

Human Behavior Classification Using Multi-Class Relevance Vector Machine

B. Yogameena, S. Veera Lakshmi, M. Archana and S. Raju Abhaikumar
Department of Electronics and Communication Engineering,
Thiagarajar College of Engineering, Madurai, Tamil Nadu, India

Abstract: Problem statement: In computer vision and robotics, one of the typical tasks is to identify specific objects in an image and to determine each object's position and orientation relative to coordinate system. This study presented a Multi-class Relevance Vector machine (RVM) classification algorithm which classifies different human poses from a single stationary camera for video surveillance applications. **Approach:** First the foreground blobs and their edges are obtained. Then the relevance vector machine classification scheme classified the normal and abnormal behavior. **Results:** The performance proposed by our method was compared with Support Vector Machine (SVM) and multi-class support vector machine. Experimental results showed the effectiveness of the method. **Conclusion:** It is evident that RVM has good accuracy and lesser computational than SVM.

Key words: Video surveillance, pose, multi class, relevance vector machines, support vector machine

INTRODUCTION

Automatic visual surveillance systems could play an important role in supporting and eventually replacing human observers. To become practical, this system needs to distinguish people from other objects and to recognize individual persons with a sufficient degree of reliability, depending on the specific application and security level. These applications allow to the task of estimating the pose of an articulated subject.

Pose estimation is applicable in the real time surveillance applications to analyze the unusual activity of an individual from the normal ones. Many methods based on computer vision have been proposed in the literature to classify people's posture earlier (Agarwal and Triggs, 2004; 2005). Ardizzone *et al.* (2000) has proposed a pose classification algorithm using support vector machine to classify different poses of the operator's arms as direction commands like turn-left, turn-right, go-straight and so on. In (Lin and Zhang, 2009) the authors proposed a method that describes a system which delivers robust segmentation of the hand gestures using a multi-class classification of the hand gestures using a SVM ensemble. Thayanathan *et al.* (2008) estimate the pose of an articulated object from a single camera using relevance vector machine. In Lian and Lu, (2006) presented a novel approach to multi-view gender classification considering both shape and

texture information to represent facial image and the classification is performed using SVM. These studies do not concentrate on full body pose and Relevance Vector Machine utilizes only relevance vectors for the training not like SVM which uses more number of vectors.

In this study a learning based approach with a particular specialization that denote the 'Relevance Vector Machine' (RVM), a model of identical functional form to the popular and state-of-the-art 'Support Vector Machine' (SVM) is introduced for the classification of human full body poses.

The study is organized as follows. The methodology, presents background subtraction and edge detection, the classification of poses using the relevance vector machine and the experimental results for classification and comparative study. Finally the conclusion is summarizes.

MATERIALS AND METHODS

Figure 1 shows an outline of the pose classification system. Each image from the camera is forwarded to the pre-processing module where the background subtracted image sequences and the edges are obtained. Then the Relevance Vector Machine classification scheme classifies the given image sequence as different kind of poses such as standing, walking and running.

Corresponding Author: B. Yogameena, Department of Electronics and Communication Engineering,
Thiagarajar College of Engineering, Madurai, Tamil Nadu, India

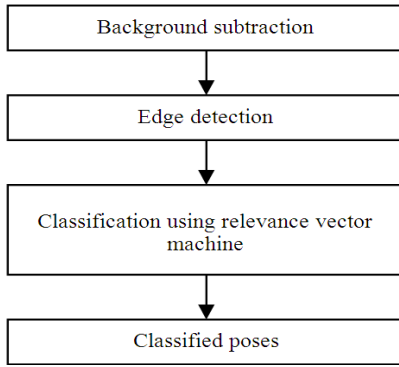


Fig. 1: System overview

Background subtraction: The first stage of video surveillance systems seeks background to automatically identify people, objects, or events of interest in various changing environments. The difficult part of background subtraction is not the differencing itself, but maintenance of a background model and its associated statistics.

