Intelligent Discipline Maintenance System with Client-Server Topology

Osama Nasir[1] , Muhammad Maaz[2] , Tahir Mehmood[3] , Waqar Ahmad[4]
Department of Electrical Engineering, UET, Lahore, Pakistan
Email: [1]osama.nasir63@gmail.com, [2]mmaaz60@gmail.com, [3]tahir.janjua97@yahoo.com, [4]m.waqar487@gmail.com

***Abstract*— This paper describes the problem of automatic discipline violation detection and client server communication with the authorities. A complete detection and reporting system is designed and evaluated on an audio database containing 150 audio signals classified among three different classes and an image database containing more than 300 sampled images distributed among 2 different classes. Special emphasis was given to robust techniques so that the system could also work independently in real time. The detection of discipline anomaly is done using both audio and video signals. Feature extraction and Dimensionality reduction are performed on the audio sample to form the feature vector. Face recognition, selection of region of interest and feature extraction are performed on video signal to form its respective feature vector. These feature vectors are then classified using rbf-kernel SVM for the former and One-class SVM for the latter. Client-Server topology is used for communication with authorities. Our results have shown that the system is capable to classify the underlying discipline anomalies with more than 85% accuracy in real time.**

Keywords*— *feature extraction, dimensionality reduction, face recognition, classification, Support Vector Machines, Client-Server topology

I. Introduction

Intelligent Discipline Maintenance System monitors non-disciplinary actions like door slamming, whistling, screaming and smoking. The system is divided into two major parts. One is responsible for audio processing and other for video processing.

The Audio detection and recognition system receives real-time audio from a microphone, extract its features and form a feature matrix for that sound sample. The extracted features are Short Term Energy (STE), Zero Crossing Rate (ZCR) and Mel-Frequency Cepstral Coefficients (MFCC). In next stage, Principal Component Analysis (PCA) is used to extract more descriptive features and reduce dimensionality of the feature matrix. Finally, SVM using Gaussian radial basis function rbf-kernel is trained and later used to recognize sounds like door slamming, whistling and screaming.

The video detection and recognition system detects smoking actions using image processing and machine learning. It receives real-time video data and extracts video key-frames. Then Face detection algorithm detects face region from these frames only. Face is detected using a Haar Classifier. Then we extract Histogram of Oriented Gradients (HOG) and Linear Binary Patters (LBP) features and input them to a One-Class SVM classifier. The output of the classifier will determine that the event should either be declared as smoking or non-smoking. When some non-disciplinary action is detected, the system automatically reports the server using a Client-Server topology. The server will inform the authorities to take the desired action.

The next section discusses what people have done so far to solve the underlying problem. Section (3) explains briefly about basic methodology and working principles of our implemented system, by dividing it into two sub systems. The experimental setup including hardware components and other experimental details is discussed in Section (4). Finally, in section (5) all the experimental results recorded during training and testing of our project are discussed, and some conclusions regarding that are mentioned in section (6).

II. Related Work

Recently, in the past 5 to 10 years people have started shifting their scope towards Artificial Intelligence and Machine Learning and many have started using intelligent machines to solve related problems. Nowadays many working systems are available which can detect and identify any desired anomaly. For example, some researchers have implemented impulsive sound detection and identification systems which can classify between glass break, gunshot, whistling, screaming, etc. The working principle of these systems are based on training and testing a machine learning classifier using a pre-prepared dataset.

Some more sophisticated machine learning example include image processing and video processing. Researchers have also started working on detecting specific actions from image or a video frame. For example, some people have implemented smoking detection systems without the aid of any smoke detectors. The system take the input in the form of a video and classify the performed action using machine learning algorithms. Some algorithms use object detection and smoke detection to classify smoking while others may use motion detection on already identified object to do the job. Although many such audio and video processing systems have been implemented individually, no one has ever combined these systems to solve the bigger problem of the discipline maintenance.

