Buletin Teknik Elektro dan Informatika
(Bulletin of Electrical Engineering and Informatics)
Vol.x, No.x, Month 2011, pp. 4xx~4xx
ISSN: 2089-3191

Speaker Recognition in Content-Based Image Retrieval for a high degree of accuracy

Suhartono[1,*] Totok Chamidy[2], Syahiduz Zaman[3]
[123]Department of Informatics, UIN Maulana Malik Ibrahim, Malang, Indonesia
*Correspondence author, e-mail: suhartonouinmlg@gmail.com

Abstrak

Pengenalan pembicara adalah proses yang dilakukan oleh komputer untuk mengenali kata yang diucapkan oleh seseorang terlepas dari identitas orang yang bersangkutan. Latar belakang dari penelitian ini adalah untuk menciptakan sebuah sistem pengenalan pembicara yang menggunakan data dinamis. Pengenalan pembicara menggunakan data dinamis; penggunaan data dinamis sulit didekati dengan rumus matematika tertentu. Metode pengenal pembicara saat ini adalah dibutuhkan untuk tingkat akurasi yang tinggi. Tujuan dari penelitian ini adalah untuk mengukur keakuratan pengenalan pembicara.

Metode yang digunakan dalam penelitian ini menggunakan metode fuzzy Mamdani, dan metode Manhattan Distance, dimana metode fuzzy Mamdani digunakan untuk identifikasi, sedangkan pada metode Manhattan Distance digunakan untuk verifikasi. Data sampel diperoleh dari fitur ekstraksi row mean pada data bentuk spektrogram digital. Dengan metode Content-Based Image Retrieval, maka data rekaman suara manusia diubah menjadi bentuk spektrogram digital. Berbagai ukuran digunakan pada penelitian ini : 256x256, 128x128, 64x64, 32x32 dan 16x16. Untuk mendapatkan fitur vektor yang memberikan sifat lebih baik, maka dilakukan proses untuk mendapatkan fitur vektor menggunakan kekre transform pada setiap sub-gambar. Fitur vektor kemudian digunakan sebagai aturan input dalam pengenalan pembicara. Sedangkan untuk aturan output adalah identitas suara dari manusia digunakan. Sistem yang dibuat dapat mengenali pembicara secara otomatis dari suaranya dan dapat memberikan ketepatan pengenalan speaker 91% pada ukuran fitur 32x32.

Kata kunci: *system fuzzy, ukuran jarak, pengenalan pembicara, identifikasi dan suara manusia.*

Abstract

Speaker recognition is a process that performed by a computer to recognize a word spoken by a person regardless the identity of the person concerned. The background of this research is to create a speaker recognition system that uses dynamic data. The pattern of speaker recognition obtained is dynamic data; dynamic data is difficult to approach with certain formulas. The speaker recognition method is currently required for a high degree of accuracy. The purpose of this research is to measure the accuracy of speaker recognition.

The method used in this research using fuzzy Mamdani method, and Manhattan distance method, in fuzzy Mamdani method used for identification, while in Manhattan distance method used for verification. The sample data obtained from features extraction row mean on spectrogram form image digital. With Content-Based Image Retrieval method, those data of the recording converted to become spectrogram form image digital. Various sizes were used 256x256, 128x128, 64x64, 32x32 and 16x16. To get vector features that give better properties, the process was performed to get vector feature using kekre transform and mean on each sub-image. The vector features then used as input rule in the speaker recognition. As for output rule, the identity of the human voice was used. The system can recognize a person automatically from his or her voice and can provide accuracy of speaker recognition 91% on the size of 32x32 features.

Keywords*: Fuzzy systems, Distance measurement, Speaker recognition, identification and Human voice.*

1. Introduction

Artificial Intelligence approach has been implemented in many fields, like in the field of decision making support system [1] and optimization [2]. One of artificial intelligence approach is fuzzy Mamdani method that can be used to map an input room to output room [3]. Where fuzzy Mamdani method can be used as one dynamic system to identify system [4]. Fuzzy Mamdani method can explain the relationship between input and output in non-linear conditioning the form of modeling [5].

Received May 9th, 201x; Revised August 3rd, 201x; Accepted August 16th, 201x

Fuzzy Mamdani method is a method based on fuzzy logic [6]. Fuzzy logic first used was to manage uncertainty [7]. While uncertainty is a problem contains doubt and unfit. The appearance of fuzzy logic doesn't mean replacing probability theory that has existed previously, but with fuzzy logic, we have found another alternative that can be used to solve problems of uncertainty.

In this research, fuzzy Mamdani method and Manhattan distance can connect between human voice records and sentence from the human voice. Where fuzzy Mamdani method can give solution related to the complex system and can give output identification for the non-linear system [8]. While Manhattan distance can similarity measurement between sample data and test data. Fuzzy Mamdani been used by researchers to the identification of Canaries Birds Chirp Quality.

The reason researchers use fuzzy Mamdani method is a fuzzy logic method that has the simplest formula compare to other fuzzy methods. In fuzzy Mamdani method was the output variable in the form of the constant or linear equation. In the research related to fuzzy Mamdani with voice, fuzzy Mamdani method was used to make a model speech recognition system [9], speech recognition pattern that has been made in order to obtain the best model for final recognition. In the research fuzzy Mamdani method used for speaker identification, the voice form human into the speech recognition system compares with some voices that have been identified in the database.

While to transform sound signal as input to become parameters that can be recognized by the system, Content-Base Image Retrieval (CBIR) method was needed. CBIR method was used to identify the speaker, stressing on feature vector selection that produced. That criterion was determined by extraction of feature row mean with Content-Based Image Retrieval method. CBIR (Content Based Image Retrieval) is feature extraction method that uses content on the image as the feature. Obtained contents were in the form of color, texture, shape, or other information. CBIR method has been used by researchers to the identification of canaries birds chirp quality. CBIR method has been used by researchers to the identification of canaries birds chirp quality. CBIR method has been used for compare source digital image with target digital image. The method to compare is image digital distance measure based feature vector.