Foreground detection then identifies pixels in the video frame that cannot be adequately explained by the background model and produces them as a binary candidate foreground mask. Piccardi (2004) have reviewed a number of background subtraction approaches. Wren *et al.* (1997) have proposed a statistical method, in which a single Gaussian function was used to model the distribution of background. Later Mittal and Paragios (2004) have proposed a novel kernel based multivariate density estimation technique that adapts the bandwidth according to the uncertainties. There are some of the practical issues concerning the use of the existing algorithm based on mixtures of Gaussians for background segmentation in outdoor scenes, including the choice of parameters (Stauffer and Grimson, 1999). The proposed system analyses the choice of different parameter values and their performance impact are obtained to get robust background model. The foreground detection used here is the Gaussian Mixture Model (GMM) which can deal with slow illumination changes, periodical motions from clutter background, (Piccardi, 2004). It describes K Gaussian distributions to model the surface reflectance value and is represented by Eq. 1:

$$P(X_t) = \sum_{i=1}^k \omega_{i,t} * \eta(X_t, \mu_{i,t}, \sum_i, t) \quad (1)$$

Where:

- K = The number of distributions
- $\omega_{i,t}$ = An estimate of the weight of the i^{th} Gaussian

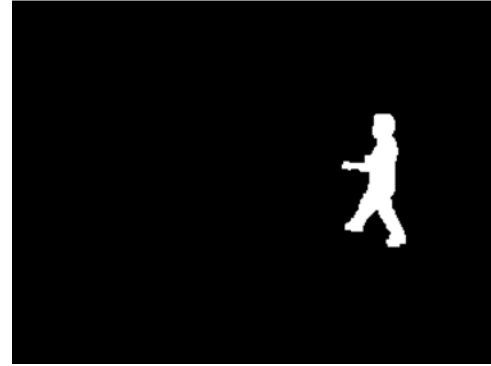


Fig. 2: An example of the background subtracted image

- μ = The mean
- \sum_i, t = The covariance matrix of the i^{th} Gaussian
- η = A Gaussian probability density function

Where:

$$\eta_{X_t, \mu_{i,t}, \sum_i, t} = \frac{1}{(2\pi)^{2/\sum_i} \sqrt{|\sum_i|}} e^{-\frac{1}{2} (X_t - \mu_{i,t})^T \sum_i^{-1} (X_t - \mu_{i,t})} \quad (2)$$

Assume the covariance matrix is of the form:

$$\sum_k = \sigma_k^2 I \quad (3)$$

This means that the red, green and blue reflectance components of the surface are independent and have the same variances, which can reduce costly matrix computation.

The weight $\omega_{i,t}$ is adjusted as shown in Eq. 4:

$$\omega_i = (1 - \alpha)\omega_{i-1} + \alpha(M_i) \quad (4)$$

Where:

- α = The learning rate
- $M_i = 1$ for the model which is matched and 0 for others

After ordering the Gaussians, the first B distributions are chosen as the background model as shown in Eq. 5:

$$B = \arg \min_b (\sum_{k=1}^b \omega_k) T \quad (5)$$

where, T is a measure of minimum models for background.

The background subtracted image obtained using GMM is shown in Fig. 2.



Fig. 3: An example for the edge extracted image

Edge detection: Following the background subtraction the edge features are extracted using the Canny edge detector (Thayanathan *et al.*, 2008). The Canny method finds edges by looking for local maxima of the gradient of the given input image. The gradient is calculated using the derivative of a Gaussian filter is given by the following relations (6) and (7):

$$G = \sqrt{G_x^2 + G_y^2} \quad (6)$$

$$\Theta = \arctan\left(\frac{G_y}{G_x}\right) \quad (7)$$

Where:

G_x = Vertical direction

G_y = Horizontal direction

The method uses two thresholds, to detect strong and weak edges and includes the weak edges in the output only if they are connected to strong edges. This method is therefore less likely than the others to be fooled by noise and more likely to detect true weak edges. Figure 3 shows an example of edge extracted image using Canny edge detector.

