automotive safety, human-computer interaction and multimedia processing. Human tracking is an important part of any automated video surveillance system [1-3]. It is used to track any previously detected human for the mapping or prediction purpose or simply for behavioural analysis. High detection rates and low false alarm rates are essential for achieving robustness in higher level vision tasks such as tracking or activity recognition. In this work, a single static camera has been used and video sequences have to be recorded as part of the data acquisition task. The environment where video recording took place is indoors with a relatively simple background scene. Only one person is assumed to be present in front of the camera at a time. The human posture recognition system, like other intelligent computer vision systems, is composed of training stage and evaluation stage. The outputs of the preprocessing are binary images which are then randomly divided into training dataset and testing (or evaluation) dataset in a certain ratio. The binary preprocessed images (henceforth referred to as training samples) are trained and evaluated with various classifiers. Each classifier has to be trained and tested one at a time. In addition, each classifier always has to be re-trained and re-tested several times in order to reach optimal parameters for that classifier [4, 5].

† Corresponding Author: Dept. of Electrical and Computer Engineering, International Islamic University Malaysia, Malaysia. (khalifa@iium.edu.my)
* School of Computing at University of Leeds, UK.
Received: October 27, 2014; Accepted: April 15, 2015

2. System Descriptions

The system implementation consists of two *stages* which are:

1. Training and evaluation stage.
2. Deployment stage.

2.1 Training and evaluation

In the training and evaluations stage as shown in Fig. 1, all the parameters of the system must be incrementally improved to optimal values so that the model would be ready for deployment.

The *video camera* is the data acquisition device which in this case is a digital camera running in video recording mode. The *video sequences* recorded from the video camera are converted into datasets of static color images (one image corresponds to one frame of a video sequence). The images are then run through the *pre-processing* step which is a combination of many algorithms and is

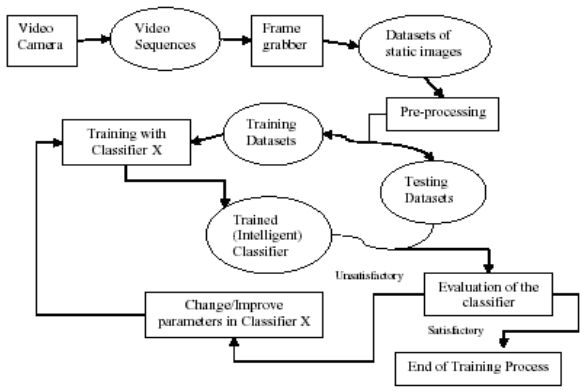

Fig. 1. Training and evaluation stage

described in detail in the next section. The outputs of the pre-processing are the binary images which are then randomly divided into *training dataset* and *testing (or validation) dataset*. The binary preprocessed images (henceforth referred to as training samples) are trained and evaluated with various classifiers whose performances are to be compared. Each classifier has to be trained and tested one at a time. In addition, each classifier always has to be re-trained and re-tested several times in order to reach optimal parameters for that classifier. In computer vision literature, there is so far no known or certain way to pre-calculate or estimate the optimal parameters which would give optimal results. Many of the textbooks suggest a systematic trial and error approach [6-8]. For example, in training and evaluating neural networks, there are numerous variables that have to be taken into account. Only a variable is allowed to vary at a time and the rest of the variables are held constant. Graphs are then plotted to identify the value which gives the highest performance. [9, 10] The reason for having two separate datasets for training and evaluation is to obtain unbiased evaluations of the results.

2.2 System deployment

After the system has been trained and evaluated many times and optimal parameters for a model have been obtained, the trained model is then ready to be deployed. Whilst the inputs to the system are static images (samples) in the dataset during the training and evaluation stage, the inputs to the system are video sequences in the deployment stage. The deployment stage is depicted in Fig. 2.

In the deployment stage, each frame of a recorded video sequence which is running at 30 frames per second (fps) is grabbed at a lower speed such as 6 fps. The reason why a lower sampling rate can be used is because, to continuously recognize human postures in a video sequence, most of the frames do not need to be processed since they are redundant or each frame contain very similar posture to the next. Thus, for any input video sequence which contains a human moving or changing from pose to pose at a reasonable speed, it is sufficient to process only a few frames in a second.

