Engineering and Technology Journal
Journal homepage: engtechjournal.org

A New Method Using Naive Bayes And RGBD Facial Identification Based on Extracted Features from Image Pixels

Wisam H. Ali [a*]

[a] Lecturer, University of technology-Electrical Dpt., Baghdad, Iraq, 30088@uotechnology.edu.iq

* Corresponding author.

Submitted: 24/11/2020 Accepted: 06/01/2021 Published: 25/04/2021

KEYWORDS

RGBD, Bayesian-SVM, lips recognition, feature extraction

ABSTRACT

Nowadays, life seems to have been resilient, particularly for those with physical disabilities. Recognition of AV letters is one of the critical and famously the difficult structures. This research has been developed based on the potential of the features in some applications than the statistical properties. While, these features have been resolved the lip movement for AV letters recognition, Naive Bayesian and Red green blue and depth RGBD have been adopted for visual letter identification. Naive Bayesian has 73.33% for usual recognition with three letters, each with ten frames, while RGBD classifier is 100%. Within that for this case, two scenarios were made with different forms of noise placed on the face of normal, normal + 10%, normal + 25% and normal + 75% noise. The first one trains and understands all classes, one after another. While the other is training 95 percent of RGBD and 83.3 percent for Naive Bayesian with recognition of one of the inflicted forms. RGBD identification is 100 percent for the second one, while 49.99 for the Naive Bayesian.

How to cite this article:, W. H. Ali “A new method using naive Bayes and RGBD facial identification based on extracted features from image pixels,” Engineering and Technology Journal, Vol. 39, Part A, No. 04, pp. 632-64, 2021.
DOI: https://doi.org/10.30684/etj.v39i4A.1936

1. Introduction

Human susceptibility to discrimination and recognition is not matched by any intelligent systems as well as discrimination in noise [1]. Speech recognition has been notoriously hard, therefore, with the present development of technologies, it becomes practically efficient. Besides, the standard of life for some people with particular needs and applications has been better [2]. One of these technologies incorporates the visual information into a speech recognition system for enhancing the accuracy of that system especially in a noise environment [3]. In this context, lip-reading has been an important area of the above incorporate [4], especially for speech recognition without speech, which is called silence or lips language. Accordingly, the motion of the lips must be recognized, while this task is difficult because it has a nonlinear model and the region of interest is noisy [5]. Despite these difficulties, the recognition can be satisfied by track the changing of lip shape and then extracted its feature. Many algorithms can lead to this purpose as well as the facial image processing morphological operation [2 and 6]. In this context, the boundaries and colors have been improved the robustness and efficiency of lip features [7].

In recent years, several significant worth studies of lip-reading systems have been presented. These studies were concentrated on three tasks, the first has been focused on the mouth area detection such as [7, 4, 3, 5], while the second has been concentrated on the speech extracted as [1, 7], finally, this task has been revived on the conformity between the visual features and particular vocabulary like[1, 8, 2, 9, 10].

A time-delay neural network (TDNN) has been presented by [1]; the recognition percentage has been improved from 51%, without acoustic effect, up to the 91% with it. While Petajan et al.[11] has been used a threshold image and hamming distance for utterances discrimination. Whilst, Pentland [12] has been using an optical flow technique for lip points estimation. In this context, [4and 8] have been proposed a segmentation technique for the lip contour extraction to perform a lip reading. While, [3] has been working on the same task, but with more lip points with the aid of fuzzy clustering. [9 and 10] has been estimated a lip contour and then color features are extracted. [5] Has been used a hybrid Discrete Cosine Transform and Dual-tree Complex Wavelet Transform for estimating the lip position and shape.

The work of this paper has been concentrated on the recognition of AV letters from a video frame sequence using a snapshot database. This recognition has been made from a segment of the mouth area. Then, the preprocessing operation has been achieved for preparing the lip area to extract the features. The probability of occurrence of the recognition later has been calculated by using Bayesian-SVM for recognizing the operative letters. Color means intensity has been erected by using RGBD for accurate segmentation, which tends to accurate recognition.

