IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 17, Issue 5, Ver. V (Sep. – Oct. 2015), PP 20-25
www.iosrjournals.org

Gender classification using face image and voice

Dharamraj yadav[1] Shashwat Shukla[2] & Bramah Hazela[2]
Dept.of Computer Science & Engineering Amity School of Engineering & Technology Amity University, Lucknow campus, India
Dept.of Computer Science & Engineering Amity School of Engineering & Technology Amity University, Lucknow campus, India

Abstract: *This paper is about gender classification using face image and voice of a speaker. The basic aim of the paper is to predict the gender of speaker through voice sample using Auto-correlation method and predict the gender of speaker through their face image using fisher face algorithm and integrated the obtained to increase the accuracy of gender classifier. The algorithm has been applied with the help of MATLAB programming. A collected database of voice samples and image samples of male and female students has been concerned. The classification algorithm is applied on the database and accuracy of gender classifier is determined.*

Keywords: *Image Processing, Gender Recognition, Pre-processing, Linear Discriminate Analysis, Principle Component Analysis.*

I. Introduction

The capacity to perform programmed acknowledgment of human gender is pivotal for various frameworks that procedure or adventure human-source data. Common cases are data recovery, human PC or human-robot collaboration. The result of a Gender Recognition framework can be utilized for producing metadata data helpful for expounding sound and feature records. Additionally, gender orientation is a vital prompt that can be misused for enhancing understand ability of man-machine connection, or just, for diminishing the inquiry space in speaker acknowledgment or reconnaissance frameworks.

The issue of Gender acknowledgment was tended to in the past by a few creators (see Section 2). In every one of these works, one and only methodology (voice and face picture) was utilized. The examinations were performed essentially under clean conditions and the power of Gender acknowledgment frameworks in certifiable situations was occasional considered. In most average applications, both sound and vision are accessible. Preferably, a Gender acknowledgment framework ought to have the capacity to endeavour both modalities to enhance power. Since every methodology has distinctive attributes, Voice-picture signals can give a more complete depiction of a subject than a solitary methodology. At long last, reconciliation of the signs may yield a Gender acknowledgment framework that is versatile to the corruption of both, or even to fleeting inaccessibility of one of the data signals.

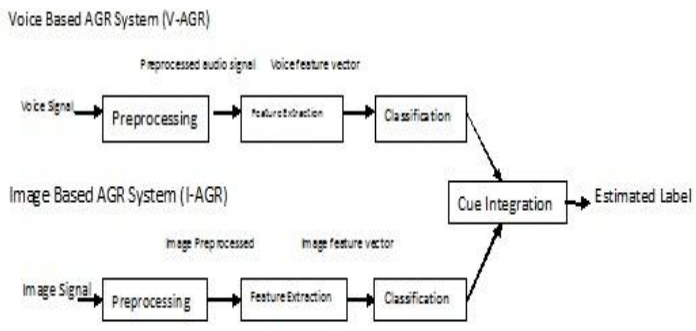

Fig. 1. Overview of the architecture of the VI-GR system. The two modalities are processed individually, and then integrated at the classifier level.

In this paper, we research a Voice-Image Gender acknowledgment (VI-Gender acknowledgment) framework prepared under clean conditions. We additionally dissect diverse component representations for each of the signal, and survey their strength to fluctuating conditions. The VI-Gender acknowledgment mulled over in this work is in light of the abnormal state mix structure. As it were, first uni-modular Gender acknowledgment frameworks are prepared, and after that the sign mix is performed by melding the proofs from

the two frameworks. Through broad studies under differing conditions, as distinctive light force of face picture, diverse stance of face picture, and lighter twisting in stable recording is utilized here for testing reason ,

we demonstrate that: (a) the sound based framework is more hearty than the vision-based framework, and (b) combination of Voice-image cues yields a versatile framework that jam the execution of the best methodology in clean conditions and, aides in enhancing the general execution in noisy conditions.