One of the research examples about features extraction CBIR with transform domain approach that have been performed was about face recognition [10]. The features extraction CBIR can be substitute solution based on color [11]. CBIR implementation to speaker recognition is to start with performing conversion process of the sound signal into images using Short Time Fourier Transform (STFT). In this research, images produced by STFT were in the shape of iris spectrogram that appears as frequency spectrum that plotted against time and amplitude in the shape of iris.

Sound processing using spectrogram analysis can be done by recording process using microphone censor [12]. Spectrogram analysis is the analysis based on frequency, where human's voice will be divided into base frequency element [13]. Transformation method Fast Fourier Transform (FFT) can be used to get spectrogram analysis. To represent the signal in time domain and frequency, Sort Time Fourier Transform (STFT) can be used [14]. The advantage of STFT is frequency component from signal can be known at any point of time.

Problem formulation in this research was on concluding the result of speaker recognition for a high degree of accuracy. The system of speaker recognition consists identification system and verification, identification system using fuzzy Mamdani method, and then verification system uses Manhattan distance method. The speaker recognition was compared voice of human recording with human recording voice on the database.

Therefore, the result of this research is to build speaker recognition system using extraction of feature row mean image with consideration of feature vector size using kekre transform as the transformation method. Therefore, the result of this research is to build speaker recognition system using extraction of feature row mean image with consideration of feature vector size using kekre transform as the transformation method.

2. Research Method

For data processing and computing were performed in Network Laboratory at BJ Habibie building, faculty of science and technology State Islamic University of Maulana Malik Ibrahim, Malang, and East Java, Indonesia. The design of speaker recognition consists identification system and verification. In figure 1 the identification proses using fuzzy Mamdani method, the identification proses were determined the speaker was part of some speaker. And

then, the verification proses claiming that sounds match the design of speaker recognition in the research can follow in figure 1.

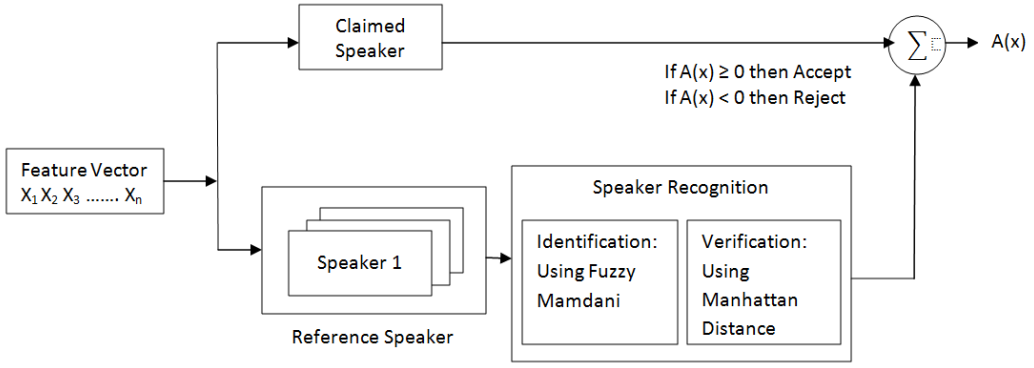

Figure 1. Diagram of speaker recognition system

The speaker recognition system is described in detail with flowcharts. The flowchart for the design of speaker recognition can follow in figure 1.

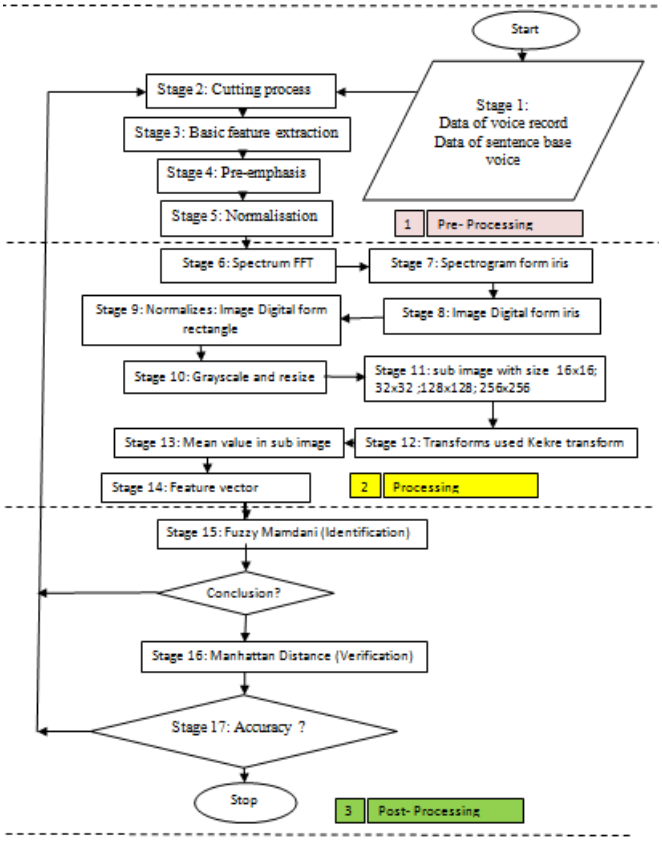

Figure 2. Stages that performed in this research

In figure 2, the process of speaker recognition in the research using three main stages. The three main stages were the pre-processing stage, processing stage, and post-processing stage. In pre-processing stage was processed extraction, in the process extraction was all human voice record data are treated the same. In the processing, stage generates feature vector. And then, the post-processing stage using method for speaker recognition.

3. Results and Analysis

In figure 2 was explained the stages performed in order to build speaker recognition usage fuzzy Mamdani method and Manhattan distance method. Those stages started from human voice record. Stages in this research as follows:

3.1. Stage 1: Data preparation

As for data of human voice, was obtained recording process and from the research using fifty-four of training data and twelve of test data. The sentence is “good morning”. The data of human voice is three speakers. Recording process on human voice was performed for two minutes during the recording process.