2. The Combining Bayesian and SVM

Combining Bayesian and SVM have been used in this algorithm. Direct Bayesian SVM has been used for multi-class classification. Where the Bayesian algorithm already converts the multi-class classification problem into a two-class problem for the intra- class classification and the extra-class classification variation. Then, SVM has been trained for two-class features. Where, in the training phase, firstly computed the image difference between images of the same class to construct the intra- class variation set as **[13];**

$$\{\Delta_I|\Delta_I\} \in \Omega_I \tag{1}$$

Where Δ_I is vector image set. In this context, calculate image difference between images of different class to construct the extra-class variation set as;

$$\{\Delta_E|\Delta_E\} \in \Omega_E \tag{2}$$

Where Δ_E is the next vector image set Then, the eigenvalue matrix Λ_I, and the eigenvector matrix $ʉ_I$, of the intra-personal subspace has been computed from the intra-class variation set. Finally, all the image difference vectors are projected and whitened in the intra-class subspace as;

$$\Delta_I' = \Lambda_I^{-\frac{1}{2}} ʉ_I^T \Delta_I \tag{3}$$

$$\Delta_E' = \Lambda_I^{-\frac{1}{2}} ʉ_I^T \Delta_E \tag{4}$$

Then, the decision function $f(\Delta)$ has been generated by training the SVM with these two vector sets (Δ_I' and Δ_E') .

In the testing phase, the difference vector Δ_I between probe vector x and each gallery vector x_i^g, and then, project and whiten the difference vector in the intra-class subspace, and then, the final classification decision is;

$$d(x) = \arg \max_{1 \le i \le c}(f(\Delta_i')) \tag{5}$$

Where;

c - number of classes in the gallery.

The larger is the value of d, the more reliable the result is.

3. Database

The database has been recorded audio-visual of isolated Audio-video AVlatters. In this work, three letters have been taken (A, B, and E) with five, eight, and ten repeating talkers respectively. Each video letter has ten frames to complete the lips pronounce for each letter. In this context, each class has been considered as all lips pronounce letters for all talkers.

4. Proposed Algorithms

Two algorithms have been proposed for recognized the AV letters; the first one is the combining Bayesian and SVM. Where the Bayesian-SVM for AVletter classification are; Bayesian stage **[14,15 and 16]**

$$p(h - AVletter/D - training\ data) = \frac{p(D/h)\, p(h)}{p(D)} \qquad (6)$$

Where;

$p(h)$- probability of occurrence of class *h*.

$p(D)$- probability of an instancec occurring (training data) *D*.

$p(h/D)$- The probability of instance *h* being in class *D*.

$p(D/h)$- The probability of generating instance *D* given class *h*.

$$p(h/D) = \frac{p(D \cap h)\backslash\, number\ of\ h\ and\ D\backslash}{p(D)\ \backslash number\ of\ D\backslash} \qquad (7)$$

$$\left.\begin{array}{r} \text{Then;} \\ \Delta'_{\mathrm{I}} = p(h/D) \\ \Delta'_{\mathrm{E}} = p(h/\overline{D}) \end{array}\right\} \qquad (8)$$

where;

$p(h/\overline{D})$- h not in class D

Then, the SVM has done, the final classification decision as;

$$\mathrm{d(x)} = \arg\max_{1\le i\le c}(f(\Delta'_{\mathrm{i}})) \qquad (9)$$

where;

$f(\Delta'_{\mathrm{i}})$ - SVM classification decision between Δ'_{I} *and* Δ'_{E}

In this context, figure (1) represents the process steps while figure (2) represents the flowchart of this algorithm