II. Related Work

The beforehand proposed answers for the Gender acknowledgment issue were in view of single methodology, either on sound or vision. The primary chips away at sound based Gender acknowledgment went for distinguishing the most suitable elements of discourse sign for the errand. Examination of voice source on the basis of pitch coordination and vocal track related segments for different vowels excluded from the clean condition voice data of more than 20 peoples has been analysed. In addition of that the graphical representation of voice data was performed. The method applied on voice data was autocorrelation reflection and cepstrum pitch determining for the same voice sample. The assessment of mel-cepstral components for distinctive gatherings of phonemes like vowels, nasal, fluids and so forth was led in [3]. Additionally, the last two studies investigated an impact of distinctive channel orders (from 8 to 20) and sorts of coefficients (static versus delta) on the execution of a Gender acknowledgment framework. All the more as of late, the comparison of Support Vector Machines (SVMs) with closest neighbour classifiers for the initial 12 cepstral coefficients on astounding recordings of 150 speakers from the ISOLET corpus was given 100% Gender acknowledgment rate for SVMs [4].

Early research in image-based Gender Recognition was focussed upon the utilization of artificial neural systems for highlight extraction and grouping on clean condition information [5, 6]. Most recent exploration investigated more unpredictable lighting and posture varieties, and for bigger arrangements of subjects, for example, in the FERET database [7]. The trial studies proposed that for the Gender acknowledgment assignment in view of visual signals, the SVMs with the RBF part are better than the direct, quadratic, fisherface straight segregate, k-closest neighbor classifiers and also to more mind boggling strategies, for example, extensive troupe RBF systems [7, 4].

In [4], correlation of column information representation with components got through primary part examination (PCA), alluded to as eigenfaces [8], was made on database comprising of 1640 frontal, unclouded face pictures.

III. The Automatic Gender Recognition System

This section presents architecture of our Voice-image Gender recognition system.

3.1. System Overview

In designing the VI- Gender recognition a two-fold approach was adapted. First, we studied the two cues separately by building voice-based and image-based Gender recognition systems (V- Gender recognition and I- Gender recognition).Second, these systems were integrated to provide the final decision based on both modalities. Overview of the system architecture is presented in Figure 1. In the proposed solution, the V-Gender recognition system utilizes speech signal. Similarly, the I- Gender recognition system exploits exclusively face images and no stature information is used. The V-Gender recognition and I-Gender acknowledgment frameworks have comparable architectures which comprise of four sections performing the accompanying capacities: (a)data gaining (b)data preparing, (c) highlight extraction, and (d) grouping. The part of the sign pre-processing piece is extraction of valuable sections of the sign. The previous studies on audio suggested that voiced phonemes are more discriminative for gender than unvoiced phonemes [2, 3]. We use a voiced/unvoiced detection to obtain the most informative parts of the signal. In case of the V-Gender Recognition system, data pre-processing includes face detection, localization, and finally segmentation. The function of the second block is extraction of features from the pre-processed signal that allow for most accurate classification of the subject. The description of this block for the V- Gender Recognition and I- Gender Recognition system is given in Sections 3.2 and 3.3, respectively. Finally, classification of an instance to one of the two possible classes (female or male) is performed.

The algorithm used here for gender recognition through voice is Autocorrelation method of voice and for the image the fisherface algorithm used. The result is divided in to two classes (male and female).The final result is sum of two individual results of separate algorithms.

3.2. Audio Features

Gender Classifier from discourse is a piece of programmed discourse acknowledgment framework to improve speaker flexibility and a piece of programmed speaker acknowledgment framework. The gender recognition model can be used as phone calls discrimination like if there is demand of only female person to call

on particular number then in that case the gender recognition can work as a filter and separate the male call and female call or block anyone of them.

Some time the way of pronounce a word gives different accent for male and female because of their voice pitch and their internal structure of voice system of genders. The arrangement of essential sound units called phonemes. A sound or the combination of mixed sound proverb to have the same limit by the speakers of a language they are called as Phoneme. An occurrence of a phoneme is/k/ sound in the words kite and knight. The same phoneme may produce different sounds or allophone during speaking these phoneme. That is varies with person to person. It happen because of the differences in shape vernacular and vocal length.

3.3. Visual Features

The standard automatic Image base gender recognition system use face image for gender recognition. The face image is trained through fisher face method. In fisher face method the PCA is used to reduce the measurement of image .hence the size of array will be reduced. However, in case of face recognition systems, this technique works efficiently only when constant face pose and lightning are preserved and tends to fail under varying conditions. To overcome this problem a technique that additionally used here that is linear discriminant analysis (LDA), referred to as the fisherface method, was introduced [12]. Both types of features, eigenfaces and fisherfaces were evaluated in this work .