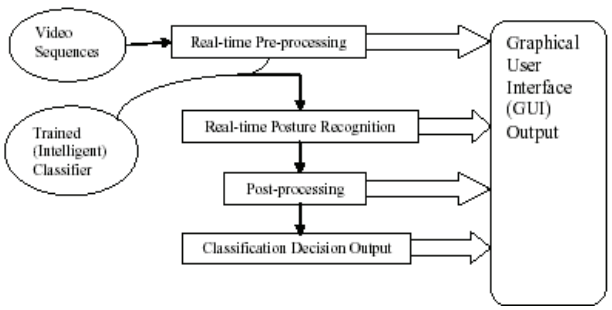

Fig. 2. Deployment stage

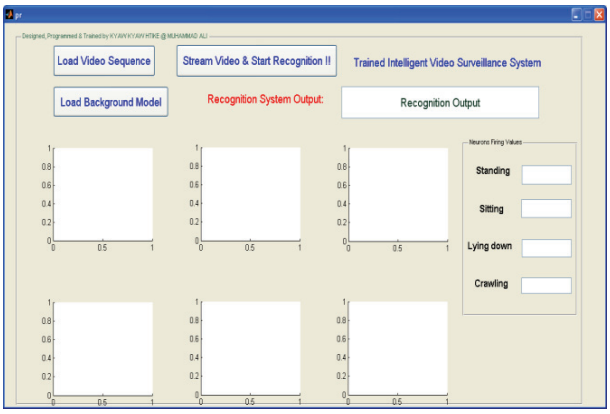

Fig. 3. The GUI of the deployed recognition system

For each frame that has been grabbed, the posture recognition (which includes pre-processing, feature extraction and classification) is performed. And then the postprocessing step is done. All the results can be seen on the Graphical User Interface (GUI) in real-time as the screenshots of the GUI of the deployed system are shown in Fig. 3.

To achieve the real-time requirement, the recognition speed of any grabbed frame must be less than or equal to the sampling rate.

3. Pre-processing and Tracking

The pre-processing component is shown in Fig. 3 and it consists of the following steps.

1. Calculating the background model. In this paper, the first frame of a video sequence is considered to be the background model for that video sequence.
2. Although this does not allow very robust pre-processing, it is extremely fast speed-wise and is sufficient for the paper.
3. Calculating absolute difference between the current image and the background image.
4. Calculating the global image threshold using Otsu's method. This chooses the threshold to minimize the intra-class variance of the black and white pixels.
5. A white silhouette of the human on the black background is obtained by assigning the pixels in the images as black or white, with black being the background and white being the foreground, by using the threshold calculated above.
6. A 2D median filtering algorithm can be applied to the resulting binary image.
7. Median filtering is a nonlinear operation often used in image processing to reduce "salt and pepper" noise. A median filter is more effective than convolution when the goal is to simultaneously reduce noise and preserve edges. Assuming that the input image can be considered as m-by-n matrix, each output pixel contains the

median value in the m-by-n neighborhood around the corresponding pixel in the input image.

8. Dilation and erosion, which are standard morphological operations in image processing, are then applied to the resulting image. The morphological structuring element used was a 'disk'.
9. Blob analysis is then performed from the image to calculate the bounding box which surrounds the (white) human foreground with minimum area. In other words, the bounding box is the smallest rectangle which contains the white pixels that make up the human blob. After obtaining it, the center of gravity of the human blob is calculated.

4. Evaluation Criteria

Each classifier has to be re-trained and re-tested numerous times until an optimal performance for that classifier has been arrived. The primary measure of performance in this chapter is *recognition rate*. It is defined as the proportion of correct classification decisions made over the total classification decisions that have to be made [3, 9, 11, 12].

Mathematically, it can be written as in Eq. (1):

$$Recognition\ Rate\ for\ classifier = \frac{No.of\ correct\ classifications\ made\ for\ all\ samples\ in\ testing\ dataset}{Total\ no.of\ samples\ in\ testing\ dataset} \quad (1)$$

Usually, the term "recognition rate" refers to percentage recognition rate and is expresses as follows:

$$Recognition\ Rate(\%) = Recognition\ Rate\ proportion \times 100\%$$

Some researchers also use the term *error rate* [9] which is simply the recognition rate subtracted from one hundred (if given in percentages). In this report, the term *recognition rate* shall be predominantly used.