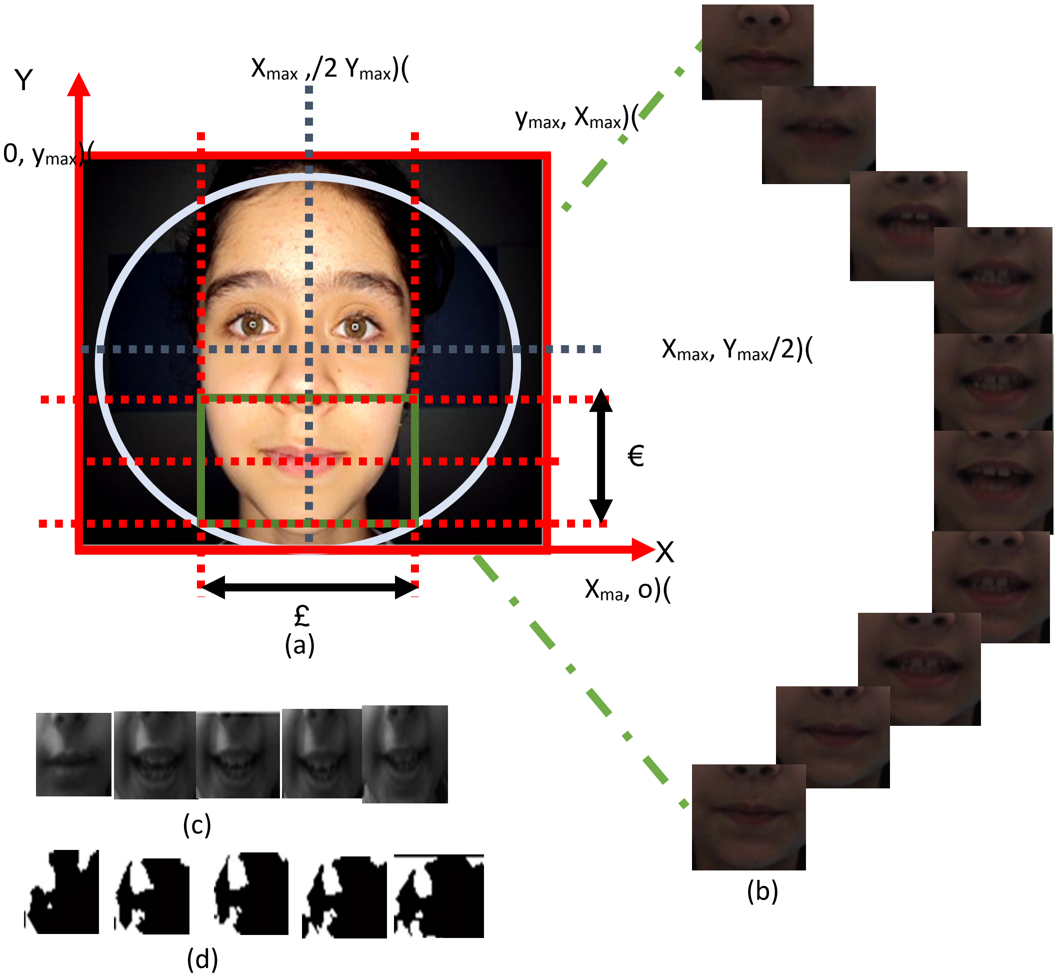


Figure 1: Algorithm Process (a) Lip Region Determination (b) Lip AVletter Frames (c) Gray Conversion (d) Binary Conversion

