Performance Analysis of Audio and Video Based Person Authentication Using Machine Learning Technique

M.Tech.Scholar Ranjana Patel ,Krishna Kumar Vishwakarma
Dept. of Electronics & Communication Engineering
ranjanapatel332@gmail.com , kri_k_2006@yahoo.co.in
Rewa Institute of Technology
Rewa M.P.,India

***Abstract*-Object detection and tracking is usually the first step in applications such as video surveillance. The main purpose of the static camera's face recognition and tracking system is designed to estimate speed and distance parameters. We propose a general use of detection and tracking method that moves object based on a visual system and using an image difference algorithm. Then recognize the person's voice to get the corresponding person's feedback .This process focuses on detecting people in the scene and then performing voice signal processing the first approach is the use of the video structures to extract scene boundary candidates from shot boundaries. Then using the MCMC method to select the true scene boundaries from these candidates, highly-accurate scene segmentation becomes possible. It should be noted that when the prior probability concerning the number of scenes in a target video sequence is given correctly, the MCMC method can provide a more accurate scene segmentation result This approach based on the classification person authentication using multi SVM approach to improve classification accuracy this simulation has performed on matlab simulation.**

***Keywords*-deepspeech speaker identification, voice biometrics voiceprint, speaker recognition, voicemapgithub, speaker recognition dataset**

I.INTRODUCTION

Authentication can be defined as verifying the validity of a user by using at least one form of the identification methods. To grant access to the system, the users' identity should be verified by determining the following factors Knowledge-based factors: It is defined as what the user knows. Some of them are any forms of a password, personal identification number, answer to the secret questions and many more [8].In recent years, the field of identity recognition has received a lot of attention. There is a growing need for reliable automatic user identification systems for secure access to buildings or services. Traditional technologies based on passwords and cards have a number of drawbacks.

The password can be forgotten or the stolen card may be lost or stolen and the system will not work between the client and the clerk. Various researchers have proposed and studied many techniques to identify users through features that are difficult to forge. Biometrics refers to areas related to human identification through physiological features, fingerprints, irises, sounds, faces, etc. Biometric identification systems can be used for human identification or verification. In the verification task, the user claims a specific identity.A large number of commercial biometric systems use fingerprint surfaces or voice. Each method has its advantages and disadvantages (discriminate ability, complexity, robustness, etc.). Fingerprint verification has been used for a long time and is based on the wrinkles of the fingertips and the local features of the wrinkles. Extract and compare features called details to determine possible matches. The image quality of fingerprints is very important for detail extraction. Matching should also solve problems, such as fingertip cutting. Through voice and face recognition naturally, it is easy for end users to accept. In recent years, a lot of work has been done in face and speaker recognition, resulting in mature technologies that can be used in applications. In the last few years, automatic face recognition has witnessed many activities.

Many new technologies have been proposed. Among the technologies that represent new trends in face recognition, features can be cited, elastic graph matching, automatic correlation, and neural networks on the back. Zhang et al. Analyzed and evaluated these three technologies. This study is probably the most effective representative and the most comprehensive study, because these algorithms are analyzed under a general statistical decision framework and evaluated in a general database containing hundreds of different topics. The experimental results of this study indicate that elastic graph matching (EGM) is better than other techniques. Face detection can be considered a special case of object class detection. In object class detection, the task is to find the location and size of all objects that belong to a given class in the image.

Examples include the upper body, pedestrians and cars. Face detection algorithms focus on detecting positive faces. It is similar to image detection, where images of people match little by little.The image matches the image stored in the database. Any changes to facial features in the database invalid the matching process. A reliable face detection method based on genetic algorithm and self-surface technology: First, it detects all possible valleys of human eyes by testing all valleys in the gray area, flat image. Then use genetic algorithms to generate all possible facial areas, including eyebrows, iris, nostrils and mouth corners. And wrinkle effect caused by movement of the head. The value of fitness for each candidate is measured based on its projection on the functional plane. After several iterations, all face candidates with high fitness values are selected for further verification.