III. Our Proposed Solution

Our proposed approach also uses machine learning techniques to solve the underlying problem. To make problem simpler, we have divided the bigger problem into three different sub problems. The first sub problem is the Noise reduction by detection of discipline anomalies due to sound signals. The second sub-problem is the detection of a smoking event from a video signal recorded by a camera. Our last sub problem would be to report the server when any such activity is detected, so that the server can communicate with authorities accordingly.

A. *Noise Reduction:*

The problem of Noise reduction can be done by using a Machine Learning approach. The first step, to design a trained machine capable to classify any noise, is to collect a sample data set for the machine on which it is trained. Data is collected from real time scenarios. The sample data contains various categories of sound like shouting, screaming, whistling and door slamming. Majority of this data is used to train the machine while some of it will be used later to test the working of the system. The second step of developing our intelligent discipline maintenance machine is to train the system on the sampled dataset. The audio signal is processed through different stages to classify it as a noise signal. To determine if a sound is noise or not, we need its feature extraction. Audio signal is passed through different filters including Short Term Energy (STE), Zero Crossing Rate (ZCR) and Mel-Frequency Cepstral Coefficients (MFCC) which extracts the required features of the audio signal. First a rbf-kernel SVM model is trained using the training samples. For test samples, the extracted features are fed to the trained model. We can now classify these samples easily as a non-noise or noise sound i.e. a discipline violation.

Once classified the system will send a signal to the main server for the detected discipline violation. In the last phase, the system is tested on some random test data sets and the output is analysed, to check the accuracy of our trained model.

B. *Smoking Detection:*

Smoking Detection can be done like the noise reduction problem. For this we will train a machine learning algorithm. The first step is to collect real-time video data and then analyse it using different machine learning techniques. Different types of gestures for smoking are included in the dataset to increase the accuracy of the machine. The system is then trained on the sample data set. The recorded video signal is fed to the system which is processed through different stages to classify its gesture. To determine a smoking gesture, we need video frame extraction. Then face detection is performed to select the region of interest. From this region, we extract Histogram of Oriented Gradients (HOG) and Linear Binary Patters (LBP) features from the given video sample and train a One-class SVM classifier. With this model, we can decide whether the given gesture matches to that of smoking or not. Once such action is classified by the system it will automatically report the main server using a Client-Server topology. The server will then inform the authorities.

C. *Client-Server Topology:*

The client-server characteristic describes the relationship of cooperating programs in an application. The server component provides a function or service to one or many clients, which initiate requests for such services. Servers are classified by the services they provide. For example, a web server serves web pages and a file server serves computer files. A shared resource may be any of the server computer's software and electronic components, from programs and data to processors and storage devices. The sharing of resources of a server constitutes a service.

In our project, the microprocessor acts as a client and a laptop computer acts as a server for this client. Whenever the client detects any discipline violation, it initializes a communication with the server and send a signal to the server regarding the discipline anomaly. Based on this signal the laptop i.e. server informs the authorities to take some action.

The combined flow diagram of all the modules running parallel in the systems is shown in figure below,