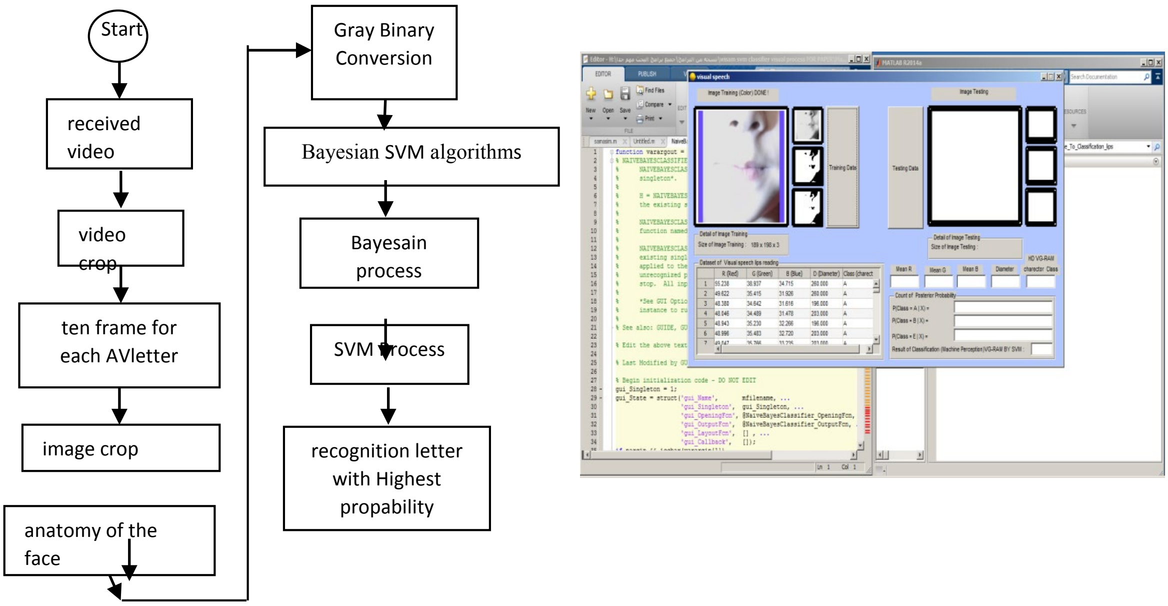


Figure 2: a) Flowchart of the Proposed Bayesian SVM Algorithm for AVletters Recognition, b) GUI

The results of this algorithm are shown in Table (I)

TABLE I: Bayesian SVM Algorithm Results

Test No.	Image size	P(A/X)	P(B/X)	P(Z/X)	Classification
	179x296x3	5.995e-008	4.5713e-007	3.688e-015	B
1	152x252x3	1.488e-005	2.0167e-009	5.5441e-008	A
2	152x252x3	8.3745e-006	1.2139e-010	3.1527e-008	A
3	152x252x3	9.4214e-006	1.0653e-010	8.6908e-009	A
4	152x252x3	2.0934e-005	1.1452e-008	3.2172e-009	A
5	160x240x3	2.4471e-013	1.0824e-005	2.8744e-007	B
6	170x250x3	2.4218e-006	3.0604e-006	3.9788e-009	B
7	170x250x3	2.7363e-006	1.7159e-006	3.7952e-010	A
8	170x250x3	1.4585e-006	2.1876e-006	1.0762e-010	B
9	17x250x3	4.1865e-007	6.244e-006	3.787e-010	B
11	170x250x3	2.9498e-010	3.3992e-005	3.5619e-008	B
12	170x240x3	6.4044e-013	1.2049e-005	4.2791e-007	B
13	170x240x3	4.2946e-014	4.9904e-006	6.4229e-007	B
14	170x250x3	1.3829e-015	3.0165e-006	9.0041e-008	B
15	170x250x3	2.2796e-013	9.9548e-006	2.8029e-007	B
16	144x236x3	8.9298e-009	2.2929e-008	2.5304e-006	Z
17	144x236x3	4.2464e-007	1.0358e-009	3.6456e-006	Z
18	144x236x3	5.8306e-007	9.4794e-012	1.208e-006	Z
19	144x236x3	1.4506e-006	4.7662e-011	1.1868e-006	A
20	144x236x3	4.0247e-006	7.576e-008	2.673e-006	A
21	144x236x3	4.1478e-010	5.1181e-006	4.8391e-006	B
22	144x236x3	8.5532e-015	2.1169e-006	6.4103e-007	B
23	144x236x3	1.0806e-019	7.4232e-009	8.7005e-008	Z

The second algorithm is the Red, Green, Blue, and diameter (RGBD) feature as in figure (3). This algorithm has been based on the effect of the luminance of the image. Where, this effect has been taken by calculating the color intensity (number of pixels of red, green, and blue of the image) for each frame, and then, calculate the diameter of crop image of the lip region, where this algorithm called RGBD feature.