1. Possession-based factors: It is defined as what the user has. Some of them are an identification card, security token, device token or any unique hardware identifier.
Inherence-based factors: It is defined based on what the user is or how he does. Some of the physiological factors are fingerprint, iris and DNA patterns and some of the behavioral factors are biometric identifiers, signatures, voice and face Authentication can be a combination of the above. The types of authentication categories include:-
Single-factor authentication: It makes use of one factor to authenticate the user trying to login to the system. It is more prone to different cyber attacks.

1.1. Two-factor authentication: It combines any two authentication factors to increase the level of security in the system. A practical example of this implementation is the real-time banking login where some banks generate a onetime password (OTP) while the user logins in by typing the correct password. Only if the password entered is valid the OTP gets generated, and if the user enters the generated number from his device correct, he gains access to the system.

1.2. Multi-factor authentication: - It combines many authentication mechanisms to form a layered approach. The plethora of functionalities offered by multi-factor authentication includes protection from intrusion, enhancement of security, and reliable false proof system. My thesis focuses more on this multi-factor authentication to develop a robust system to identify the users via using machine learning algorithms. The idea is to add three factors of authentication by entering the correct password, verifying the device, and identifying the users typing pattern. Hundreds of models exist for classification. In fact, it’s often possible to take a model that works for regression and make it into a classification model. This is basically how logistic regression works. We model a linear response WX + b to an input and turn it into a probability value between 0 and 1 by feeding that response into a sigmoid function. We then predict that an input belongs to class 0 if the model outputs a probability greater than 0.5 and belongs to class 1 otherwise.

Another common model for classification is the support vector machine (SVM). An SVM works by projecting the data into a higher dimensional space and separating it into different classes by using a single (or set of) hyperplanes. A single SVM does binary classification and can differentiate between two classes. In order to differentiate between K classes, one can use (K – 1) SVMs. Each one would predict membership in one of the K classes.

2. Support Vector Machines: The Support Vector Machine (SVM) represents a new class of classification machines that have been effectively used for organizing [16]. It can perform the binary arrangement task. Given a set of n labeled data points {(x1, y1), ..., (xn, yn)}, where yi = ± 1, SVM learns w by thoroughly separating hyperplanes, x + b = 0 decision Function where xi∈ Rn, wrn and br In the linear case, SVM finds the hyperplane using the margin by reducing 1/2 • ‖. - → w ‖ 2, obey yi (≺w, xi + b)> = 1, i = 1, 2, ... n. In the case of linear separability, the optimal hyperplane is calculated by adding the relaxation variables εi = 1, 2, ..., n and the penalty parameter C. The optimization difficulty is expressed as: min

$$\min \frac{1}{2} \cdot ||\overrightarrow{w}||^2 + C \sum_{i=1}^{n} \varepsilon_i$$

$$subject\ to\ y_i(\prec w, x_i \succ +b) >= 1 - \varepsilon_i,\ i = 1, 2, ...n$$

Using Lagrangian formula, these optimization problems can be resolved by presenting a new unidentified scalar mutable αi (called Lagrangian multiplier), which is presented for each restraint or forms a multiplier coefficient Linear mixture. The problem is expressed as follows:

$$L_d = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i x_j$$

After solving the optimization difficult, the data points related with nonzero ai correspond to specific data (xi, yi) called provision courses. This data is used to calculate conclusion function while discarding the rest of the data. In non-linear cases, surface extrication types are not linear, so the idea is to convert the data points into added high-dimensional function space where the difficult can be separated linearly. If the change to a high-dimensional space is ϕ, Lagrange purpose can be expressed as:

$$Ln = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \varphi(x_i) \varphi(x_j)$$

The dot produce(xi) ϕ(xj) in that high dimensional universe explains a kernel function k(xi, xj). Within greatest common kernel purposes, we cite:

Linear: $x_i . x_j$,
Polynomial of degree d: $(x_i . x_j + 1)^d$,
Radial Basis Function (RBF): $exp(\frac{-||x_i - x_j||^2}{2\sigma^2})$.

Once provision courses have been resolute, SVM choice purpose has form:

$$f(x) = \sum_{j=1}^{Supportvectors} \alpha_j y_j K(x_j, x)$$

II. RELATED WORK

In existing approach, the first approach is the use of the video structures to extract scene boundary candidates from shot boundaries. Then using the MCMC method to select the true scene boundaries from these candidates, highly-accurate scene segmentation becomes possible. It should be noted that when the prior probability concerning the number of scenes in a target video sequence is given correctly, the MCMC method can provide a more accurate scene segmentation result. Therefore, in the second approach of the proposed method, the parameter utilized in the prior probability is set to the optimal value by using Multiple Regression Analysis (MRA).