3.4. Voice-image Gender recognition System

The VI-Gender recognition system is created by fusing evidences from the two modalities at the high level, after the single-cue classification is performed. The structural design of the VI-Gender recognition system is presented in Figure 1. The a posteriori probabilities provided by the single-cue classifiers are combined using the sum or product rule to provide the final decision based on both modalities. Theoretical studies show that these two rules are most suitable for the two-class problem [13]. Additionally, we considered equal or unequal weighting of modalities during the experiments. The latter were performed in order to answer the question of different importance of the modalities in the correct classification.

IV. Setup Of Proposed Work

4.1.image and voice Database

The Gender acknowledgment framework was assessed on the understudy database having the gathering of male and female pictures and there relating voice. The information procurement was performed through a camera and mouthpiece having a decent nature of obtaining of information. Foundation of picture is not consider just the fecal piece of picture is trimmed. The picture database is isolated into two gatherings, one is male and another is female. For the voice information the voice is recorded through receiver in such a situation where slightest foundation clamor . The information were isolated into two sets utilized for training and testing data. Tranning and testing sets are rendamly divided in to two category of data. Each of them have the face of male and female. As here we used 20 subject. At each subject there are 4 face images. So the total number of image is 80.

These images are randomly divided in to two categories after conversion of Eigen value.

Similarly the voice sample of each person is collected so the total number of voice sample is 20.

The database collected from 20 college student used for the testing and gender recognition.

4.2. Analysis of voice Data

In the voice information investigation first we record the sound influx of the recurrence 8000 casing for every second then spare the information with the name of individual to whom voice is recorded. At that point the auto connection technique is connected on sound wave document. The relationship among two waveforms is a quantify of their comparability. The waveforms are dissected at various time intervals, and their "comparability" is registered at each interval. The delayed consequence of a relationship is a measure of equivalence as a component of time slack between the beginnings of the two waveforms. The autocorrelation limit is the relationship of a waveform with itself. One would expect exact comparability immediately slack of zero, with extending contrast as the time slack augmentations. The pitch extraction strategy of a talk sign can be in light of handling the brief while autocorrelation limit of the talk signal. Brief Time Auto-relationship for a talk sign is given by

$$Rn(k) = \sum_{M=-\infty}^{\infty} \begin{matrix} x(m)(n-m)] \\ x(m+k)w(n-m-S)] \end{matrix}$$

Where,

Rn(k) = Auto-correlation(short time)

x = voice Signal
w = Window size
S = Sample time at which auto-correlation was calculated
For audio portions of speech, the Short-Time Auto-relationship capacity shows periodicity of the discourse. Rn(k) diminishes with k as the reckoning procedure proceeds. The figure 2 and 3 exhibited the autocorrelation diagram of male and female sample.

