Human Posture Recognition and Classification

Othman O. Khalifa
Department of Electrical and Computer Engineering
International Islamic University Malaysia (IIUM)
Kuala Lumpur, Malaysia
khalifa@iium.edu.my

Kyaw Kyaw Htike
Department of Electrical and Computer Engineering
International Islamic University Malaysia
Kuala Lumpur, Malaysia

Abstract—Human posture recognition is gaining increasing attention *in the field of 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 is a challenging task which is part of the more general problem of video sequence interpretation. This paper presents a novel an intelligent human posture recognition system for video surveillance using a single static camera. The training and testing were performed using four different classifiers. The recognition rates (accuracies) of those classifiers were then compared and results indicate that MLP gives the highest recognition rate. Moreover, results show that supervised learning classifiers tend to perform better than unsupervised classifiers for the case of human posture recognition. Furthermore, for each individual classifier, the recognition rate has been found to be proportional to the number of postures trained and evaluated. Performance comparisons between the proposed systems and existing systems were also carried out.*

Index Terms—Human Postures, Human Behavior Recognition, Pattern Recognition.

I. INTRODUCTION

All Human posture refers to the arrangement of the body and its limbs. According to Oxford Dictionary (2009), posture is defined as the particular position of the body and the way in which a person holds his or her body. There are several agreed types of human postures such as standing, sitting, squatting, lying, kneeling and other unusual positions such as standing on the arms, standing on the head, being "on all fours" and etc. Some examples of human postures are shown in Figure 1 [1][2].

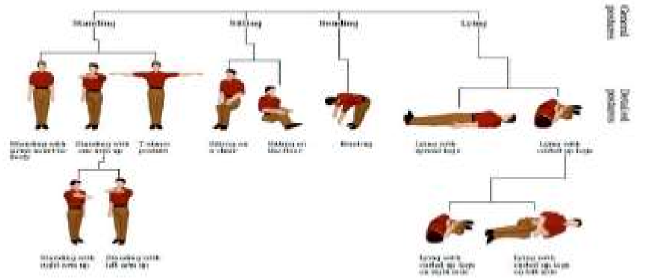

Figure 1: Examples of different human postures [2]

II. SYSTEM OVERVIEW

The posture recognition *system* is made up of two *stages* which are

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

II.1 Training and Evaluation

In the training and evaluations stage as shown in *Figure* 2, all the parameters of the system must be incrementally improved to optimal values so that the model would be ready for deployment.

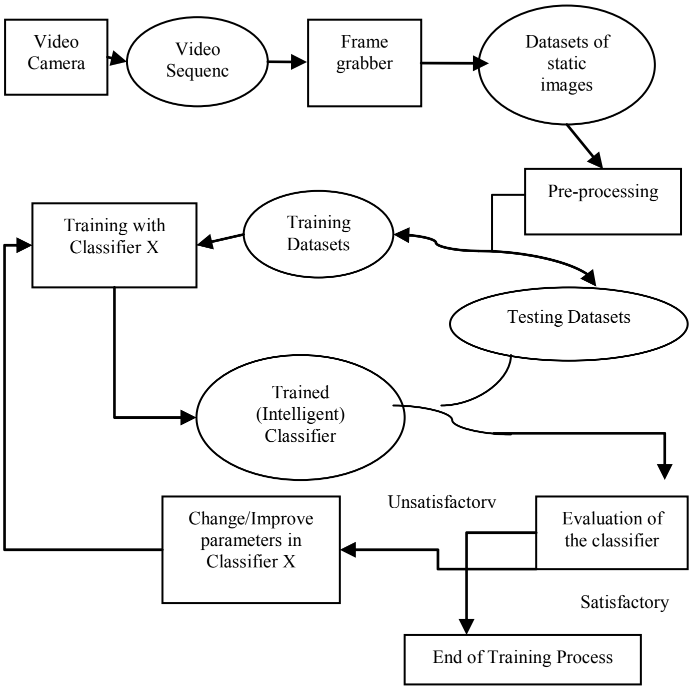
Video Camera
Video Sequenc
Frame grabber
Datasets of static images
Pre-processing
Training with Classifier X
Training Datasets
Testing Datasets
Trained (Intelligent) Classifier
Unsatisfactory
Evaluation of the classifier
Change/Improve parameters in Classifier X
Satisfactory
End of Training Process

Figure 2: Training and Evaluation Stage

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

978-1-4673-6232-0/13/$31.00 ©2013 IEEE

combination of many algorithms and is 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. The classifiers used are:

1. Multilayer Perceptron Feed-forward Neural Networks (MLP-NNs)
2. Self-Organizing Maps (SOMs)
3. K means
4. Fuzzy C Means (FCM)
5. K nearest neighbor (only as a bench mark for Iris dataset)

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. 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. The reason for having two separate datasets for training and evaluation is to obtain unbiased evaluations of the results.