III. PROPOSED SYSTEM

We propose a face tracking algorithm with temporal-spatial information and confidence trajectories. The whole process is divided into video and voice association. The high confidence path is associated with the detection result of the current frame when it is locally associated, and the low confidence path is associated with the detection result of the current frame when it is globally associated. We use a combined model for digital image processing and digital signal processing to determine correlation results. The most important steps performed are detection and identification. Automatic person recognition by machine can be categorized into person identification and authentication. The objective of a person identification system is to determine the identity of a test subject from the set of reference subjects. The performance of the person identification system is quantified in terms of identification rate or recognition rate. On the other hand, a person authentication or verification system should accept or reject the identity claim of a subject.

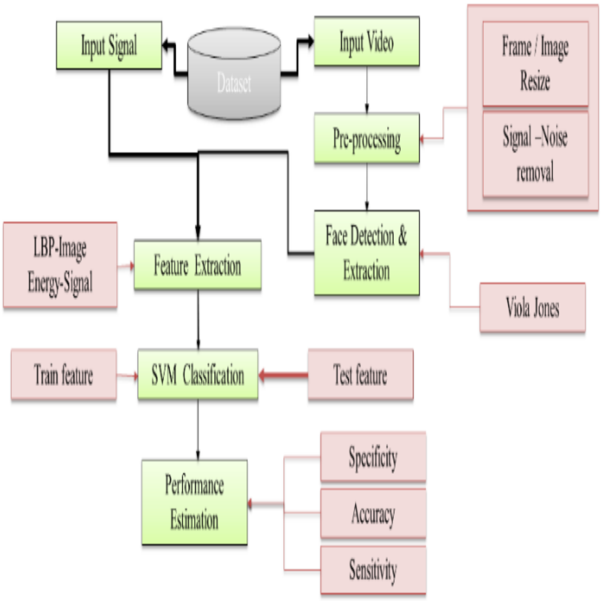


Fig. 1. proposed flow chart.

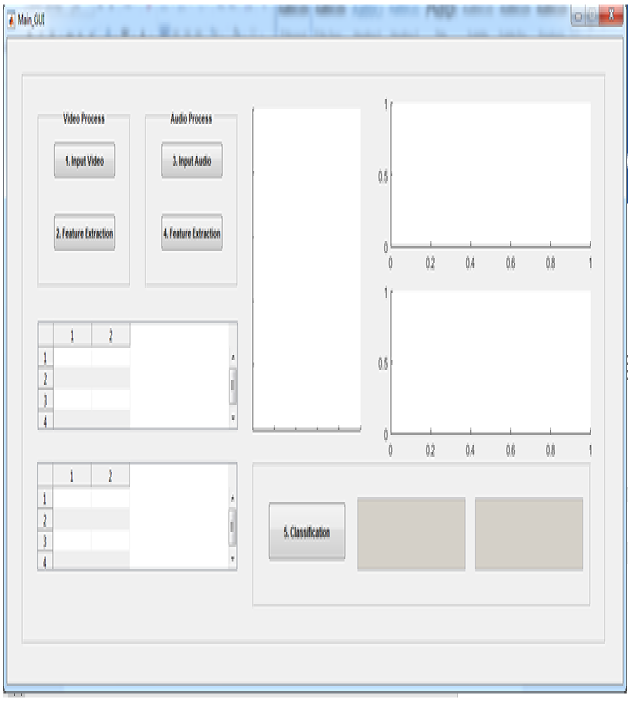


Fig 2. Input GUI windows.