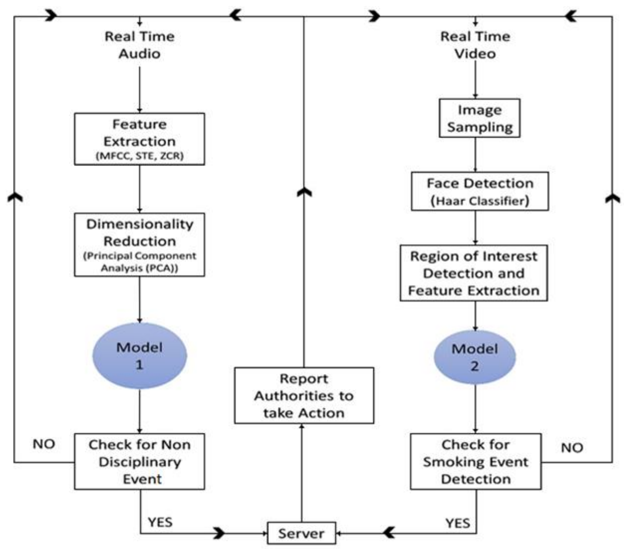

Fig. 1. Flow Diagram of the complete working system

IV. Experiments

A. *Audio Processing System:*

In our database all the sound samples were recorded at 44.1 kHz frequency using single channel and later they were truncated to exactly 5 seconds each. We have extracted 26 MFCCs, STE and ZCR using 20msec overlapping windows, for each audio sample, for training the classifiers.

In our case, the actual number of features extracted from our raw database were 2600 features from MFCCs, 250 features from STE and 250 features from ZCR, thus our final feature vector has a total of 3100 features per sample. Using PCA, we have successfully reduced our 3100-dimensional feature space into only 70-dimensional feature space having only 70 principal components such that the variance of our given data is maximum on these principal components.

Using three binary KNNs to train an audio discipline identifier, we achieved 89.45% accuracy using all features, and only 88.41% accuracy using 70 PCA derived features. For multi-class KNN, we achieved 89.72% accuracy using all features, and only 89.33% accuracy using only 70 PCA components. Using other techniques (like SVM classifier) we improved these accuracies.

We have trained our dataset using SVM machine. As the data separation seems non-linear, the linear SVM gave very poor results. But the non-linear SVM with radial basis function (rbf) kernel function gave much better testing accuracies. 94.26% accuracy was achieved using all the features, while 94.12% accuracy was achieved using 70 PCA features. Using multi-class SVM did not actually improved these accuracies, but had a significant effect on real-time results. A comparison of the two classifiers show us that using multi-class SVM with PCA is a wise choice.

B. *Real-Time Testing for Audio Processing System:*

In the last stage of developing our final model for Audio Processing System, we tested all the trained classifiers using our test set. At the end, the performance of all classifiers is evaluated and compared. The classifier with better testing accuracy is preferably used. In our case the testing accuracy for SVM trained model was highest.

The classifier with the best accuracy on test set is usually chosen for real time implementation. Sometimes both test-set accuracy and real-time test accuracy are used for comparison between different trained models. In our case the multi-class SVM with rbf kernel gave maximum accuracies, 93.6% for test-set and 89% for real time data.

C. *Video Processing System:*

All the video in our database are filmed at around 30 FPS. To convert our problem from video processing domain to image processing domain, we have transformed our database into image key-frames extracted from each video. We have used the face detection implemented by Haar classifier to detect the face region of all the images present in the database. This is used to select our region of interest. Then we extracted LBP and HOG features for all the images present in the data set. This gave us a feature matrix to train our classifiers.

We trained our One-Class SVM classifier for those images which contain smoking action only. When an image is passed to this classifier, it will check if this image belongs to the one class defined by the model. If not, the image will be classified as non-smoking, otherwise it will be classified as smoking. Using this one-class SVM model we achieved 88.6% accuracy.

D. *Real-Time Testing for Video Processing System:*

Once we have trained our classifiers, the next step is to check how efficiently these classifier works. For this purpose, we tested our trained classifiers against different test sets. At the end, the performance of all classifiers is evaluated and compared. The classifier with better testing accuracy is preferably used. In our case the testing accuracy for One-Class SVM trained model was highest.

Sometimes both test-set accuracy and real-time test accuracy are used for comparison between different trained models. In our case the SVM with rbf kernel gave maximum accuracy of 87% for test-set and 83% for real time data while the the One-class SVM gave maximum accuracy of 88.6% for test-set and 84.85% for real time data.