3.2. Stage 2: Cutting process

Perform cutting process of recording data from two minutes into fifteen seconds. The cutting of records was performed using code of Matlab software. Then complete program codes were%-to read the sound file; [x, fs] = wavread ('statement_1.wav'); x=x (:1);
x=x(1:900) ;

3.3. Stage 3: Basic feature extraction process.

This stage was performed after obtaining sample data after cutting process; therefore, the purpose of this process is data samples have the same time, and then performed feature extraction process from the records of the human voice in order to get the characters in every sentence from the human voice. Features obtained from extraction process are signal length, time vector, and data samples. Then complete program codes were %-to read the sound file; [x, fs] = wavread ('statement_1.wav'); x=x (:1); %-- to read data samples; N=length(x); %--to read signal length; t= (0: N-1)/fs; %--to read time vector;

3.4. Stage 4: Perform pre-emphasis process.

This process is to dismiss the DC components. Dismissing DC components by counting the average of voice data samples, then deduct by every data sample. Program codes for pre-emphasis process were u = mean(x); %--obtaining average value/mean; x=x-u; %-- dismissing component;

3.5. Stage 5: Normalisation process

Normalisation process is the process that can be used to normalize degraded sample value that caused by the distance of human voice and microphone recorder. In every record, the human voice has the different shape and also different amplitude level. Therefore, to level the highest amplitude value from every record, normalization process was performed. Normalisation amplitude process was performed by dividing all value digital signals with an absolute highest value of data sample. Generally, normalization process program code with Matlab was Kn = 255/maxval; x=Kn*x;

3.6. Stage 6: Short Term Fourier Transform (STFT) process.

STFT process is the process to change sound signal into spectrogram images. Simply, STFT process contains frame blocking process, windowing, and fast Fourier transform [15]. Human voice signal that has been normalized then will go to sort time Fourier transform (STFT) process. This process is the process of transforming Fourier on stationary period. Example, the Human voice signal is x[i], where i=1, 2, 3 …, N then to determine spectrum frequency it multiplies by the length of window Nm obtained FFT. The forming process of iris spectrogram from sound signal can be seen in figure 3.

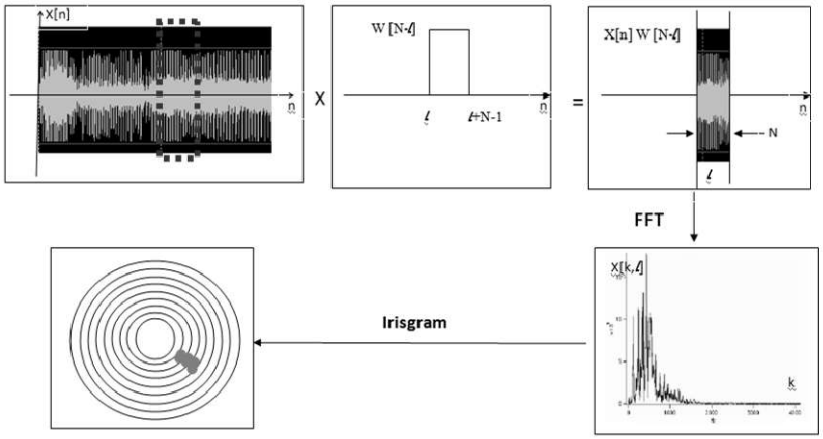

Figure 3. The process of forming iris spectrogram from the sound signal using STFT.

Mathematic formula from STFT, using equation (1).

$$X[k,l] = \mathrm{DFT}\{[x[l]w[0] \ldots . x[l+N-1]w[N-1]]\} \quad (1)$$

Where k is the index that shows frequency, 1 is time and w (t-т) is window function to cut signal.

3.6.1 Stage 6.1: Frame blocking process

Frame blocking process purpose is to divide the sound signal into smaller pieces that will ease the process of calculating and sound analysis. Frame blocking Process by blocking sound signal to become few frames as many as N samples, some frames are close to space M (M<N). The first frame consists of N first sample. Second frame with M sample after first frame and overlap with N-M sample. With the same way, the third frames start 2M sample after first frame (or M sample after the second frame) and overlap with N-2M sample. Frame Blocking was performed continuously until all signals were processed. Generally, Frame blocking process program code with Matlab were fs=16000; segment=fs∗.025; olp=fs∗.015; winlen=segment+olp;

3.6.2 Stage 6.2: Hamming window.

Frame blocking process causes sound signal discontinue/non-stationary that can create new frequencies while the process FFT was being performed. That is why, to prevent that, the process of changing sound signals from discontinuing to continuing was needed. One of the ways is by using hamming window process. Frame blocking process program code with Matlab were winlen = 2096; % window length; win = hamming (winlen, 'periodic');

3.6.3 Stage 6.3: Fast Fourier Transform (FFT).

Fast Fourier Transform transformed method to change sound signal in time domain into sound signal into frequency domain [16]. The result of these transforms was in the form of frequency spectrums. The formula of FFT like shown on mathematic equation (2) and to find the value of frequency spectrum, mathematic equation (3) was used.

$$X[k] = \frac{1}{N}\sum_{n=1}^{N} x(n)\left(\cos\left(\frac{2\pi kn}{N}\right) - j\sin\left(\frac{2\pi kn}{N}\right)\right) \quad (2)$$

$$|F(k)| = |R^2 + I^2|^{1/2} \quad (3)$$

Where, X[k] is the result of FFT process, x(n) is the sound signal, F(k) is frequency spectrum, R is the real number from calculation result and I is imaginary number calculation result. Then complete program for Fast Fourier Transform was %-x-signal in the time domain; fft(x);

3.6.4 Stage 6.4: Spectrogram

Spectrogram is visual representative from sound frequency spectrum (or other signal) toward time or other variables. then complete program codes were % x - signal in the time domain; nfft = 4096; % number of fft points; winlen = 2096; % window length; olp = 3/4*winlen; % overlapping; win = hamming (winlen, 'periodic'); [~, f, t, P] = spectrogram (x, win, olp, nfft, fs);

3.7 Stage 7: Irisgram.

Where this research used the visual form in the shape of iris spectrogram (Irisgram). Irisgram is a circular plot of the spectrogram where the time increase azimuthally (circumferentially) clockwise, the frequency increase radially and the signal level is given axially with a color. The exact location of a given point on the plot could be observed using the data cursor. To create iris spectrogram, Matlab software was used and spectrogram form iris used the function from [17], then complete program codes for spectrogram form iris were %function irisgram; irisgram (x, win, olp, nfft, fs). The result of iris spectrogram form function can follow in figure 4.