1. Pre-Processing: Image processing is a method of converting an image into digital form and performing some operations on it to get an enhanced image or extract some useful information from the image. Datasets may require preprocessing techniques to ensure accurate, efficient or meaningful analysis. Image preprocessing means "preparing" samples / images to introduce them into algorithms for specific tasks, such as tracking targets, recognition, features Excerpts, etc. RGB to grayscale image:

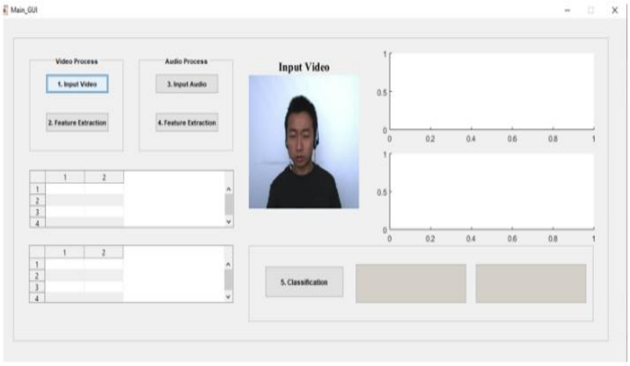
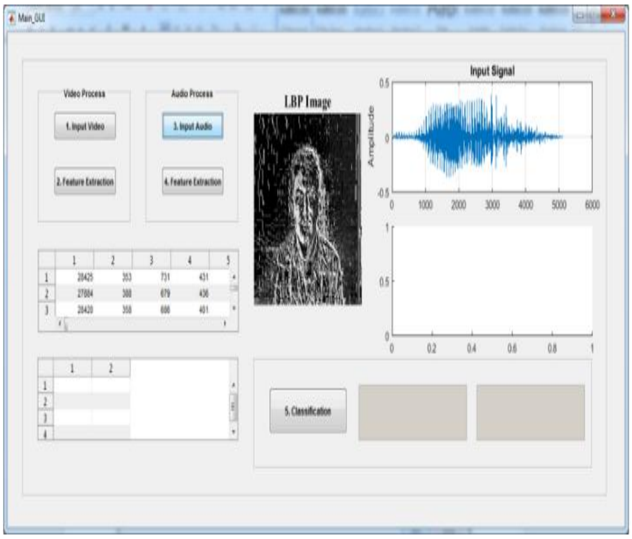


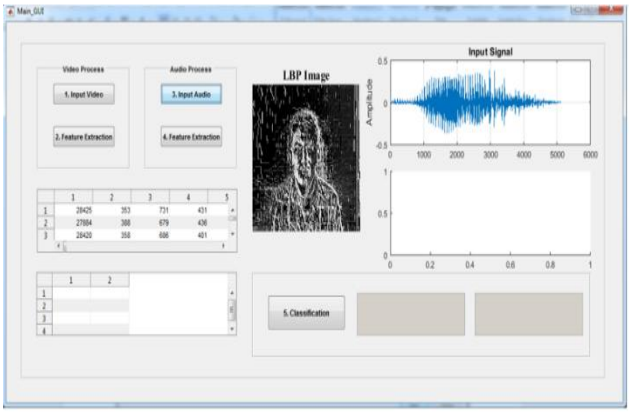
Fig 3. Input video GUI window.



	1	2	3	4	5
1	28425	353	731	431	
2	27884	388	679	436	
3	28420	358	686	401	

Fig 4. Feature Extraction Window.

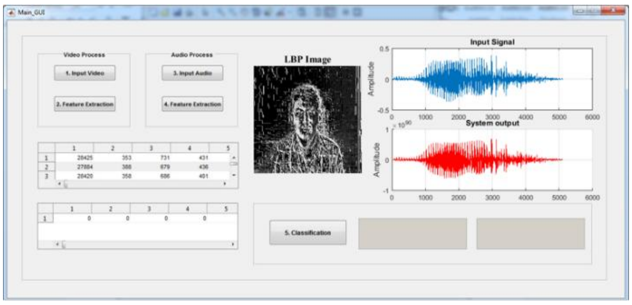
Face detection is a computer technology used in various applications to recognize faces in digital images. Face recognition also refers to the psychological process that people find and look at faces in visual scenes .Although it can be trained to detect different object categories, its main motivation is face detection.



	1	2	3	4	5
1	28425	353	731	431	
2	27884	388	679	436	
3	28420	358	686	401	

Fig 5. Audio Input GUI window.

Collect datasets with face images and audio signals. Face image and voice data sets are implemented as input. The input image is recorded in .jpg or .png format and the signal is recorded in ".wav" format



	1	2	3	4	5
1	28425	353	731	431	
2	27884	388	679	436	
3	28420	358	686	401	

	1	2	3	4	5
1	0	0	0	0	

Fig.6. Audio Feature Extraction.