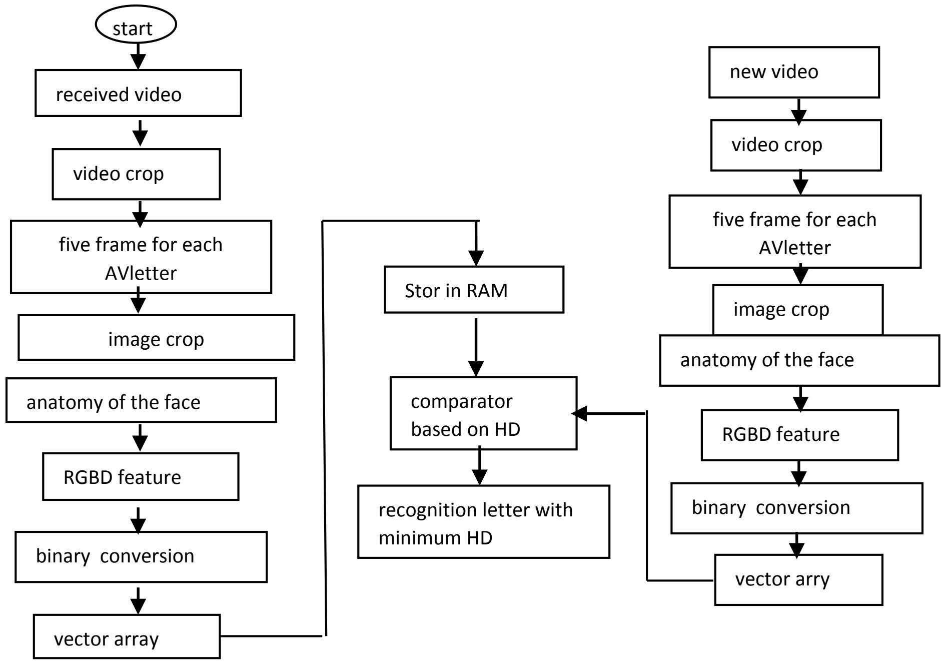


Figure 3: Flowchart of the Proposed RGBD Algorithm for AVletters Recognition

This algorithm has been based on several steps; these are

1. The Subsector of image (video crop) then image crop at each frame as in figure (4-b). This cropping has been based according to the anatomy of the face.

The anatomy of the face as in figure (4-a) was based on the size of the image (X_{max} , Y_{max}). Where, divided it into the vertical half ($X_{max}/2$, Y_{max}) - ($X_{max}/2$, 0). Then at the horizontal half at ($X_{max}/2$, $Y_{max}/2$) - (X_{max} , $Y_{max}/2$). In this context, take the vertical lines from the pupil with a distance between them is (£), and from the hidden lip line center of the upper and lower is (€). Then, the lip area is (£ × €).

2. The intensity of colors (RGB) and diameter (D) of the lip area has been calculated as (RGBD) features according to the Eqs. (10-13).

2. Convert the RGBD features to binary for each image frame of AV letter (ten frames for each AV letter).

3. Then, ten vectors for each AV letters have been saved in RAM as in figure (3).

4. Repeat the above steps for each training AVletters and save it in a particular RAM.

5. In the testing phase, with the new AV letter apply the above steps, and then, compare the resultant vectors with that storage. Then, apply Hamming distance (HD) for deciding the corrected one that has a smaller HD as in figure (3).