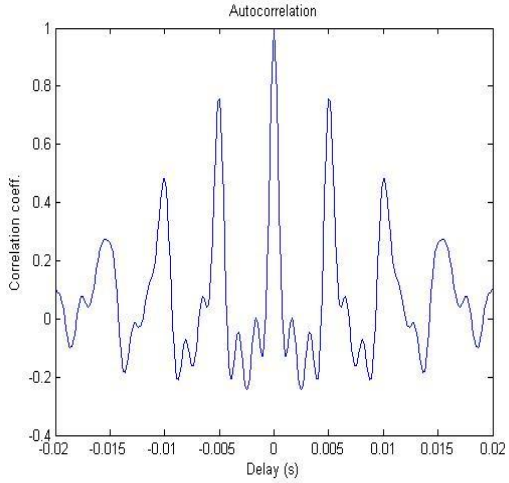

Fig. 2: Autocorrelation graph plot of a Female voice sample

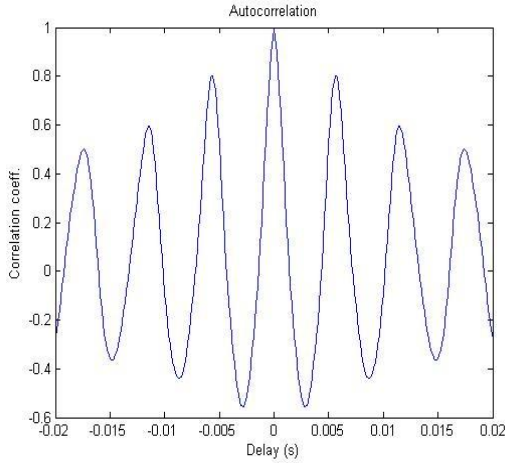

Fig. 3: Autocorrelation graph plot of a male voice sample

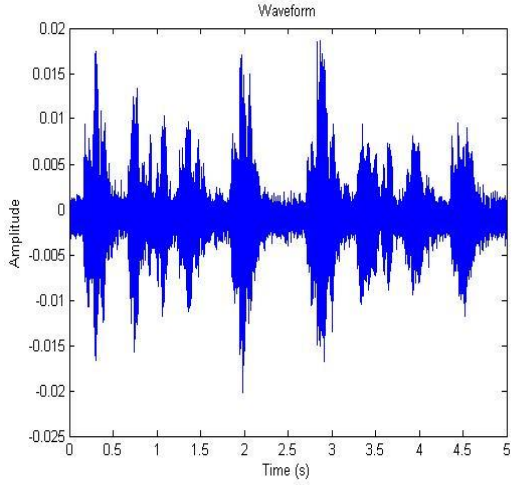

Fig. 4: waveform of a female voice sample

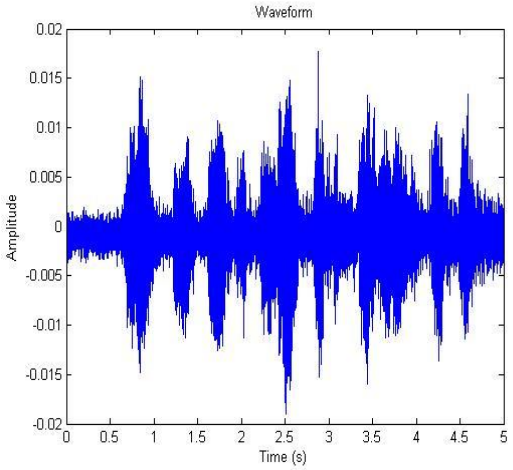

Fig. 5: waveform of a male voice sample

4.3. Analysis of Image Data

In order to extract the face region from an image , first the image is cropped in such a way manually that only eyes , nose and lips are to be the part of image and rest is removed Then, every picture was trimmed to a size of 65x80. The picture is then partitioned into two gatherings (Male and Female).then after that the Shuffling of picture. The PCA is utilized here to decrease the quantity of segment. It decreases the span of exhibit into X to N-C where as x=Array of all pictures and N is the quantity of section in X and C is the quantity of gatherings. While applying the PCA on picture information it diminishes the extent of cluster to 46 from 48 subjects for every pictures. Then the fisher face features were obtained through LDA. The LDA features were encoded in to n-1 vectors, where n represent as the number of classes (here the number of classes is only two), each image was represented using only one feature.

4.4. System

We used the resultant output of the I-gender recognition and V-gender recognition and integrate both the result and check that either result is matching with the actual gender of particular person or not. And also we calculate the number of time system fail to produce correct result. On those bases the system accuracy can be evaluated.

V. Results And Discussion

This section presents an assessment of the Gender recognition systems under varying conditions.

5.1. Voice-Based Gender recognition System

If there should be an occurrence of the voice-Gender acknowledgment framework, we consider the two things: voice recurrence (Fs) and Autocorrelation of the voice test. The voice recorded through a typical voice recorder having outside clamor additionally accessible. The trial performed on 20 voice test (counting male and female). What's more, out of 20 voice test the voice recognizer perceive 17 example effectively and just 3 specimen anticipated wrong. Accordingly the voice acknowledgment framework anticipated 85% effectively The after-effect of acknowledgment framework can be enhanced on the off chance that we utilize high calibre of voice recorder. The execution of the Gender acknowledgment framework diminishes with expanding seriousness of conditions.

5.2. Image-Based Gender recognition System

If there should be an occurrence of the V-Gender acknowledgment framework, we looked at two sorts of components: eigenfaces and fisherfaces. Results got under diverse conditions are displayed in Figure 3. For both elements execution of the V- Gender acknowledgment framework diminishes with corruption of the sign. Of course, the fisherfaces are better than eigenfaces under controlled conditions. The distinction in execution of 2.1% (document precision) was watched. Then again, eigenfaces are superior to fisherfaces by 4.0% and 0.7%under debased and unfavourable conditions, separately.