In machine learning, pattern recognition, and image processing, the extraction of the feature starts with a set of initial measurement data and constructs derived information (features) designed to provide information and non-redundancy to facilitate subsequent learning and generalization steps.SVM performs mapping from input space to feature space to support nonlinear classification problems. Nuclear tricks help by allowing the lack of accurate representations of mapping features that can lead to dimensional curse. This makes the linear classification in the new room (or function room) equal to the non-linear classification in the original room (or input room). SVM realizes these functions by mapping the input vector into a higher-dimensional space (or function space) where the largest spaced hyperplane is constructed.

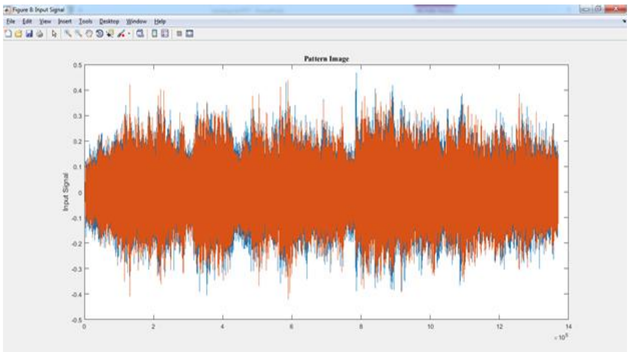


Fig.7 Audio Input Signal.

2. Modules

- Input
- Preprocessing
  - ✓ Image Resize
  - ✓ Noise removal of signal.
- Face detection & extraction
- Feature extraction
- Classification
- Performance

2.1. Input: The dataset which is collection of face images and audio signals. The face image & Speech dataset are implemented as input. The input images are taken in the format .jpg or .png and the signals are taken in the format '.wav'.

2.2. Pre-Processing: Image processing is a method to convert an image into digital form and perform some operations on it, in order to get an enhanced image or to extract some useful information from it. Data sets can require preprocessing techniques to ensure accurate, efficient, or meaningful analysis.Pre-processing of an image means "preparation" of the sample/image to introduce it to an algorithm for specified tasks like tracking targets , recognition, feature extraction, etc.RGB to Gray image: First is the conversion of RGB images into a Gray scale image using suitable command for the brighter appearance of the optic disc than its background. Proper resizing of image is also done. Image interpolation occurs when you resize or distort your image from one pixel grid to another. [7]Image resizing is necessary when you need to increase or decrease the total number of pixels. Images are often corrupted by random variations in intensity, illumination, or have poor contrast and can’t be used directly. Filtering process transforms pixel intensity values to reveal certain image characteristics.

- Enhancement: improves contrast
- Smoothing: remove noises
- Template matching: detects known patterns
- Signal Denoise : It is to remove noise from the original signal. For this task, we used Butterworth filter.

2.3. Face Detection & Extraction: Face detection can be regarded as a specific case of object-class detection. In object-class detection, the task is to find the locations and sizes of all objects in an image that belong to a given class. Examples include upper torsos, pedestrians, and cars. Face-detection algorithms focus on the detection of frontal human faces. It is analogous to image detection in which the image of a person is matched bit by bit. Image matches with the image stores in database. Any facial feature changes in the database will invalidate the matching process.

A reliable face-detection approach based on the genetic algorithm and the eigen-face technique: Firstly, the possible human eye regions are detected by testing all the valley regions in the gray-level image. Then the genetic algorithm is used to generate all the possible face regions which include the eyebrows, the iris, the nostril and the mouth corners. Each possible face candidate is normalized to reduce both the lightning effect, which is caused by uneven illumination; and the shirring effect, which is due to head movement.[3] The fitness value of each candidate is measured based on its projection on the eigen-faces. After a number of iterations, all the face candidates with a high fitness value are selected for further verification. The Viola-Jones algorithm is a widely used mechanism for object detection. The main property of this algorithm is that training is slow, but detection is fast. This algorithm uses Haar basis feature filters, so it does not use multiplications. The efficiency of the Viola-Jones algorithm can be significantly increased by first generating the integral image.[2]

