Clustering Techniques for Human Posture Recognition: K-Means, FCM and SOM

MALEEHA KIRAN[1], LAI WENG KIN[1], KYAW KYAW HITKE ALI[1, 2]
[1]Centre for Multimodal Signal Processing,
Malaysian Institute of Microelectronics Systems (MIMOS Berhad)
Technology Park Malaysia, 57000 Kuala Lumpur
MALAYSIA
malehaa.kiran@mimos.my; lai@mimos.my; ali.kyaw@gmail.com
http://www.mimos.my;
[2]Department of Electrical and Computer Engineering
Faculty of Engineering, International Islamic University Malaysia, Jalan Gombak,
53100 Kuala Lumpur, Malaysia
http://www.iium.edu.my

Abstract: - An automated surveillance system should have the ability to recognize human behaviour and to warn security personnel of any impending suspicious activity. Human posture is one of the key aspects of analyzing human behaviour. We investigated three clustering techniques to recognize human posture. The system is first trained to recognize a pair of posture and this is repeated for three pairs of human posture. Finally the system is trained to recognize five postures together. The clustering techniques used for the purpose of our investigation included K-Means, fuzzy C-Means and Self-Organizing Maps. The results showed that K-Means and Fuzzy C-Means performed well for the three pair of posture data. However these clustering techniques gave low accuracy when we scale up the dataset to five different postures. Self-Organizing Maps produce better recognition accuracy when tested for five postures.

Key-Words: - Surveillance systems, posture recognition, Clustering, K-Means, fuzzy C-Means, Self-Organizing Maps

1 Introduction

Automated surveillance systems are rapidly becoming a vital tool in providing security. Such systems are increasingly expected not to only scan the images captured, but also to perform some form of real-time analysis on the scene being recorded. As surveillance cameras and video monitors are increasingly being deployed for security purposes, it becomes more and more challenging to adequately monitor and effectively analyze the data resulting from them. Hence the principal goal of automated surveillance system is to automate video monitoring and this is also crucial in combating operator fatigue especially in circumstances where the number of monitors being used for surveillance is huge.

At present most automated surveillance systems are being used to analyze events after they have occurred (for example in post-crime scene analysis). However an intelligent security system should be able to preemptively alert security personnel whenever any suspicious behaviour or event is detected by the system. This would mean the system should have the capability to analyze human behavior and to alert the operators accordingly.

One key aspect of analyzing human behaviour is correctly recognizing different types of posture. Ideally, we want to recognize postures with only one static camera and in real time. These constraints placed on the system can be rationalized because in majority of applications for video surveillance, only one static camera is used to observe the scene and any analysis of the captured images is done in real time [1]. This would mean the processing speed of the system is also a crucial aspect of such systems. However for our research the first priority is achieving high accuracy of correct posture recognition. Only after a satisfactory level of accuracy has been achieved, the system will be optimized to improve its speed.

2 Problem Formulation

Human posture refers to the arrangement of the body and its limbs [2]. There are several agreed types of human postures such as standing, sitting, squatting, lying, kneeling. However for the purpose of our investigation we focus on recognizing five postures known as lying, jumping, fighting (pushing/punching), climbing and pointing. A sample set of these postures is shown in figure 1. Initially the system is designed and tested to recognize three pairs of postures where the chosen pairs were [climbing, pointing], [jumping, climbing], [lying down, jumping]. Afterwards the system will be retrained and tested with all five posture data shown in figure 1.

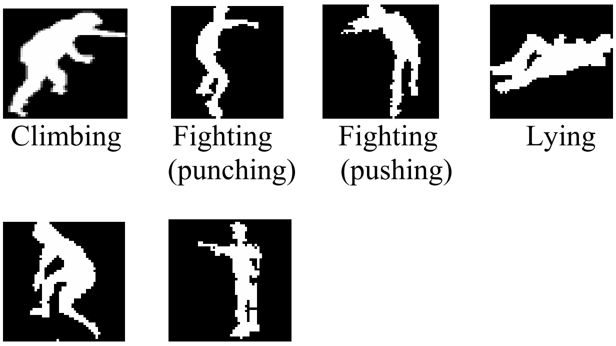


Fig. 1: A sample of posture data

When collecting data for the posture recognition system, we ensured that no occlusion occurs and the entire body of the person is visible. Furthermore it is essential that the the distance of the person from the camera remains constant. These limitations will ensure that the system is more orientated towards posture recognition and does not spend unnecessary computation time on preprocessing work. Similar constraints have been imposed by researchers working in this area as many techniques and algorithms are at an experimental stage where they perform very well in tightly controlled laboratory settings [3][4].