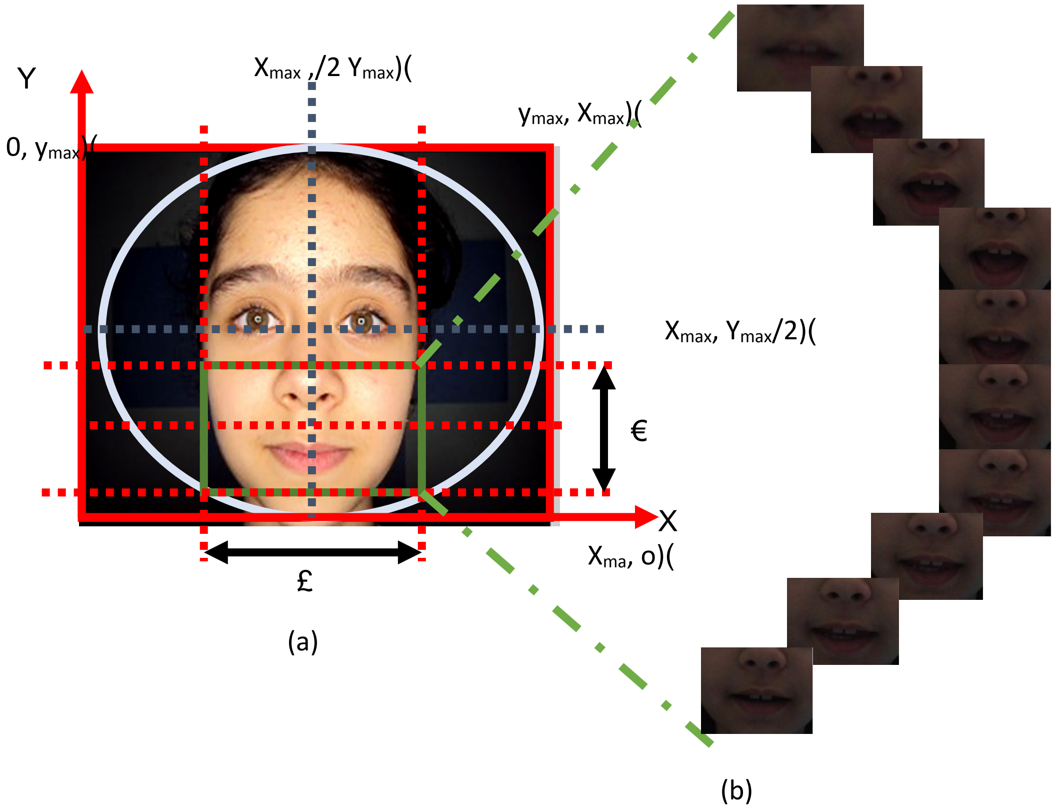


Figure 4: Algorithm Process (a) Lip Region Determination (b) Lip AVletter Frames

$$red\ color\ intensity = \frac{number\ of\ red\ pixels}{total\ number\ of\ pixels} \tag{10}$$

$$green\ color\ intensity = \frac{number\ of\ green\ pixels}{total\ number\ of\ pixels} \tag{11}$$

$$blue\ color\ intensity = \frac{number\ of\ blue\ pixels}{total\ number\ of\ pixels} \tag{12}$$

$$Diameter = D(lip\ area(\ £\ \times\ €) \tag{13}$$

The results of this algorithm are shown in Table (II) below.

TABLE II: The Results of RGBD Algorithm

Test No.	R(Red) intensity	G(Green) intensity	B(Blue) intensity	D(Diameter) pixels	RGBD AVletter recognition
1	67.194	56.287	53.199	260.000	A
2	61.949	50.163	47.427	233.000	A
3	60.739	49.228	46.295	233.000	A
4	60.899	49.157	46.087	236.000	A
5	63.105	51.157	47.764	240.000	A
6	73.510	60.490	56.708	226.000	B
7	66.557	54.245	51.397	241.000	B
8	66.455	54.037	51.184	245.000	B
9	66.808	54.438	51.719	247.000	B
10	67.838	55.297	52.387	245.000	B
11	70.958	58.499	55.574	235.000	B
12	72.751	60.080	57.160	225.000	B
13	73.281	60.782	57.747	221.000	B
14	74.057	62.137	58.919	227.000	B
15	72.572	60.901	57.645	226.000	B
16	64.448	55.944	51.928	203.000	Z
17	61.280	52.592	49.031	210.000	Z
18	58.903	49.906	46.750	216.000	Z
19	59.481	50.186	47.046	219.000	Z
20	62.963	53.026	49.772	222.000	Z
21	68.804	58.424	54.986	215.000	Z
22	72.139	62.252	58.740	219.000	Z
23	74.310	64.184	60.241	204.000	Z

Th
e comparison between these two algorithms are shown in Table (III).