2.4. Feature Extraction : In machine learning, pattern recognition and in image processing, feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations. Feature extraction is a dimensionality reduction process, where an initial set of raw variables is reduced to more manageable groups (features) for processing, while still accurately and completely describing the original data set. When the input data to an algorithm is too large to be processed and it is suspected to be redundant (e.g. the same measurement in both feet and meters, or the repetitiveness of images presented as pixels), then it can be transformed into a reduced set of features (also named a feature vector). Determining a subset of the initial features is called feature selection.[The selected features are expected to contain the relevant information from the input data, so that the desired task can be performed by using this reduced representation instead of the complete initial data. Local Binary Pattern (LBP) features have performed very well in various applications, including texture classification and segmentation, image retrieval and surface inspection. In the feature extraction process, we can implement the LBP for pattern extraction in image, and MFCC, Energy features extraction in the speech signal. Local binary patterns (LBP) are a type of visual descriptor used for classification in computer vision.

It has since been found to be a powerful feature for texture classification; it has further been determined that when LBP is combined with the Histogram of oriented gradients (HOG) descriptor, it improves the detection performance considerably on some datasets. Mel-frequency cepstral coefficients (MFCCs)[1] are coefficients that collectively make up an MFC. They are derived from a type of cepstral representation of the audio clip (a nonlinear "spectrum-of-a-spectrum"). The difference between the cepstrum and the mel-frequency cepstrum is that in the MFC, the frequency bands are equally spaced on the mel scale, which approximates the human auditory system's response more closely than the linearly-spaced frequency bands used in the normal cepstrum. This frequency warping can allow for better representation of sound, for example, in audio compression.[10][11]

2.5. Classification: A Support Vector Machine (SVM) is discriminative classifiers formally defined by a separating hyper plane.We give trained image features and target data as an input to SVM. A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyper plane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyper plane which categorizes new examples.SVMs does the mapping from input space to feature space to support nonlinear classification problems. The kernel trick is helpful for doing this by allowing the absence of the exact formulation of mapping function which could cause the issue of curse of dimensionality. This makes a linear classification in the new space (or the feature space) equivalent to nonlinear classification in the original space (or the input space). SVMs do these by mapping input vectors to a higher dimensional space (or feature space) where a maximal separating hyper plane is constructed. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyper plane which categorizes new examples. In this step we implement the Multi SVM Classifier is used to recognize the person and then display the result.[1]

2.6. Performance Measure: The performance of the process is measured in terms of performance metrics like Accuracy, Sensitivity, and Specificity.
Terms associated with performance measure:

- TP- True Positive (correctly identified)
- TN-True Negative(correctly rejected)
- FP-False Positive(incorrectly identified)
- FN-False Negative(incorrectly rejected).

2.6.1. Accuracy: Accuracy in classification problems is the number of correct predictions made by the model over all kinds predictions made.

$$\text{Accuracy} = (TP+TN) / (TN+TP+FN+FP)$$

2.6.2. Sensitivity: The ability of a test to correctly identify those with the disease (true positive rate).Measures the proportion of actual positives that are correctly identified.

$$\text{Sensitivity} = TP / (TP+FN)$$

Specificity: The ability of the test to correctly identify those without the disease (true negative rate). Measures the proportion of actual negatives that are correctly identified.

$$\text{Specificity} = TN / (TN+FP)$$

2.6.3. F-measure: The F measure (F1 score or F score) is a measure of a test's accuracy and is defined as the weighted harmonic mean of the precision and recall of the test.

f_measure = 2*((precision*recall)/ (precision + recall));

2.6.4. Gmean: The geometric mean is a mean or average, which indicates the central tendency or typical value of a set of numbers by using the product of their values (as opposed to the arithmetic mean which uses their sum)
gmean = sqrt (tp_rate*tn_rate);

The data in Table 1 . shows that the model in this work in a similar type of training set is poor and in some cases exceeds the existing model. However, it is impossible to compare the accuracy of recognition because the size of the model training samples provided in different publications cannot usually be compared to the size of the model used to perform this study.

Table 1. Comparisons of the Results Obtained With Existing Solutions.