It might be an outcome of the befuddle between inside of class changes that are fundamentally higher in the test than preparing set.

5.3. V I-Gender recognition System

For the VI- Gender acknowledgment framework, we assessed mixes of voice (F0, PLPs) and Image (eigenfaces, fisherfaces) highlights. Results acquired for the total and item run the show.

For the combination of voice and image of database containing 20 different person's image where each person folder have 4 image .Each person voice sample also used here so the total voice sample is 20.The VI-gender recognition system predicted 90% correctly. The VI-gender recognition system only produces false result only on 2 samples. The result of the VI- gender recognition system can be improved further by using the high definition camera and a good quality of microphone. These are the factors that affect the result of the system slightly and by the removal of these degradation the result of VI- Gender recognition system can produce 95% correct result.

The result of the system is lower to 90% because of the fact that the I-gender recognition system was inferior to the V-gender recognition system. Future work will attend to the problem of improving the robustness of the V-gender recognition system.

VI. Conclusion

Conclusion of this paper is based on proposal of a model having two type of data are used together for identify the gender. Here we use both voice and Image as the input of gender recognition system so this features make the system more robustness. Individually gender recognition through voice is more reliable as compare to image of the database we used here. The integration of voice and image in to single model for gender recognition improve the result from there individual result. And the VI-Gender recognition system can be used on the worst case where if anyone of the type of input data is unavailable either image or voice then even in that condition it will work and produce result.

References

[1]. D. G. Childers and K. Wu, "Gender Recognition from speech. part II: Fine analysis," *JASA*, vol. 90, pp. 1841–1856, 1991.
[2]. B. Moghaddam and M-H. Yang, "Gender classification with support vector machines," *Proc. FG*, 2000.
[3]. J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in *Advances in Large Margin Classifiers*, 1999.
[4]. P.J. Phillips, H. Wechsler, J. Huang, and P.J. Rauss, "The FERET Database and Evaluation Procedure for Face Recognition Algorithms," Image and Vision Computing J., vol. 16, no. 5, pp. 295-306, 1998.
[5]. A. Bladon, "Acoustic phonetics, auditory phonetics, speaker sex and speech recognition: a thread," in *Computer Speech Proc.*, 1985.
[6]. A.M. Burton, V. Bruce, N. Dench (1993) What's the difference between men and women? Evidence from facial measurement, Perception, 22:153-176
[7]. L. Sirovich et al., "Low-dimensional procedure for the characterization of human faces," *JOSA*, vol. 2,, 1987.
[8]. D.M. Tax et al., "Combining classifiers by averaging or multiplying?," *Pattern Recog.*, vol. 33, no. 9, 2000.
[9]. Tomi Kinnunen " Spectral Feature for Automatic Voice-independent Speaker Recognition"Depertment of Computer Science, Joensuu University,Finland. December 21, 2003
[10]. E. Bailly-Baillire et al., "The BANCA database and evaluation protocol," *LNCS*, vol. 2688, 2003.
[11]. Pattern Based Gender Classification Omveer Singh, Gautam Bommagani, Sr. Reddy Ravula. Vinit Kumar Gunjan,' International Journal of Advanced Research in Computer Science and Software Engineering' .
[12]. K. Wu and D. G. Childers, "Gender Recognition from speech. part I: Coarse analysis," *JASA*, vol. 90, pp. 1828–1840, 1991.
[13]. J. W. Fussell, "Automatic sex identification from short segments of speech," in *Proc. ICASSP*, 1991.
[14]. S.Marcel et al., "On the recent use of local binary patterns for face authentication," *EURASIP JIVP*, 2007, Accepted.
[15]. L.Walawalkar et al., "Support vector learning for gender classification using audio and visual cues: A comparison.," in *SVM*, 2002, vol. 2388 of *LNCS*,.
[16]. G. Mallikarjuna Rao, G. R. Babu, G. Vijaya Kumari and N.Krishna Chaitanya, "Methodological Approach for Machine based Expression and Gender Classification, IEEE International Advance Computing Conference", pp. 1369-1374, 6-7 March 2009,.