V. RESULTS

A. *Experimental Results for Audio Processing System:*

For the Audio Processing system, we have trained all the classifiers using three one-versus-one binary classifiers, and a single multi-class classifier with and without Feature extraction technique (PCA). The test set and real-time accuracies for all these classifiers are given in table below,

TABLE I. TEST ACCURACIES FOR BINARY CLASSIFIERS OF AUDIO PROCESSING SYSTEM

Classifier	**Without PCA**		**With PCA**	
	Test Data Accuracy	Real Time Accuracy	Test Data Accuracy	Real Time Accuracy
SVM	94.26	83.86	94.12	85.35
KNN	89.45	79.39	88.41	81.02

The test set accuracies of binary class SVM and KNN classifier using multiple number of PCA Components for the trained model are shown in table below,

TABLE II. TEST ACCURACIES FOR BINARY CLASSIFIERS USING PCA

PCA Components	10	20	30	40	50	60	70 and above
SVM	22.77	38.14	54.96	71.20	82.37	87.92	94.20
KNN	32.53	46.81	56.38	74.12	82.85	88.31	88.43

We tested our binary SVM and KNN classifier multiple times, using different number of PCA components. The following graph showing their accuracy comparison,

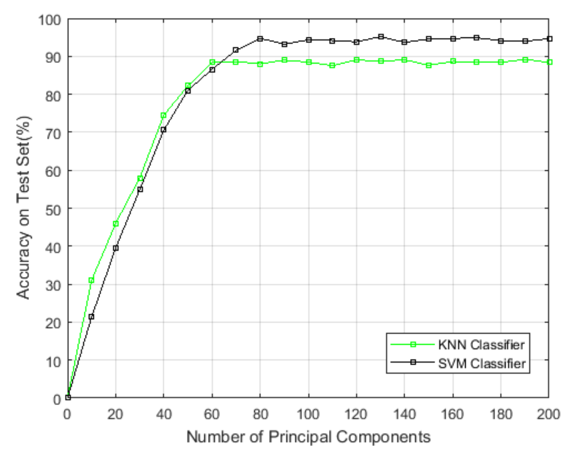

Fig. 2. Comparison of Accuracy for SVM and KNN Binary Classifiers using PCA

Following table contains the Test data set accuracy and real-time accuracy using a single Multi-class SVM and KNN classifier without any feature extraction technique along with the Test data set accuracy and real-time accuracy of the same classifiers using PCA as a feature extraction technique.

TABLE III. TEST ACCURACIES FOR MULTI-CLASS CLASSIFIERS OF AUDIO PROCESSING SYSTEM

Classifier	Without PCA		With PCA	
	Test Data Accuracy	Real Time Accuracy	Test Data Accuracy	Real Time Accuracy
SVM	92.48	87.29	93.61	89.14
KNN	89.72	83.20	89.33	83.85

The test set accuracies for Multi-class SVM and KNN classifier using multiple number of PCA Components for the trained model are shown in table below,

TABLE IV. TEST ACCURACIES FOR MULTI-CLASS CLASSIFIERS USING PCA

PCA Components	10	20	30	40	50	60	70 and above
SVM	22.86	37.29	53.31	74.50	87.44	92.86	93.59
KNN	39.11	52.63	67.85	74.39	86.68	89.14	89.30

We have trained and tested our multi-class SVM and KNN classifier multiple times, using different number of PCA components. The graph showing the accuracy comparison for both these classifiers is shown in figure below,