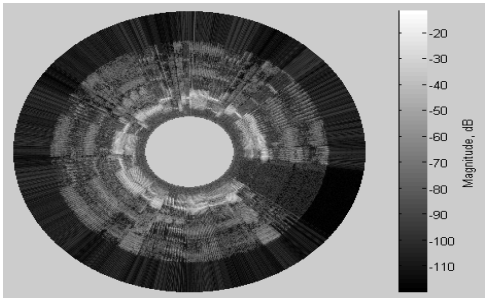

Figure 4. The result of iris spectrogram using Matlab software.

3.8 Stage 8: Spectrogram shape iris.

After the iris shape spectrogram has been obtained, an iris shape spectrogram stored in the digital image. Then program for stored in digital image were %-save to bmp; saveas (figure,'image_digital_irispectrogram.bmp');

3.9 Stage 9: Normalization on Iris spectrogram.

The features of Iris spectrogram was contained important content to be processed. The important contents were lines and spots. The important contents will be used as the reference for speaker recognition process. The normalization on iris spectrogram can facilitate the extraction of the digital image. The normalization is remapped every point on the polar coordinate iris area with formula I(r,θ) into the Cartesian coordinate area with formula with formula I(y,x) [18]. The normalization process can follow in figure 5.

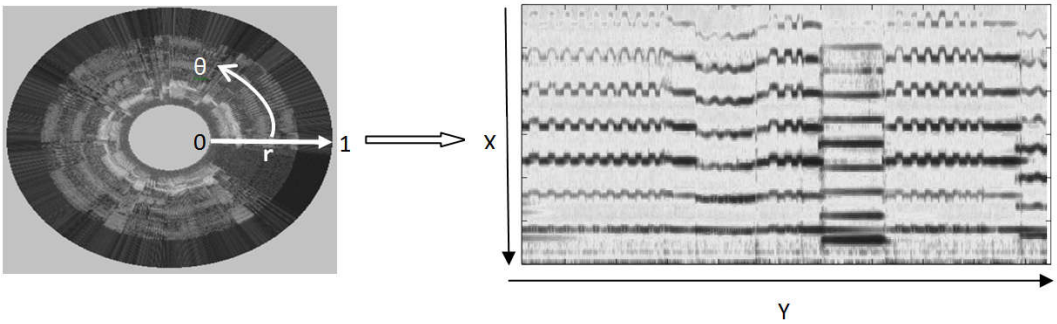

Figure 5. The illustration of Daugman Rubber Sheet Model

The proses normalization on Spectogram was changed change the polar form into a Cartesian. The formula to change can follow in (4).

$$I\big(y(r,\theta),x(r,\theta)\big) \rightarrow I(r,\theta) \quad (4)$$

The formula of y (r, θ) in (5) and the formula x(r,θ) in (6) were linear combination of pupil boundary points on image digital, the formula of ($y_p(\theta)$, $x_p(\theta)$) in (7) and the formula of ($y_i(\theta)$, $x_i(\theta)$) in (8) were the outer iris border points , (y_o,x_o) was the center of the iris/pupil, and r_x was radius of iris/pupil.

$$y(r,\theta) = (1-r)y_p(\theta) + ry_i(\theta) \quad (5)$$
$$x(r,\theta) = (1-r)x_p(\theta) + rx_i(\theta) \quad (6)$$
$$y_x = y_0 + r_x * \sin\theta \quad (7)$$
$$x_x = x_0 + r_x * \cos\theta \quad (8)$$

3.10 Stage 10: Grayscale and Resize.

After Iris spectrogram was obtained, and then the grayscale process was performed [19]. Grayscale is conversion process from RGB into grayscale with the purpose is to ease computing, and the resizing process is to normalize the size of each spectrogram that formed [20]. Program for conversion process to grayscale were %---conversion process from RGB to grayscale; Gray4=0.3*red+0.5*green+0.2*blue;

3.11 Stage 11: Various sizes

Various sizes used were 256x256, 128x128, 64x64, 32x32 and 16x16. The extraction process of image digital can be shown in figure 6.

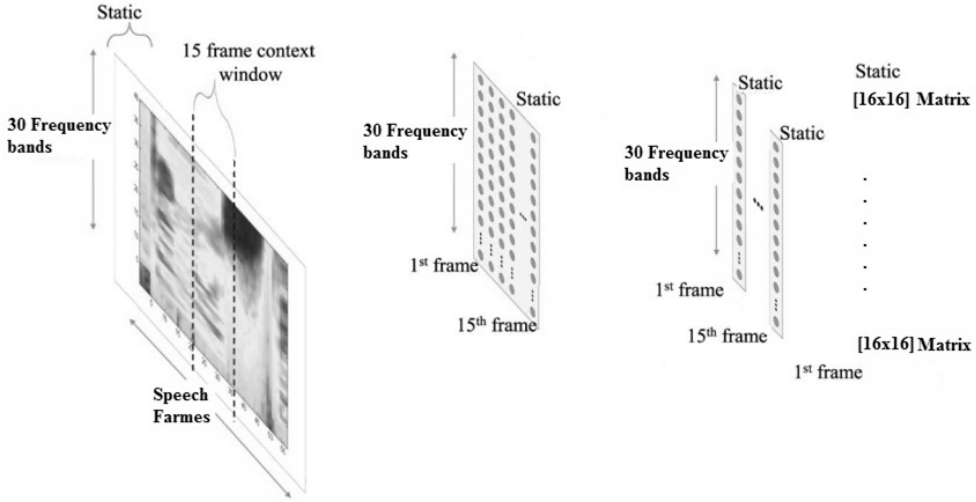

Figure 6. Matrix 16x16 used to organize speech input features to the system.