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. 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. The reason for having two separate datasets for training and evaluation is to obtain unbiased evaluations of the results.

III. DEPLOYMERNT

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 Figure .

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.

For each frame that has been grabbed, the posture recognition (which includes pre-processing, feature extraction and classification) is performed. And then the post-processing step is done. All the results can be seen on the Graphical User Interface (GUI) in real-time. To achieve the real-time requirement, the recognition speed of any grabbed frame must be less than or equal to the sampling rate.

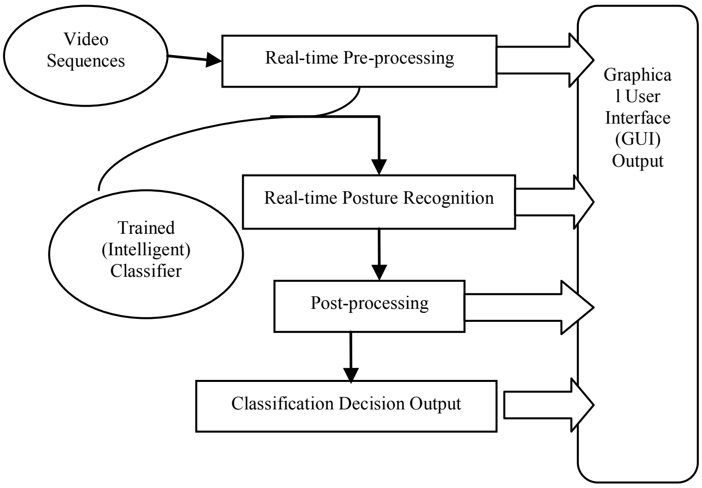
Video Sequences
Real-time Pre-processing
Graphica l User Interface (GUI) Output
Real-time Posture Recognition
Trained (Intelligent) Classifier
Post-processing
Classification Decision Output

Figure 3: Deployment Stage

VI. PRE-PROCESSING & TRACKING

The pre-processing component is shown in Figure 1 and it consists of the following steps.

1. Calculating the background model. In this project, the first frame of a video sequence is considered to be the background model for that video sequence. Although this does not allow very robust pre-processing, it is extremely fast speed-wise and is sufficient for the project.
2. Calculating absolute difference between the current image and the background image.
3. 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.
4. 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.

5. A 2D median filtering algorithm is then applied to the resulting binary image. 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.
6. 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’.
7. 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 centre of gravity of the human blob is calculated.

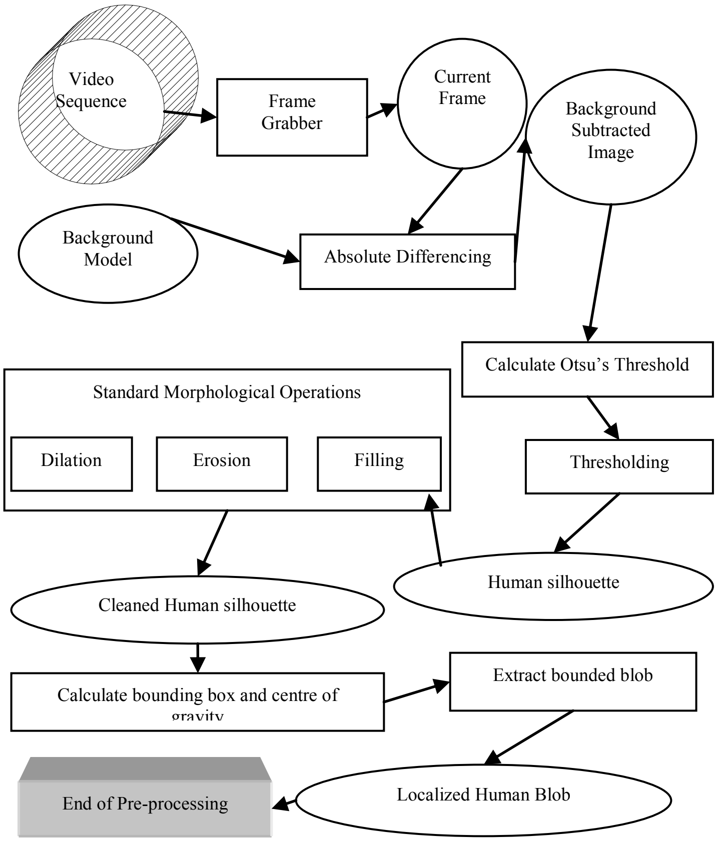

Figure 1: The preprocessing component

V. TRAINING & RESULTS

The training process was done for the three datasets separately.