We focused on three clustering techniques to recognize human posture and the results from these are analyzed to determine which is most suitable for the purpose of posture recognition. In the long term we believe that the correct classification of these ‘static’ postures can be projected onto a time frame and combined with other algorithms to recognize activities such a ‘falling down’ or ‘tripping’.

3 Problem Solution

The aim here was to investigate clustering algorithms such as K-means, fuzzy C means and Self Organizing Maps (SOMs) to recognize human posture accurately. These techniques have the advantage of being simple to implement and practical for real-time applications. The input are silhouettes generated by the pre-processing and feature extraction steps.

Pre-processing is necessary because the data acquired from the external environment always contains noise, lighting and other distortions [5]. In the pre-pre-processing step, measures (such as background subtraction, noise reduction and shadow suppression) are taken so that the input is ready to be used by the feature extraction step. In this phase it is also necessary to reduce the size or dimensions of the input data because this will enable the data to be processed faster as well as represent it using the most significant features.

In recent years human behaviour and activity recognition has received a lot of interest due to its potential applications in the area of surveillance and security. A fast and reliable approach is presented by Spagnolo et al. [6] to estimate body postures in outdoor visual surveillance. The sequences of images coming from a static camera is trained and tested for recognition. The system uses a clustering algorithm and therefore manually labeling of the clusters is required after the training stage. The features extracted are the horizontal and vertical histograms of binary shapes associated with humans. After training the Manhattan distance is used for building clusters and for recognition. The main strengths of their method is high classification performance and small computational time which allows the system to perform well in real time. On the other hand Fuzzy C-Means (FCM) has been used by Korde et al. [7] to classify hand gestures. Testing results reveal a recognition accuracy of 85.83%. The main strength of the system is fast speed and an acceptable level of accuracy.

Finally Buccolieri et al. [8] used active contours and neural networks for their posture recognition system. With regards to feature extraction, localization of moving objects in the image and human posture estimation are performed. The classification is performed by the Radial Basis Functions Neural Network. Their approach has some advantages such as low sensitivity to noise, fast processing speed, and the ability to handle some degree of occlusion. However, the system is limited to recognizing only three postures, namely standing, bending and squatting postures.

However our system it is organized into two main phases: the learning phase and the recognition phase. In the learning phase, the system is trained using different variations of each posture data using the different learning algorithms. In the recognition phase, the system will recognize the different postures presented. In the beginning the system will

ISSN: 1790-5109 ISBN: 978-960-474-114-4

be trained to recognize pairs of postures. Subsequently the dataset will be expanded to include five types of postures and the system will be tested on these five postures. The system learns with a different clustering algorithm each time, for example it is first trained with K-means and then the recognition rate of the postures is obtained. Then this process is repeated with fuzzy C-means and finally with SOM.

3.1 K-Means

This is a well-known clustering technique which will cluster *n* objects based on attributes into *k* partitions, where $k < n$. Its main aim is to find the centers of natural clusters in the data and assumes that the object attributes form a vector space [9]. The objective it tries to achieve is to minimize total intra-cluster variance, or, the squared error function as shown:

$$V = \sum_{i=1}^{k} \sum_{x_j \in S_i} \left(x_j - \mu_i\right)^2 \quad \text{……………... (1)}$$

where there are k clusters S_i, i = 1, 2, ..., k, and μ_i is the centroid or mean point of all the points $x_j \in S_i$.

3.2 Fuzzy C-Means (FCM)

This is a data clustering technique where every data point in the dataset belongs to every cluster to a certain degree. It starts with an initial guess for the cluster centres, which are intended to mark the mean location of each cluster. The initial guess for these cluster centres is most likely incorrect. Next, FCM assigns every data point a membership grade for each cluster [10].

By iteratively updating the cluster centers and the membership grades for each data point, FCM iteratively moves the cluster centres to the right location within the data set. This iteration is based on minimization an objective function that represents the distance from any given data point to a cluster centre weighted by that data point's membership grade. The objective function is:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2 \quad , \quad 1 \le m < \infty \quad \text{…….. (2)}$$

where u_{ij} is the degree of membership of x_i in the cluster j. Essentially, FCM is seen as a derivative of the K-means. Instead of having each data point belonging completely to only one cluster, the data points in FCM belong to each and every cluster with varying degrees of association.

3.3 Self-Organizing Maps (SOM)

Self-Organizing Map (SOM) helps to understand high dimensional data by reducing the dimensions of the data to a map. The map preserves the topological properties of the input space. This makes SOM useful for visualizing low-dimensional views of high-dimensional data. In addition SOM represents clustering by grouping similar data together. Clustering in this case, is performed by having several units compete for the current object.