Classification of poses using multi-class relevance vector machine: The extracted features are given to Relevance Vector machine for learning to classify poses.

Relevance vector machine: In recent years, machine learning methods have become prevalent to examine the pattern of structural data. Kernel methods such as Support Vector Machines (SVM) are widely used to classify the poses of faces, hands, different human body parts and robots. Relevance vector machine yields a formulation similar to that of a support vector machine

and it also uses hyper parameters instead of margin/Costs. The RVM is a Bayesian regression framework, in which the weights of each input (Thayanathan *et al.*, 2008; Chen and Tang, 2005; Williams *et al.*, 2003; Tipping, 2001). These hyper parameters describe the posterior distribution of the weights and are example of governed by a set of hyper parameters estimated iteratively during training. Most hyper parameters approach infinity, causing the posterior distributions of the effectively setting the corresponding weights to zero. The remaining examples with non-zero weights are called relevance vectors. It minimizes the number of active Kernel Functions to reduce computation time. No need to find “inappropriate” parameters as in SVM (Stauffer and Grimson, 1999; Allili *et al.*, 2007).

The extension of the relevance vector machine classification scheme to a multi-class relevance vector machine classification algorithm presents training of multiple mapping functions to reduce the estimation error and creates sparser template sets. Additionally, the total training time is reduced because the RVM training time increases quadratically with the number of data points and the samples are divided among the different RVMs.

In our framework, we are focused our attention on the pose of the whole human body. This kind of gestures can be described very well through the analysis of the body contour. Moreover, to provide a detailed description of a shape, one has to take into account the whole body contour. The image features were shape-contexts descriptors of silhouette points and pose estimation was formulated as a one-to-one mapping from the feature space to pose space (Wren *et al.*, 1997; Mittal and Paragios, 2004), as shown in Fig. 4. The pose of an articulated object, a full human body, is represented by a parameter vector x given in Eq. 8:

$$x = w^k \phi(z) + \xi_k \quad (8)$$

Where:

x = The input for the system

w^k = The weight of the basis function

$\phi(z)$ = The vector of the basis function

ξ^k = The Gaussian noise vector

In order to learn about the multiple RVM technique an expectation maximization algorithm has been adopted.

It is used to minimize the cost function of the multiple RVM regression shown in 9:

$$L^k = \sum C_K^{(n)} \left(y_k^{(n)} \right)^T S^k \left(y_k^{(n)} \right) \quad (9)$$

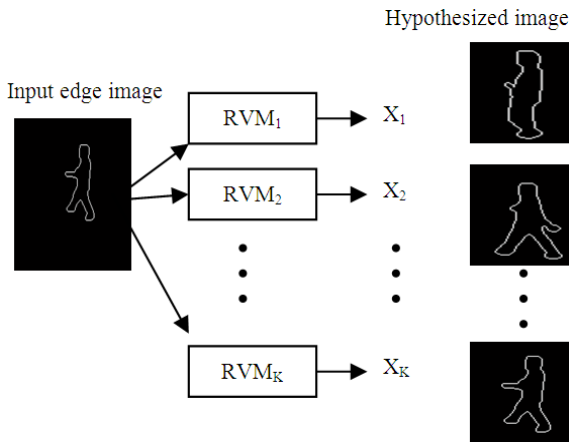


Fig. 4: Method of RVM mapping

Where:

$$y_k^{(n)} = x^{(n)} - W^k \phi(z^{(n)}) \quad (10)$$

Here:

- $Y_k^{(n)}$ = The output with n sample points belongs to the mapping function k
- W^k = The weight of the basis function
- $\Phi(z^{(n)})$ = The design matrix of vector of the basis function
- S^k = The diagonal covariance matrix of the basis function
- $C_k^{(n)}$ = The probability that the sample point n belongs to the mapping function k

RESULTS AND DISCUSION

The efficiency of the algorithm proposed has been evaluated by carrying out extensive works on the simulation of the algorithm using Matlab 7.0. In this study the video file used is both for indoor and outdoor which is captured in our college campus and image processing laboratory. The proposed method processes about 24 frames sec^{-1} for color images and the total number of frames is equal to 128 at size of 240×320 each on PC with a 2.5 GHz Pentium IV CPU to demonstrate the performance of the algorithm.