The results shown in Table 1 are all optimal results. Most of the times, it took several trials and errors (sometimes even days) to obtain optimal results. For the sake for simplicity, only optimal results are shown.

Table 1: Summarized results of Recognition Rates using various classifiers

ID	Technique Used	No. of postures trained	Recognition Rate (%) for each posture						
			climbing	fighting	lying	jumping	pointing	unknown	Mean
1	MLP	6	93.3	100	100	86.7	96.7	96.6	95.5
2	MLP	5	95	100	100	87	98		96
3	SOM	6	100	95	80	85	40	25	70.8
4	SOM	5	100	100	65	85	80		86
5	SOM	2	100	90					95
6	SOM	2	100			96			98
7	K Means	5	65	15	20	25	30		31
8	K Means	2	90				70		80
9	K Means	2	85			80			82.5
10	FCM	5	75	10	35	25	20		33
11	FCM	2	75				95		82.5
12	FCM	2			95	55			75

As can be seen in the table above, the MLP classifier was trained and evaluated for both 6 and 5 postures in the Dataset A. The SOM classifier was trained and evaluated for 6, 5 and 2 postures. Furthermore, both FCM and K Means classifiers were experimented for 5 and 2 postures. Regardless of the number of postures trained and evaluated, MLP gives the highest recognition rate, followed by SOM. In contrast, K Means results in the lowest recognition.

In the graph shown in Figure 2, the horizontal axis shows labels such as MLP [x]. The ‘x’ refers to the number of postures simultaneously trained and evaluated. Furthermore, ‘KM’ stands for K Means.

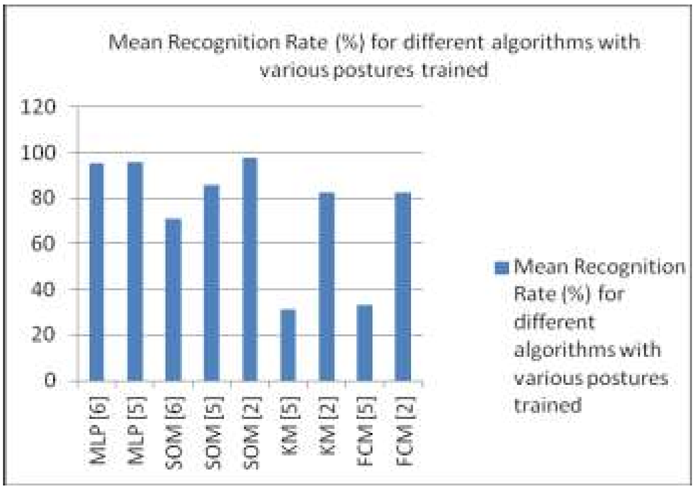

Figure 2: Mean recognition rate for different classifiers with various trained postures

CONCLUSION

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. Supervised learning classifiers tend to perform better than the unsupervised counterparts for the task of human posture recognition.

REFERENCES

[1]. S. Russell, P. Norvig, J. Canny, J. Malik, and D. Edwards, *Artificial intelligence: a modern approach*, Prentice hall Englewood Cliffs, NJ, 1995.

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

[3]. N. Tahir and A. Hussain, "PCA–Based Human Posture Recognition," Jurnal Teknologi D, UTM, 2007

[4]. Girondel, V.; Bonnaud, L.; Caplier, A.; Rombaut, M., "Static human body postures recognition in video sequences using the belief theory," *Image Processing, 2005. ICIP 2005. IEEE International Conference on* , vol.2, no., pp. II-45-8, 11-14 Sept. 2005.

[5]. L.B. Ozer and W. Wolf, "Real-time posture and activity recognition," *Proc. of Workshop on Motion & Video Computing*, 2002, pp. 133–138.

[6]. ZHENG XIAO, 3D HUMAN POSTURES RECOGNITION USING KINECT, 2012 4TH INTERNATIONAL CONFERENCE ON INTELLIGENT HUMAN-MACHINE SYSTEMS AND CYBERNETICS (IHMSC), 26-27 AUG. 2012, PP. 344 - 347 .

[7]. Maleeha, K., Sin, L.T., Chan, C.S., Lai, W.K.: Human Posture Classification Using Hybrid Particle Swarm Optimization. In: Proceedings of the Tenth International Conference on Information Sciences, Signal Processing and their application (ISSPA 2010), Kuala Lumpur, Malaysia, May 10-13 (2010)