3.12 Stage 12: Feature Vector Extraction.

Using kekre transform for features extraction in this research involved kekre transform as a method to screen possibilities features that give good properties [21]. Determine vector of features using row mean the image that implemented using kekre transform like on equation (6).

$$[A] = [K][I][K]^T \quad (6)$$

Where [A] is the result of transformation, [K] is kekre matrix, [I] is image, and $[K]^T$ is matrix kekre transpose.

3.13 Stage 13: Row means image.

Features extraction that used in this research is row mean in sub-image. These features extractions were obtained by taking mean value (average) pixel on every row and the result was stored as features. Features extraction process row mean image can be shown in figure 7.

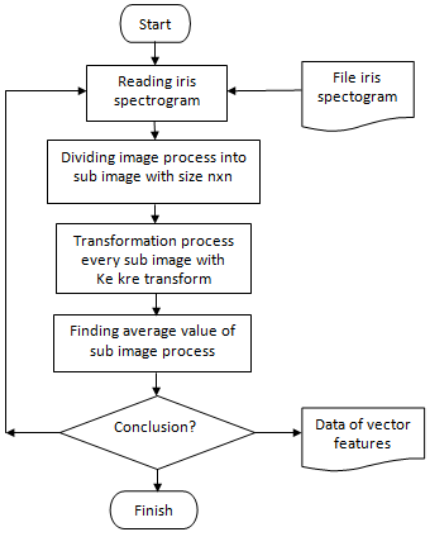

Figure 7. Flowchart of features extraction row mean image

3.14 Stage 14: **Fuzzy Mamdani method.**

Mamdani method is often known as the Max-Min method. This method was introduced by Ebrahim Mamdani in 1975. The flowchart of fuzzy Mamdani method can follow in figure 8, The input of fuzzy Mamdani method was feature vector. The feature vector was the average value of the result of the transformation kekre. The input of kekre transformation was sub-image of the digital image. In figure 8, the digital image goes into the identification process, one of the identification processes was the evaluation of error rate, if the error rate was too high then the digital image is not passed to the verification process, but the digital image is returned again in the proses of feature vectors.

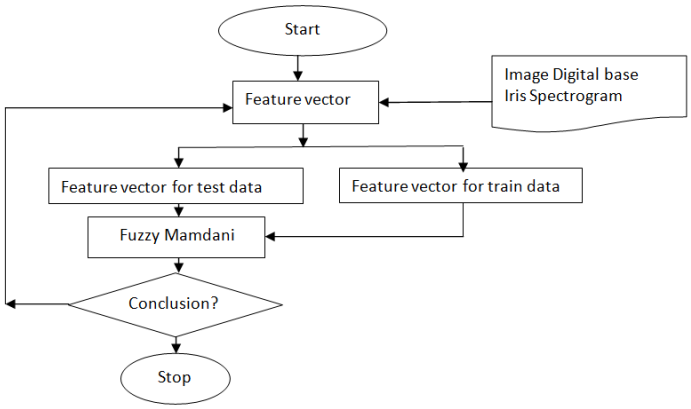

Figure 8. Flowchart for identification using fuzzy Mamdani method.

The fuzzy Mamdani method used for process identification, the fuzzy Mamdani method consists of four stages can be seen in the figure 9.

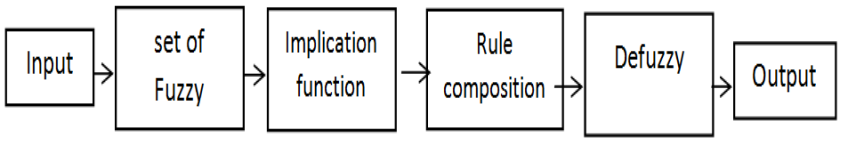

Figure 9. The explained that there are four stages, those are.

3.14.1 Stage 14.1: Formation set of fuzzy.

Features extraction result using vector features on each sub-image as input variables on fuzzy Mamdani method. Output variables were sentenced from the human voice. Value of input and output that were obtained then mapped into crisp value (numeric) into the set of fuzzy and determine membership degree in the set of fuzzy. All input and output data were processed based on set fuzzy theory. Table of the input variable for vector features sub image1 can be seen in table 1.

Table 1. Input variable for vector features sub image1

Code	Set of input fuzzy feature vector sub image 1		Domain
	Name	Notation	
1	Low	r	[0, 35]
2	Medium	s	[25, 35, 45]
3	High	t	[35,65]

In table 1, degree of membership function linearly decrease represent set of fuzzy low and degrees of membership function linearly increase for set of fuzzy high. Degree of membership function triangle used to represent set of fuzzy medium. Program codes for fuzzy mamdani method starts with creating FIS variable and adding input variable, for variable of feature_vector_sub_image_1, formation set of fuzzy with matlab were %-----Create FIS variable; a=newfis('speakerrecognition'); %---Add input feature_vector_sub_image_1; a=addvar(a,'input','feature_vector_sub_image_1',[0 65]); % Add membership function feature_vector_sub_image_1: Low, Medium, High; a=addmf(a,'input',1,'Low','trimf',[0 35]); a=addmf(a,'input',1,'Medium','trimf',[25 35 65]); a=addmf(a,'input',1,'High','trimf',[35 65]);%plotinput feature_vector_sub_image_1 to see the result; plotmf(a,'input',1);

3.14.2 Stage 14.2: Application of implication function.

After obtained the set of fuzzy for input and output, then implication function process was performed to get the output in the form of IF-THEN rule. On the input part is the degree of truth, part of antecedent and fuzzy set on the consequences part. Implication function that used was minimum. For one implication function in the form of IF-THEN rule can be seen in figure 10.