Like most artificial neural networks, Self-Organizing Maps operate in two modes: training and mapping. Training builds the maps using input examples and this is done through a competitive process known as vector quantization. Mapping automatically classifies a new input vector. Associated with each node in the network is a weight vector of the same dimension as the input data vectors and a position in the map space. The nodes in the SOM are placed in a regular fashion in either a hexagonal or rectangular grid.

The procedure for placing a vector from data space onto the map is to find the node with the closest weight vector to the vector taken from the data space and to assign the map coordinates of this node to the input vector. The neurons with the weight vector most similar to the input is called the best matching unit (BMU). The weights of the BMU and neurons close to it in the SOM lattice are adjusted towards the input vector. The magnitude of the change decreases with time and with distance from the BMU. The update formula for a neuron with weight vector Wv(t) is:

$$\mathbf{Wv}(t+1) = \mathbf{Wv}(t) + \Theta(v, t)\, \alpha(t)(\mathbf{D}(t) - \mathbf{Wv}(t)), \quad \text{…(3)}$$

where α(t) is a monotonically decreasing learning coefficient and D(t) is the input vector [10][11].

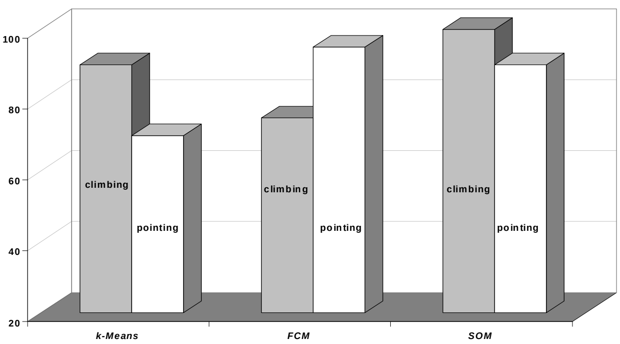
The goal of learning in self-organizing maps is to cause different parts of the network to respond similarly to certain input patterns. The network must be fed a large number of example vectors that represents, as close as possible, the kinds of vectors expected during mapping. The examples are usually administered several times and the training utilizes competitive learning.

4. Experimental Results

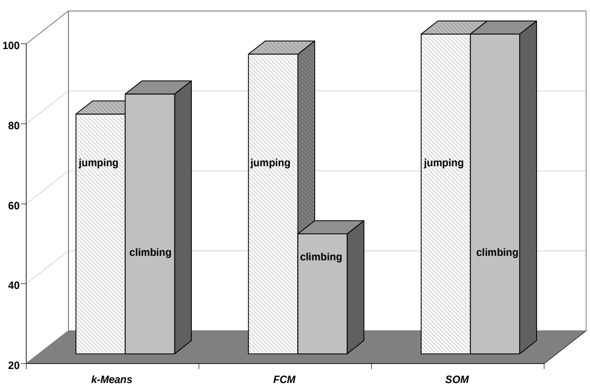
The recognition accuracy of postures was obtained by testing on a dataset of fixed size. Each test image contained in the dataset is 50x50 pixels in size. Each of the learning algorithms was trained with 100 samples per postures. The system was tested on a different set of test images. The recognition accuracy was calculated using the following formula:

Recognition accuracy = no. of correctly recognized postures / total no. of images

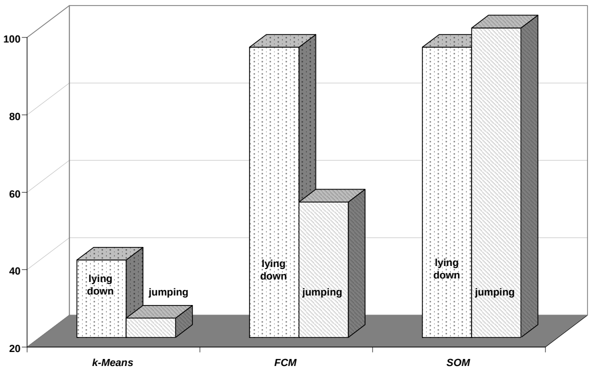
The figures 2(a) to 2(c) show the recognition rate of the three pairs of human postures with each of the clustering techniques.



(a) Climbing and pointing



(b) Climbing and jumping



(c) Jumping and lying down

Fig. 2 Results for various pairing of postures

The experimental results indicate that the k-means and FCM were not able to identify the postures well, even though there are only two clusters in each case. Other than the two postures of lying down and jumping which are quiet distinctly separable, the simpler k-Means have been able to identify the climbing pose at more than 80%. It is interesting to note that both the FCM and k-Means show poor results when identifying the lying down and jumping postures.