There is no universal agreement on the 'goodness' of recognition rates. Different scientists seem to possess different judgments on that front. The most crucial reason for the absence of a universal agreement is owing to the fact that each system is built with dissimilar assumptions, limitations, robustness and scopes. Therefore comparison of recognition rates only applies locally, i.e. within the same system. Nevertheless, if the datasets used are exactly the same, the recognition rates can be compared across different systems, which is why scientists prefer to use standard datasets in the field of computer vision. In case of when one is a pioneer of a particular area, the person would usually make his own datasets available to other scientists via a communication channel, usually the Internet. In this paper, one of the datasets used is Iris Flower dataset which is a universally accepted standard for testing the recognition rate of a new classifier. This particular dataset has been included to make an estimate of the performance of the classifiers on a global basis.

5. Results and Analysis

After the training and evaluation stages, MLP neural networks was chosen to be the classifier for the deployment because it gives the highest accuracy as can be seen from

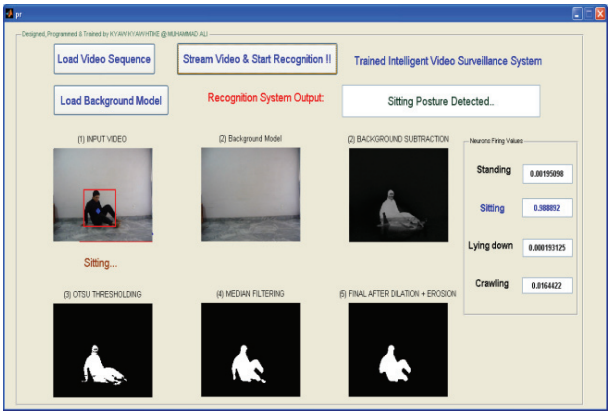

Fig. 4. Sitting posture detected by the system

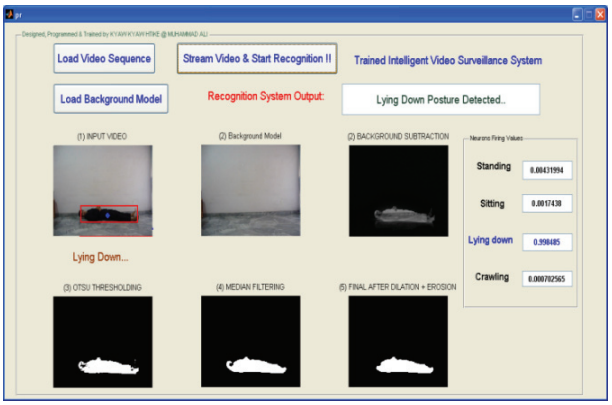

Fig. 5. Lying down posture detected by the system

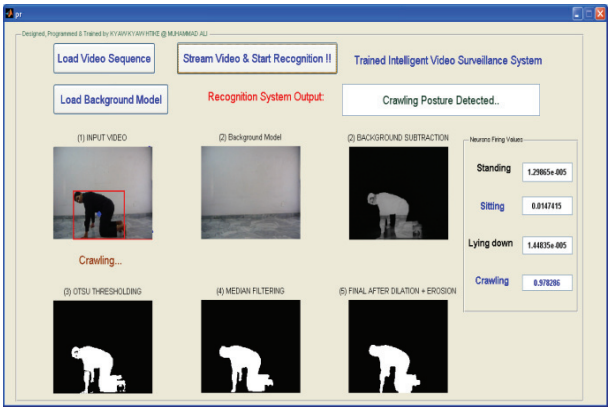

Fig. 6. Crawling posture detected by the system

the results. The screenshots of the GUI of the deployed system are shown in Fig. 4 to Fig. 6.