The Fig. 5 shows the background subtracted image sequences of input frames (113, 246, 320) for the classification task.

After foreground pixels obtained, canny edges are extracted from them using a derivative of a Gaussian filter as shown in Fig. 6.

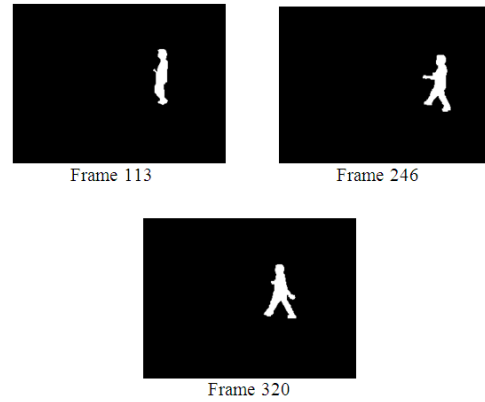


Fig. 5: Background subtracted images

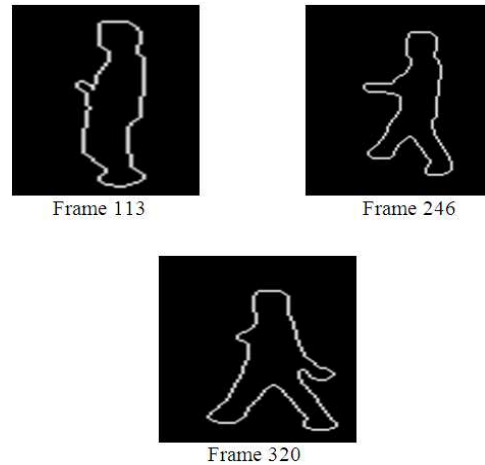


Fig. 6: Extracted edge images

Then the edge image is given as input for the classification approach. The Fig. 7 shows the initial stage of the training stage. Then the Fig. 8 shows the result after the iteration. The circles in the Fig. 8 denote the relevance vector obtained after the iteration. The maximum number of number of iterations used here is ten. Thus the resultant vector obtained here is closer to the original template and there is no misclassification of vectors has been taken place. This has been the main advantage in this proposed method.

To evaluate the proposed approach the classification accuracy has been computed and compared with existing state of art methods. On an average the performance of for the linear SVM classifier was 82.34%, for all activities, with a standard deviation of 0.56% and the performance of the proposed method was 94.67% as shown in Fig. 9. The majority of the abnormal actions are classified except the Weizmann bending. In Weizmann bending the action is classified as normal action.