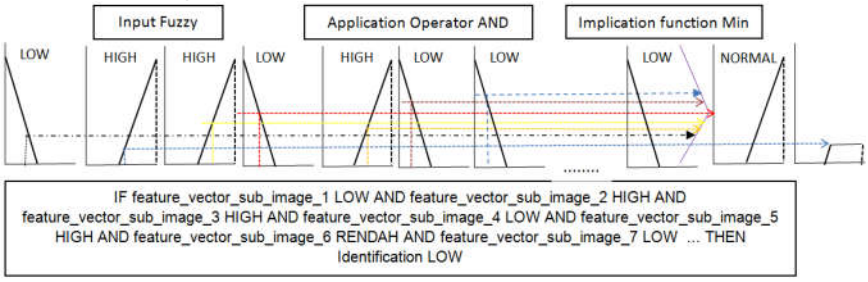

Figure 10. MIN implication function

3.14.3 Stage 14.3: Rule composition

Rule composition system that built consists of fifty-four rules, and then inference was obtained from set and correlation of fifty-four rules. In this research, inference method system fuzzy that used was max. The solution for the set of fuzzy was obtained by taking maximum value rule, then use it to modify fuzzy zone, and apply to output zone using operator OR (union). If all propositions have been evaluated, then the output will be filled up with the set of fuzzy that reflects contribution from every proposition. Generally, can be written in mathematics formula (7).

$$\mu sf[xi] \longleftarrow \max(\mu sf[xi], \mu kf[xi]) \quad (7)$$

μsf[xi] = value of solution fuzzy membership until rule number-i;

μkf[xi] = value of membership consequent fuzzy number-i;

If there are 3 rules (proposition) as follow in program codes to create rules in matlab were % -Create rules; % Rule1: IF feature_vector_sub_image_1 Low AND feature_vector_sub_image_2 High AND feature_vector_sub_image_3 High AND

feature_vector_sub_image_4 Low AND feature_vector_sub_image_5 High AND feature_vector_sub_image_6 Low AND feature_vector_sub_image_7 Low ... THEN Identification Low; % Rule2: ; % Rule3: ; rule1 = [1 1 1 1 ... 2];rule2 = [2 0 2 1 ... 0];rule3 = [3 2 3 1 ... 2];

3.14.4 Stage 14.4: Inference process.

Inference process using Max method in performing rules composition, like can be seen in figure 11. Program code for rules composition in matlab were %--------- Input rules ; listRules= [rule1;rule2;rule3];a = addrule(a,listRule);

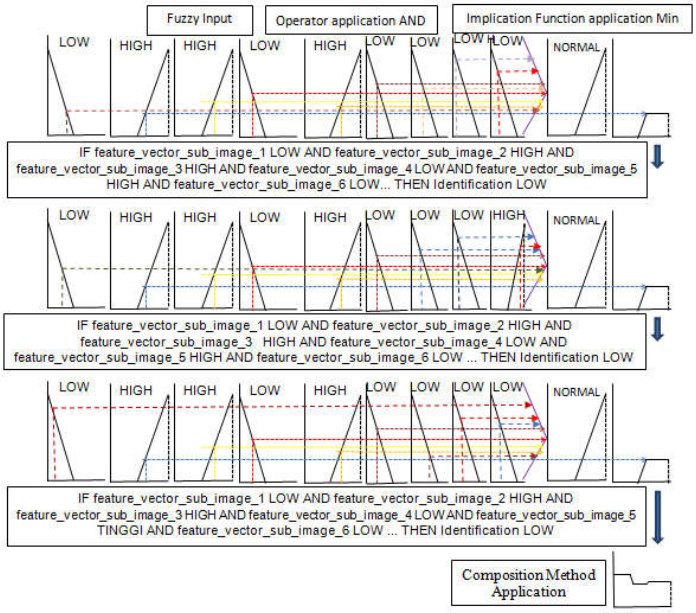

Figure 11. Fuzzy rules composition: MAX Method

3.14.5 Stage 14.5: Affirmation (defuzzy).

Input from defuzzification is a set of fuzzy that was obtained from rules of fuzzy composition, as for the output resulted from a number in the set fuzzy domain. So, if set of fuzzy in the certain range, then the crisp value must be taken as the output. Defuzzification method on rules composition rules of fuzzy Mamdani method in this research was Centroid Method (Composite Moment). In this method, the crisp solution was obtained by taking the center point (z*) fuzzy area. Generally formulized with mathematic equation (8)

$$z^* = \frac{\int_{z1} z\mu(z)dz}{\int_z \mu(z)dz} \qquad (8)$$

To get evaluation from identification system for human voice, program code were % Perform evaluation for feature_vector_sub_image_1 = 7 AND feature_vector_sub_image_2 = 8 AND feature_vector_sub_image_3 = 9 AND feature_vector_sub_image_4 = 10 AND feature_vector_sub_image_5 = 11 AND feature_vector_sub_image_6 =12 AND feature_vector_sub_image_7 =13...; evalfis ([7 8 9 10 11 12 13 …], a)

3.14.6 Stage 14.6: Evaluation of fuzzy Mamdani method.

The process evaluation of fuzzy Mamdani was performed towards form twelve of test data. Speaker recognition will be compared according to the size of data capture from the image. The process evaluation was made as error rate like can be seen in table 2, an error rate was the difference between digital image new and digital image train. In table 2, the identification was accepted if error below or equal to 13%, and then the identification proses was rejected if error above 13%, in this research the lowest error rate is 9.34%, in the table can follow with circle sign, can be indicated as the most suitable speakers. The 3rd speaker in process evaluation include accepted because the 3rd speaker has an error rate of error below or equal to 13%, but the 3rd speakers need to be verified in the next process.