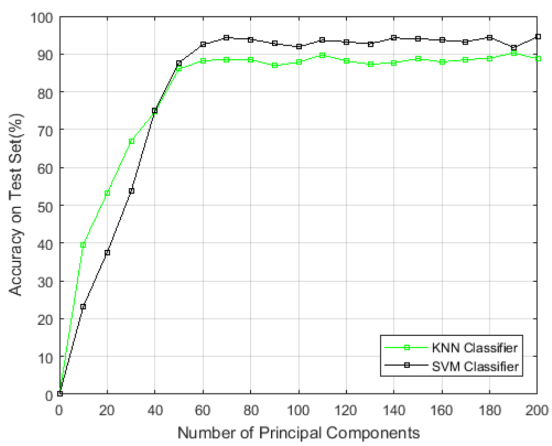

Fig. 3. Comparison of Accuracy for SVM and KNN Multi-Class Classifiers using PCA

We have trained and tested two different SVM classifiers for the Video processing system. The first SVM classifier uses an rbf kernel for classification and the second SVM classifier solves the problem using one-class classification. We have trained both the classifiers using only HOG features, LBP features and both HOG and LBP features. The following table contains the test set and real-time accuracies for all these classifiers.

TABLE V. TEST ACCURACIES FOR MULTI-CLASS CLASSIFIERS OF AUDIO PROCESSING SYSTEM

Classifier	Data Set	HOG	LBP	HOG & LBP
SVM(rbf)	Test Data	83.86	80.38	87.41
	Real-Time Data	80.61	74.18	83.92
SVM(One-Class)	Test Data	85.23	81.91	88.57
	Real-Time Data	81.76	78.04	84.85

ACKNOWLEDGMENTS

All praise to Allah Almighty, the most gracious, the benefactor of all. Because it is He who endows us with everything including knowledge, health and wealth. We would then like to appreciate Dr. Kashif Javed who guided us in understanding the concepts of Design and implementation of Intelligent Discipline Maintenance System with Client Server Topology and parameters related to it. We are grateful to him for provision of apparatus and facilities required for the project. His support at each step from beginning to the end and guidelines for project report has made us able to present our project in the best form. Secondly, we acknowledge dedication of our teachers at Department of Electrical Engineering, University of Engineering and Technology, Lahore. We would also like to thank our families and friends for their continuous encouragement and moral support.

REFERENCES

[1] R. Biondi, G. Dys, G. Ferone, T. Renard and M. Zysman. "Low Cost Real Time Robust Identification of Impulsive Signals." *International Journal of Computer, Electrical, Automation, Control and Information Engineering,* vol:8, 2014.

[2] P. Foggia, N. Petkov, A. Saggese, N. Strisciuglio and M. Vento. "Reliable detection of audio events in highly noisy environments." University of Salerno, June 2015.

[3] A. Dufaux, L. Besacier, M. Ansorge, and F. Pellandini,"Automatic Sound Detection And Recognition For Noisy Environment", University of Neuchatel, 2001.

[4] O. Lartillot and P. Toiviainen. "A Matlab Toolbox For Musical Feature Extraction From Audio." in *Conference on Digital Audio Effects (DAFx-07),* Bordeaux, France, Sep. 2007.

[5] J. Maxime. "Sound representation and classification benchmark for domestic robots." in Proceedings of *the IEEE International Conference on Robotics and Automation (ICRA),* 2014, pp. 6285-6292.

[6] J. C. Wang, H. P. Lee, J. F. Wang and C. B. Lin. "Robust Environmental Sound Recognition for Home Automation." *Automation Science and Engineering, IEEE Transactions,* vol. 1, 2008, pp. 2531.

[7] F. G. Barbosa and W. L. Silva, "Multiple Support Vector Machines and MFCCs Application on Voice Based Biometric Authentication Systems." in *IEEE International Conference on Digital Signal Processing (DSP),* 2015.

[8] V. Tiwari. "MFCC and its applications in speaker recognition. "*International Journal on Emerging Technologies",* vol. 1, 2010, pp. 19-22.

[9] S. Valencia, M. V. Smith and A. Atyabi. "Mobile ascertainment of smoking status through breath: A machine learning approach." in *Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON),* Oct. 2016.

[10] Kavitha, Chandan, Gayathri Devi, Pradeep Kumar and Reshma. "Automated System For Smoking Detection." *International Journal Of Current Engineering And Scientific Research,* vol. 4, 2017.