TABLE III: Recognition Comparison

algorithm	RGBD feature	Bayesian- SVM
Recognition %	100%	73.33%

Then two scenarios have been made for improving the robustness of the proposed algorithms. The training classes are; Normal, normal+10%, normal+25%, and normal+75%. Where these increments to the normal case represent the lighting inflicted on the face of classes as shown in figure (5).

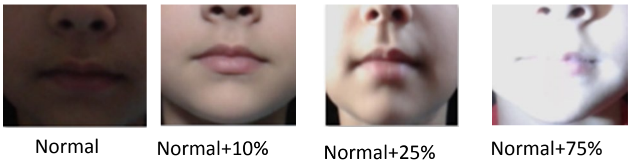


Figure 5: Different Lighting Inflicted on the Face of Classes

The first scenario has been made by training all classes (Normal, normal+10%, normal+25%, and normal+75%) which represent three letters A, B, and Z. Where, each class has ten frames. Then, at the testing phase, the testing has been made to the individual class. The results of this scenario are shown in Table (IV).

TABLE IV: The Results of the First Scenario

Training	Test	Noise Level	Charter	Recognition	
				RGBD feature Algorithm	Bayesian SVM Algorithm
All classes	Normal	0.0	Z	Z	A
		0.2	Z	Z	A
		0.4	Z	Z	B
		0.6	Z	Z	Z
		0.8	Z	Z	Z
		0.9	Z	Z	Z
	Normal+10%	0.0	Z	Z	B
		0.2	Z	Z	Z
		0.4	Z	Z	Z
		0.6	Z	Z	Z
		0.8	Z	Z	Z
		0.9	Z	Z	Z
	Normal+25%	0.0	Z	Z	Z
		0.2	Z	Z	Z
		0.4	Z	Z	Z
		0.6	Z	Z	Z
		0.8	Z	Z	Z
		0.9	Z	Unknown	Z
	Normal+75%	0.0	Z	Z	Z
		0.2	Z	Z	Z
		0.4	Z	Z	Z
		0.6	Z	Z	Z
		0.8	Z	Z	Z
		0.9	Z	Z	Z

While the second scenario has been made by testing all classes when training one class. The results of this second scenario are shown in Table (V).

TABLE V: The Results of the Second Scenario

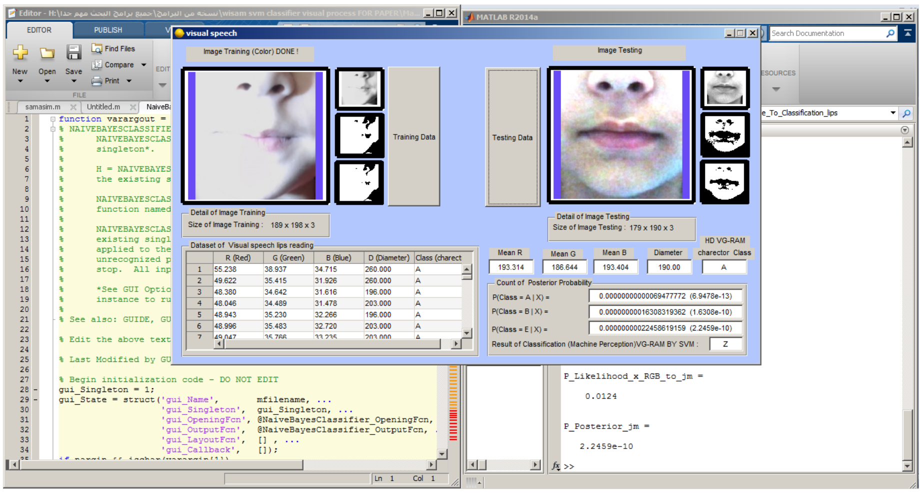
Training	Testing	Character	Recognition	
			RGBD feature	Bayesian SVM
Normal	Normal+10%	Z	Z	A
	Normal+25%	Z	Z	A
	Normal+75%	Z	Z	A
Normal+10%	Normal	Z	Z	B
	Normal+25%	Z	Z	B
	Normal+75%	Z	Z	B
Normal+25%	Normal	Z	Z	A
	Normal+10%	Z	Z	B
	Normal+75%	Z	Z	B
Normal+75%	Normal	Z	Z	A
	Normal+10%	Z	Z	A
	Normal+25%	Z	Z	B

Table (VI) represent the differences in the recognition decision of these two algorithms with the two scenarios as stated before.