Table 2. Error rate for identification of speaker recognition

Code	Speaker	Speaker recognition Size	Identification

		16x16	32x32	64x64	128x128	256 x 256	
1	Speaker 1	19.33	09.34	17.66	15.27	15.32	Accepted
2	Speaker 1	18.76	09.01	15.17	13.92	19.72	Accepted
3	Speaker 1	15.43	12.90	14.21	17.33	19.00	Accepted
4	Speaker 1	14.76	08.75	15.33	16.87	18.95	Accepted
5	Speaker 2	13.41	13.21	13.36	15.33	17.21	Rejected
6	Speaker 2	17.97	13.99	16.22	17.99	18.81	Rejected
7	Speaker 2	14.55	16.03	13.11	15.76	17.23	Rejected
8	Speaker 2	16.32	14.00	16.42	14.44	18.88	Rejected
9	Speaker 3	17.47	13.71	16.22	17.99	18.81	Rejected
10	Speaker 3	16.52	12.67	16.17	16.93	18.76	Accepted
11	Speaker 3	15.54	13.44	18.99	20.01	21.44	Rejected
12	Speaker 3	14.65	14.11	16.66	20.11	20.11	Rejected
	Mean	16.22	12.43	15.79	16.82	18.68	

Calculation result. To count error rate, in this research used Mean Absolute Percentage Error (MAPE) method, with mathematic equation (9).

$$MAPE = \frac{\sum \frac{(x_i - y_i)}{x_i} x\ 100\%}{n} \quad (9)$$

With X_i is actual data number-i, is the data of recording result and F_i is forecasting data number-i, is the result of system calculation.

In table 4, the lowest mean of error rate was size 32x32; the frame is the best frame size. The last research explains process of speech recognition is good if the frame size is 32x32 [22]. In this research mean of error rate was 12.43. Level of error rate (MAPE) that less than 40% is said to be good and dependable [23].

3.15 Stage 15: Manhattan Distance Method.

Manhattan distance method is also called "city block distance". This method is the sum of the distances from entire attributes [24]. Generally, can be written in mathematics formula (10).

$$d_{ik} = \sum |x_{ik} - c_{ik}| \quad (10)$$

d_{ik} = distance between $x_{ik}$and c_{ik}
x_{ik} = target
c_{ik} = Comparative
k = Number of attributes in each case
i = Individual attributes from 1 to n

Manhattan Distance is a similarity measurement that is most suitable for speaker recognition that represents cases relevant to natural numbers or quantitative data [25].

Manhattan Distance method used to verify a new digital image. The verify proses knowing the distance between digital images new with digital image train. The closest distance is accepted. the unacceptable digital image will be returned to the process to get the vector feature. The flowchart of Manhattan distance method can follow in figure 12,

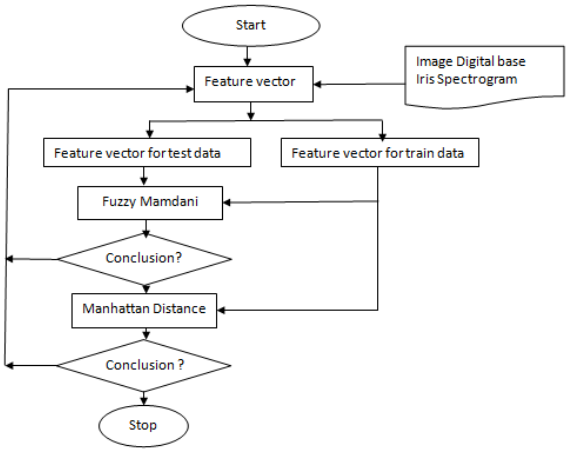

Figure 12. Flowchart for verification using Manhattan distance

3.15.1 Stage 15.1: Calculate the distance between.

Steps to be done: calculate the distance between the query and all case samples. For example, if a new case can follow table 3 and table 4.

Table 3. Calculate the distance for train data

Code	FV_1	FV_2	FV_3	FV_4	FV_5	FV_6	FV_7	Distance	K
1	10	9	8	9	4	5	9	22	1
2	4	5	7	8	9	3	4	14	2

Table 4. calculate the distance for test data

Code	FV_1	FV_2	FV_3	FV_4	FV_5	FV_6	FV_7	Distance	K
1	5	6	9	6	12	5	7	22	1

Where; FV_i = feature vector of sub image-i, K= Clasification

The distance between the new case and the old case is

|5-10|+|6-9|+|9-8|+|6-9 |+|12-4 |+|5-5 |+|7-9| = 22.

|5-4|+|6-5|+|9-7|+|6-8 |+|12-9 |+|5-3 |+|7-4| = 14.

3.15.2 Stage 15.2: Evaluation of Manhattan distance method

The evaluation process of Manhattan distance was performed towards form six test data. Speaker recognition will be compared according to the size of data capture from the image. That process was made as distance rate, like can be seen in table 5. The verification proses of speaker recognition was accepted if distance rate below or equal to 1,000, and rejected if the distance above 1,000. In table 5 obtained that the group of speakers accepted, can follow in table with dark color, but that declared most suitable has the least distance, in table was 965 with a round sign. And then the mean of low distance for size was 32x32 sizes. This size was best in the process of speaker recognition; the lowest distance is 998.75, in table with a round sign. The 3rd speaker on identification proses accepted, but the 3rd speaker in verification rejected, because the distance was above 1,000.

Table 5. Distance rate for verification of speaker recognition

Code	Speaker	Distance of Speaker recognition					Verification
		Size					
		16x16	32x32	64x64	128x128	256 x 256	
1	Speaker 1	1,079	965	1,189	1,186	1,295	Accepted
2	Speaker 1	1,285	974	1,099	1,288	2,100	Accepted
3	Speaker 1	1,088	973	1,298	1,176	2,106	Accepted
4	Speaker 1	1,108	978	1,152	1,226	1,907	Accepted
5	Speaker 3	1,195	1,083	1,101	1,090	1,109	Rejected
	Mean	1,161.75	998.75	1,171.75	1,185.75	1,652.5	

4. Conclusion

In this research, can be concluded as follow: (1). on the system for speaker recognition, has been successfully extracted features from the sample of human voice using features extraction row mean image. (2). Various size used were 256 x256, 128x128, 64x64, 32x32, and 16x16. (3). to screen possibilities features that give good properties then features vector selection using kekre transform. (4). Speaker recognition result that has been performed, was obtained introduction percentage of the human voice is as big as 87.57 % using fuzzy Mamdani method for identification and Manhattan distance method for verification. (5). Feature size 32x32 as the best feature size in this research with mean error 12.43 % using fuzzy Mamdani methods.