In the next phase the system was trained and tested with all five postures and the results are shown in figure 3. With five postures, the SOM has achieved significantly better results than either the k-Means and FCM. It was also able to identify the climbing and fighting postures very well. As in the previous results, it has not been able to recognize the lying down posture well, even though this particular posture is significantly different from the rest.

Future work will investigate how recognition accuracy of the pair of lying down and jumping postures, even though quite distinctive, can be further improved. In addition we intend to investigate the effect of using a better selection of feature set to represent the postures, will have on improving the recognition accuracy of the system.

ISSN: 1790-5109

ISBN: 978-960-474-114-4

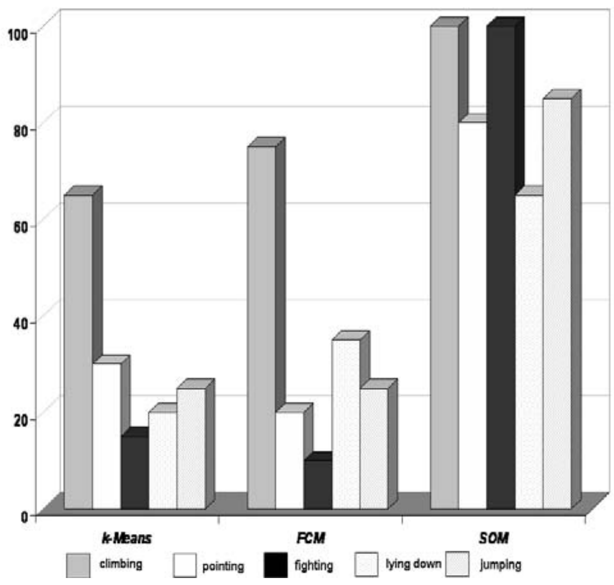


Fig. 3 Results from 5 different postures

5. Conclusion

Automated surveillance systems are increasingly being required to analyze human behaviour to detect any suspicious activity. Being able to correctly recognize human postures contributes towards the detection of such an activity. We investigated three clustering techniques to recognize human postures namely K-Means, fuzzy C-Means and Self-Organized Maps. The system was first trained to recognize pairs of postures and this was carried out for three different pairs of human posture. Finally the system was trained to recognize five postures together. The results show that K-Means and FCM perform well for the three pair of posture data. However these two clustering techniques gave low recognition accuracy when tested for all five postures. Self-Organizing Maps produced the highest recognition accuracy when tested using the three pair of human postures. Likewise it performed very well in recognizing the more complex dataset comprising of five human postures.

References:

[1] B. Boulay, F. Bremond and M. Thonnat, “Human Posture Recognition in Video Sequence”, *Joint IEEE International Workshop on Visual Surveillance, Performance Evaluation of Tracking & Surveillance (VS-PETS)*, pp.23-29, 2003

[2] V. Girondel, L. Bonnaud, A. Caplier and M. Romaut, “Static Human Body Postures Recognition in Video Sequences Using the Belief Theory”, *International Conference on Image Processing (ICIP 2005)*, Genoa, Italy, 2005.

[3] P. Guo, Z. Miao, “Projection Histogram Based Human Posture Recognition”, *Proceedings of 8th International Conference on Signal Processing*, Guilin, China, 2006.

[4] V.Kellokumpu, M. Pietikainen, J.Heikkila, “Human Activity Recognition using Sequences of Postures”, *MVA2005 IAPR Conference on Machine Vision Applications*, May 16-18, 2005 Tsukuba Science City, Japan

[5] E. Herrero-Jaraba, C. Orrite-Urunuela, F. Monzon, D. Buldain, "Video-based human posture recognition," *Proceedings of the 2004 IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety, 2004. CIHSPS 2004.*, July 2004

[6] P. Spagnolo, M. Leo, A. Leone, G. Attolico, and A. Distante, “Posture Estimation in Visual Surveillance of Archaeological Sites”. *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*, 2003.

[7] S.K. Korde, and K. C. Jondhale, “Hand Gesture Recognition System Using Standard Fuzzy C-Means Algorithm for Recognizing Hand Gesture with Angle Variations for Unsupervised Users”. *Proceedings of the 2008 First international Conference on Emerging Trends in Engineering and Technology - Volume 00* (ICETET), 2008.

[8] F. Buccolieri, F. Distante, C. Leone, A., "Human posture recognition using active contours and radial basis function neural network," *IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS 2005),* 2005.

[9] L. Costa, and R. M. Cesar, *Shape Analysis and Classification: Theory and Practice*, CRC Press, 2001

[10] R. O. Duda, P.E. Hart, D. G. Stork, *Pattern Classification*, Wiley-Interscience; 2nd edn. 2000.

[11]Alpaydin, E., *Introduction to Machine Learning*, The MIT Press, 2004