Among the different classifiers trained and evaluated, MLP consistently gives the highest recognition rates whilst K Means results in the lowest recognition rates. This means that K Means is not 'sophisticated' enough for complex datasets such as human posture datasets. However, for a simple dataset such as the Iris flower dataset, the recognition rate of K Means is quite high. Furthermore, for simple datasets, FCM seems to have a much better performance than K Means. Also, For each classifier, recognition rate has been found to be proportional to the number of postures trained and evaluated. For example, if a posture dataset contains only two postures, the resulting recognition rate would generally be higher than with the dataset which consists of 5 postures. This observation confirms the hypothesis that the higher the complexity of a dataset, the lower the recognition rate would be.

6. Conclusion

An intelligent human posture recognition system has been designed and implemented in MATLAB using supervised and unsupervised classifiers. The system consists of training and evaluation stage and deployment stage. Both of the stages consist of complex sub-stages. The goal of the training stage is to obtain optimal parameters of the system (or model), resulting in the highest recognition rate. There were two datasets used in the work. The first dataset was trained using four different classifiers which are MLP Neural networks, SOMs, FCM and K Means. The recognition rates (accuracies) of those classifiers were then compared and results indicate that MLP gave the highest recognition rate (95.5%). The mean speed of the pre-processing process was 12.4 frames per second (fps) whilst that of the overall deployed system (which consists of all the stages including preprocessing and simulating the trained architectures) was 6.1 fps. Assuming that the sampling rate for the input recognition system is only 1 in 10 frames, the system is able to process in real-time.

References

[1] B. Boulay, "Human posture recognition for behaviour understanding," Phd Thesis, Universite de Nice-Sophia Antipolis, 2007.

[2] Guo and Z. Miao, "Projection histogram based human posture recognition," in *Signal Processing, The 8th International Conference on*, vol. 2, 2006.

[3] S. Iwasawa, K. Ebihara, J. Ohya, and S. Morishima, "Real-time human posture estimation using monocular thermal images," *Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998. Proceedings*, 1998, pp. 492-497.

[4] L.H.W. Aloysius, G. Dong, H. Zhiyong, and T. Tan, "Human posture recognition in video sequence using pseudo 2-D hidden Markov models," *Control, Automation, Robotics and Vision Conference, 2004. ICARCV 2004 8th*, 2004.

[5] M. Rahman and S. Ishikawa, "Human Posture Recognition: A Proposal for Mean Eigenspace," *SICE-ANNUAL CONFERENCE*-, SICE; 1999, 2002, pp. 2456-2459.

[6] B. Boulay, "Human posture recognition for behaviour understanding," Phd Thesis, Universite de Nice-Sophia Antipolis, 2007.

[7] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification* (*2nd Edition*), 2nd ed. Wiley-Interscience, November 2000.

[8] J.C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *Cybernetics and Systems*, vol. 3, 1973, pp. 32-57.

[9] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp M. Finocchio, R. Moore, and et al. "Real-time human pose recognition in parts from single depth images". In Proc. of IEEE CVPR, 2011.

[10] H., Zhao and Liu, Z., Recognizing Human Activities Using Non-linear SVM Decision Tree. Journal of Computational Information Systems, 2011. 7(7): pp. 2461-2468.

[11] Thi-Lan Le, Minh-Quoc Nguyen and Thi-Thanh-Mai Nguyen, Human posture recognition using human skeleton provided by Kinect, The Inter-national Conference on Computing, Management and Telecommunications (Commantel 2013).

[12] Z. Zequn, L. Yuanning, A. Li, and W. Minghui, A novel method for user-defined human posture recognition using Kinect, 7th International Congress on Image and Signal Processing (CISP), 2014, pp. 736-740.

Kyaw Kyaw Htike is obtained a Bachelor of Engineering (Electronics-Computer and Information) degree (First Class Honours) in 2010 from International Islamic University Malaysia (IIUM). He is currently a Research Fellow in the Computer Vision Research Group (Institute for Artificial Intelligence and Biological Systems) in the School of Computing at University of Leeds.

Othman O. khalifa received his Bachelor's degree in Electronic Engineering from the Garyounis University, Libya in 1986. He obtained his Master degree in Electronics Science Engineering and PhD in Digital Image Processing from Newcastle University, UK in 1996 and 2000 respectively. His area of research interest is Communication Systems, Information theory and Coding, Digital image / video processing, coding and Compression, He published more than 350 papers in international journals and Conferences including 11 Books.