	Technique	Accuracy	Specificity	Sensitivity
Proposed Work	Multi-SVM	99.68	99.93	99.63
Previous Work [1]	KNN	73.1	-	-

CONCLUSION

Biometrics is a field related to human identification through physiological features, fingerprints, irises, voices, face, etc. Biometric identification systems can be used for human identification or verification. This paper records human faces and voice signals. Therefore, the LBP feature description extracts the exact features of pixels in the image. We use SVM for classification. Therefore, Multi-SVM classification for binary classification requires only a few positive examples compared to other classifiers. To enhance the power of speech recognition system, it is required to design speech recognizers in local languages. Multilingual is new evolving field in area of speech recognition. Although this field has gained a wide approval to automate the services and applications but there are several parameters which affect the accuracy and efficiency of speech recognition system. in this work accuracy of the system improve as per existing system

REFERENCES

[1] V. Zatonskikh, Georgii I. Borzunov, Konstantin Kogos Development of Elements of Two-Level Biometric Protection Based on Face and Speech Recognition in the Video Stream Efim Department of Cryptology and Cybersecurity National Research Nuclear University MEPhI (Moscow Engineering 978-1-5386-4340-2/18/$31.00©2018 IEEE Moscow, Russia

[2] Saswati Debnath PinkiRoyMulti-modal authentication system based on audio-visual dataTENCON 2019 - 2019 IEEE Region 10 Conference (TENCON) Year: 2019 ISBN: 978-1-7281-1895-6 DOI: 10.1109/IEEEKochi, India, India

[3] AakarshMalhotra ; Richa Singh ; MayankVatsa ; Vishal M. PatelPerson Authentication Using Head Images2018 IEEE Winter Conference on Applications

of Computer Vision (WACV)Year: 2018 ISBN: 978-1-5386-4886-5 DOI: 10.1109/IEEELake Tahoe, NV, USA

[4] Mahesh R. Pawar ;Imdad Rizvi IoT Based Embedded System for Vehicle Security and Driver Surveillance 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)Year: 2018 ISBN: 978-1-5386-1974-2 DOI: 10.1109/IEEECoimbatore, India

[5] NayansukhPatil ; Rachana PatilAchieving Flatness: with Video Captcha, Location Tracking, Selecting the Honeywords2018 International Conference on Smart City and Emerging Technology (ICSCET)Year: 2018 ISBN: 978-1-5386-1185-2 DOI: 10.1109/IEEEMumbai, India

[6] MaheenZulfiqar ; Fatima Syed ; Muhammad Jaleed Khan ; KhurramKhurshidDeep Face Recognition for Biometric Authentication2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)Year: 2019

[7] MithunDutta ;KangkhitaKaem Psyche ; Tania Khatun ; Md. Ashiqul Islam ; Md. AzijulIslamATM Card Security Using Bio-Metric and Message Authentication Technology2018 IEEE International Conference on Computer and Communication Engineering Technology (CCET)Year: 2018 ISBN: 978-1-5386-7437-6 DOI: 10.1109/IEEEBeijing, China

[8] R. Divya ; Raja LavanyaA Systematic Review on Gait Based Authentication System2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)Year: 2020 ISBN: 978-1-7281-5197-7 DOI: 10.1109/IEEECoimbatore, India, India

[9] Chandrakant P. Divate ; Syed Zakir Ali Study of Different Bio-Metric Based Gender Classification Systems2018 International Conference on Inventive Research in Computing Applications (ICIRCA)Year: 2018 ISBN: 978-1-5386-2456-2DOI: 10.1109/IEEECoimbatore, India

[10] Abdul Razaque ;PrudhviSagarSreeramoju ; Fathi H. Amsaad ; Chaitanya Kumar Nerella ; MusbahAbdulgader ; HarshaSaranuMulti-biometric system using Fuzzy Vault2016 IEEE International Conference on Electro Information Technology (EIT)Year: 2016 ISBN: 978-1-4673-9985-2 DOI: 10.1109/IEEEGrand Forks, ND, USA

[11] A.Muthu Kumar ; A. Chandralekha ; Y. Himaja ; S. MounikaSaiLocal Binary Pattern based Multimodal Biometric Recognition using Ear and FKP with Feature Level Fusion 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS) Year: 2019DOI: 10.1109/IEEETamilnadu, India, India.