References

[1]. Altrock, Constantin Von. Fuzzy Logic and NeuroFuzzy Applications Explained. Prentice Hall PTR, 1995:103-104

[2] Meng, Lei, Shoulin Yin, and Xinyuan Hu. An Improved Mamdani Fuzzy Neural Networks Based on PSO Algorithm and New Parameter Optimization. *Indonesian Journal of Electrical Engineering and Computer Science*, 2016, 1(1): 201.

[3] Wang, Li-Xin. A Course in Fuzzy Systems and Control. Prentice Hall PTR, 1997:56-58

[4] Ljung. Lennart. System Identification: Theory for the User. Prentice Hall PTR, 1999:89-90

[5] Takagi T., and Sugeno M. Fuzzy Identification of Systems and Its Applications to Modeling and Control, *IEEE Transactions on Systems, Man, and Cybernetics SMC-15,* 1985:1:116–32.

[6] Dutu L. C., Mauris G., and Bolon P. A Fast and Accurate Rule-Base Generation Method for Mamdani Fuzzy Systems, *IEEE Transactions on Fuzzy Systems,* 2017; 99:1–1.

[7] Meng, Lei, Shoulin Yin, and Xinyuan Hu. An Improved Mamdani Fuzzy Neural Networks Based on PSO Algorithm and New Parameter Optimization. *Indonesian Journal of Electrical Engineering and Computer Science*, 2016, 1(1):201.

[8] Yilmaz, Atinc, and Kursat Ayan. Cancer Risk Analysis by Fuzzy Logic Approach and Performance Status of the Model, *Turkish Journal of Electrical Engineering & Computer Sciences*, 2013; 21(3): 897–912.
[9] Silva W. L. S., and Serra G. L. d. O. *Proposal of an Intelligent Speech Recognition System*. In 2012 Third Global Congress on Intelligent Systems, 2012; 1:356–359.
[10] Kekre H. B., and Shah K. *Performance Comparison of Kekre's Transform with PCA and Other Conventional Orthogonal Transforms for Face Recognition*. In 2009 Second International Conference on Emerging Trends in Engineering & Technology,2009;1:873–79.
[11] Kekre, H. B., Vaishali Kulkarni, Sunil Venkatraman, Anshu Priya, and Sujatha Narashiman. Speaker Identification Using Row Mean of DCT and Walsh Hadamard Transform, *International Journal on Computer Science and Engineering* , 2011;3(1):47-56
[12] Permana, Inggih. A Comparative Study on Similarity Measurement in Noisy Voice Speaker Identification. *Indonesian Journal of Electrical Engineering and Computer Science*, 2016, 1(3): 590.
[13] Aragonda H., and Seelamantula C. S. *Riesz-Transform-Based Demodulation of Narrowband Spectrograms of Voiced Speech*, In 2013 IEEE International Conference on Acoustics, Speech and Signal Processing,2013;1: 8203–7.
[14] Krawczyk M., and Gerkmann T. STFT Phase Reconstruction in Voiced Speech for an Improved Single-Channel Speech Enhancement, *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* 2014; 22(12):1931–40.
[15] Thepade S. D., and Kalbhor M. M. *Image Cataloging Using Bayes, Function, Lazy, Rule, Tree Classifier Families with Row Mean of Fourier Transformed Image Content*, In 2015 International Conference on Information Processing (ICIP),2015;1:680–84.
[16] Ezhil, Shenbaga SS. Real Time Application of Fourier Transforms. *Indonesian Journal of Electrical Engineering and Computer Science*, 2017, 8(2): 574–577.
[17] https://www.mathworks.com/matlabcentral/fileexchange/64882-spectrogram-visualization-with-matlab-implementation?requestedDomain=true
[18] Daugman, J. How iris recognition works. *IEEE Transactions On CircuitsAnd Systems For Video Technology*, 2002; 14(1):103-114
[19] Kekre H. B., Mishra D., and Saboo R. S. *Comparison of Image Fusion Techniques in RGB & Kekre's LUV Color Space*, In 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE). 2011; 1:114–20.
[20] Patil S., Bhangale U., and More N. *Comparative Study of Color Iris Recognition: DCT vs. Vector Quantization Approaches in Rgb and Hsv Color Spaces*. In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI).2017; 1:1600–1603.
[21] Kekre, H. B., and Dhirendra Mishra. Sectorization of Full Kekre's Wavelet Transform for Feature Extraction of Color Images. *International Journal of Advanced Computer Science and Applications*. (2011;2(2): 69–74.
[22] Kulkarni, V., Kekre, H.B. Gaikar, P. and Gupta, N., 2012, Speaker Identification using Spectrogram of Varying Frame Sizes, *International Journal of Computer Applications.*2012; 50(20):27-33.
[23] Brooks, Peter. Metrics for Service Management. Van Haren, 2012:301
[24] Muhammad K. Farooq, Malik Jahan Khan, Shafay Shamail, and Mian M. Awais. *Intelligent project approval cycle for local government: case-based reasoning approach.* In Proceedings of the 3rd international conference on Theory and practice of electronic governance (ICEGOV '09), Tomasz Janowski and Jim Davies (Eds.). ACM, New York, NY, USA. 2009; 1; 68-73.
[25] Cheung, Sai On, Peter Shek Pui Wong, Ada S.Y. Fung, and W.V. Coffey. Predicting Project Performance through Neural Networks. *International Journal of Project Management*. 2006; 24(3):