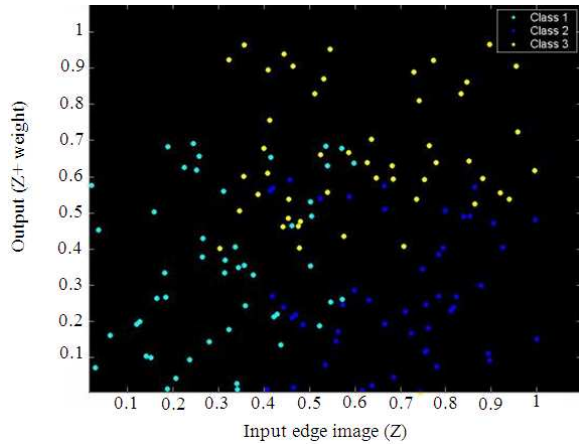


Fig. 7: Initial stage

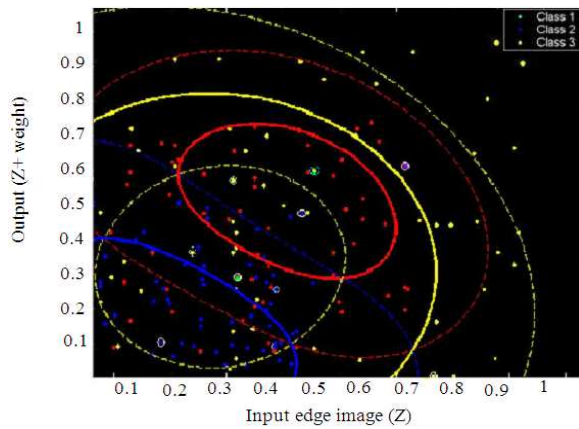


Fig. 8: Result for classification

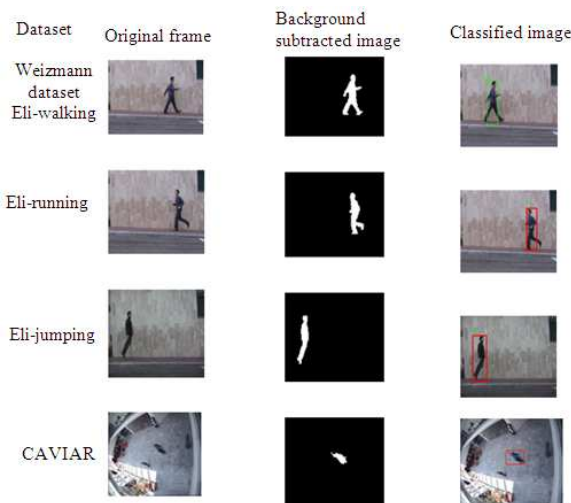


Fig. 9: Results for benchmark datasets

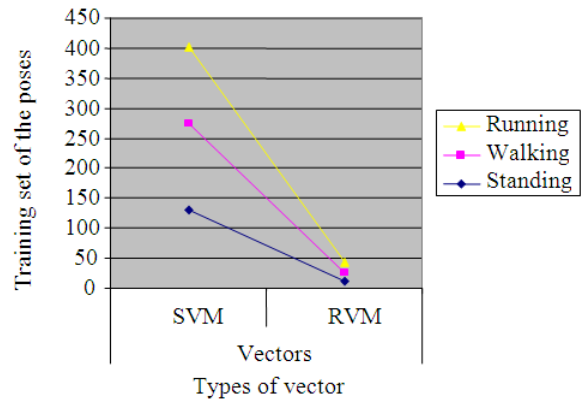


Fig. 10: Comparison of RVM and SVM vectors

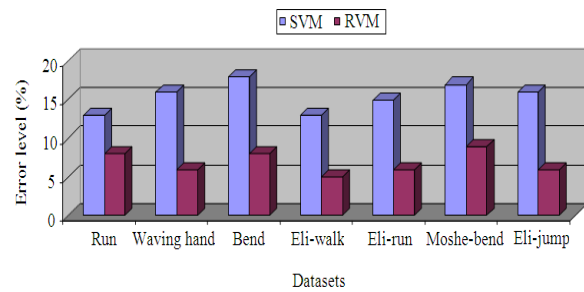


Fig. 11: Comparison of RVM and SVM errors

The results obtained with the relevance vector machine are compared with support vector machine. The number of vectors for the training is too reduced and at the same time error is also reduced. Figure 10 shows the comparison of error and training vectors are reduced for RVM than SVM. The error level for the above mentioned datasets are compared between the SVM and RVM classifiers as shown in Fig. 11. The percentage of error level is considered for both training and testing sequences of the dataset in terms of True Positive and False Negative. Higher error level leads to poor action classification. In all the datasets the percentage of error level of SVM is higher than proposed RVM classification. Moreover it is inferred from Fig. 8 that, as the number of iteration increases the better convergence of relevance vectors is achieved for the experimental datasets.

CONCLUSION

This study has introduced the framework of multi-class relevance vector machine for the classification of different kind of human poses for the application of video surveillance. To this end the given image sequences are classified as standing, running and so on.

The method has been used as a component of a system for estimating the classified poses. The experimental results have been compared between the multiclass RVM and the multiclass SVM. It is evident that RVM has good accuracy and lesser computational than SVM.

REFERENCES

- Agarwal, A. and B. Triggs, 2004. 3D human pose from silhouettes by relevance vector regression. Proceeding of the Conference on Computer Vision and Pattern Recognition, June 27-July 2, IEEE Xplore Press, USA, pp: 882-888. DOI: 10.1109/CVPR.2004.1315258
- Agarwal, A. and B. Triggs, 2005. Monocular human motion capture with a mixture of regressors. Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 25-25, IEEE Xplore Press, San Diego, California, pp: 72-72. DOI: 10.1109/CVPR.2005.496
- Ardizzone, E. and A. Chella, R. Phone, 2000. Pose classification using support vector machines. Proceeding of the International Joint Conference on Neural Networks, Aug. 6-6, IEEE Xplore Press, Como, pp: 317-322. DOI: 10.1109/IJCNN.200.859415
- Allili, M.S., N. Bouguila and D. Ziar, 2007. A robust video foreground segmentation by using generalized Gaussian mixture modeling. Proceeding of the 4th Canadian Conference on Computer and Robot Vision, May28-30, IEEE Xplore Press, Montreal, Que., pp: 503-309. DOI: 10.1109/CRV.2007.7
- Thayanathan, A., R. Navarednam, B. Stenger, P.H.S. Torr and R. Cipolla, 2008. Pose estimation and tracking using multivariate regression. *Patt. Recog. Lett.*, 29:1302-1310. DOI: 10.1016/j.patrec.2008.02.004
- Lian, H.C. and B.L. Lu, 2006. Multi-view gender classification using local binary patterns and support vector machines. *Lecture Notes Comput. Sci.*, 3972: 202-209. DOI: 10.1007/11760023_30
- Lin, Y. and P. Zhang, 2009. Vision-based human computer system using hand gestures. Proceeding of the International, Conference, on Computational Intelligence and Security, Dec. 11-14, IEEE Computer Society, Beijing, China, pp: 529-532. DOI: 10.1109/CIS.2009.29
- Mittal, A. and N. Paragios, 2004. Motion-based background subtraction using adaptive kernel density estimation. Proceeding of the Computer Society Conference on Computer Vision and Pattern Recognition, June 27-July 2, IEEE Xplore Press, Washington DC., USA., pp: 302-309. DOI: 10.1109/CVPR.2004.164
- Piccardi, M., 2004. Background subtraction techniques: A review. Proceeding of the IEEE International Conference on Systems, Man and Cybernetics, (SMC'04), IEEE, USA., pp: 3099-3104. DOI: 10.1109/ICSMC.2004.1400815
- Stauffer, C. and W.E.L. Grimson, 1999. Adaptive background mixture models for real time tracking. Proceeding of the IEEE International Conference Computer Vision and Pattern Recognition, June 23-25, IEEE Xplore Press, Fort Collins, Colorado, pp: 246-252. DOI: 10.1109/CVPR.1999.784637
- Chen, Z. and H. Tang, 2005. Sparse Bayesian approach to classification. Proceeding of the International, Conference, on Networking, Sensing and Control, Mar. 19-22, IEEE Xplore Press, USA., pp: 914-917. DOI: 10.1109/ICNSC.2005.1461315
- Tipping, M.E., 2001. Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.*, 1: 211-244. DOI: 10.1162/15324430152748236
- Williams O., A. Blake and R. Cipolla, 2003. A sparse probabilistic learning algorithm for real-time tracking. Proceeding of the 9th International Conference on Computer Vision, Oct. 13-16, IEEE, Nice, France, pp: 353-360.
- Wren, C.R., A. Azarbayejani, T. Darrell and A.P. Pentland, 1997. Pfunder: Real-time tracking of the human body. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19: 780-785. DOI: 10.1109/34.598236