TABLE VI: Recognition Comparison

algorithm	RGBD feature	Bayesian SVM	Scenario
Recognition %	95,8%	83.3%	1
	100%	49.997%	2

Figure (6), shows snapshot of the GUI program built to test the unknown visual lips image after training the pronounce of the training letters. While figure (7) represent the two proposed classifier output throughout the testing phase.



```
function varargout =
% NAIVEBAYESCLASSIFIE
%       NAIVEBAYESCLAS
%       singleton*.
%
%       H = NAIVEBAYES
%       the existing s
%
%       NAIVEBAYESCLAS
%       function named
%
%       NAIVEBAYESCLAS
%       existing singl
%       applied to the
%       unrecognized p
%       stop.  All inp
%
%       *See GUI Optio
%       instance to ru
%
% See also: GUIDE, GU

% Edit the above text

% Last Modified by GU

% Begin initialization code - DO NOT EDIT
gui_Singleton = 1;
gui_State = struct('gui_Name',       mfilename, ...
                   'gui_Singleton',  gui_Singleton, ...
                   'gui_OpeningFcn', @NaiveBayesClassifier_OpeningFcn, ...
                   'gui_OutputFcn',  @NaiveBayesClassifier_OutputFcn, ...
                   'gui_LayoutFcn',  [] , ...
                   'gui_Callback',   []);
```

Detail of Image Training

Size of Image Training : 189 x 198 x 3

Dataset of Visual speech lips reading

	R (Red)	G (Green)	B (Blue)	D (Diameter)	Class (charect
1	55.238	38.937	34.715	260.000	A
2	49.622	35.415	31.926	260.000	A
3	48.380	34.642	31.616	196.000	A
4	48.046	34.489	31.478	203.000	A
5	48.943	35.230	32.266	196.000	A
6	48.996	35.483	32.720	203.000	A
7	49.047	35.766	33.235	203.000	A

Detail of Image Testing

Size of Image Testing : 179 x 190 x 3

Mean R	Mean G	Mean B	Diameter	HD VG-RAM charector Class
193.314	186.644	193.404	190.00	A

Count of Posterior Probability

P(Class = A | X) = 0.00000000000069477772 (6.9478e-13)

P(Class = B | X) = 0.00000000016308319362 (1.6308e-10)

P(Class = E | X) = 0.00000000022458619159 (2.2459e-10)

Result of Classification (Machine Perception)VG-RAM BY SVM : Z

```
P_Likelihood_x_RGB_to_jm =

    0.0124


P_Posterior_jm =

   2.2459e-10

>>
```

Figure 7: (a) NaiveBayes- SVM and (b) RGBD

5. Discussion of the Results

From the above results in Tables (I-VI), especially tables (3 and 6), it is clear that the RGBD algorithm is progressing (better classification rate) other than the Bayesian SVM algorithm in recognition of the lips movement AV letters. The main reason for reducing the success rate in the Bayesian SVM algorithm is that this algorithm depends on the facial features region not only on the image pixel so there is interference between these features on selected images. Also, there is another important reason, where, this type of application has many patterns in each class. Where some of these patterns have a similarity with the other in another class. These similarities mean having the same statistical properties, then, tend to miss the recognition.

6. Conclusion

Many points can be concluded from this work. For the proposed application, the RGBD algorithm is more accurate in classifying the pronounced word rather than a Bayesian SVM algorithm. While the Bayesian SVM algorithm is faster than RGBD algorithm. On the other hand, the proposed application (identifying the pronounced word based on image pixel) needs to extract many visual features from the sets of the isolated image that represent the pronounced word for more accurate recognition. Therefore, the RGBD algorithm has the advancing than the SVM algorithm by 1.42 times with the same type of visual feature.

References

[1] D.G. Stork, G. Wolff, E. Levine, Neural network lip reading system for improved speech recognition, Int. Jt. Conf. Neural. Netw., 2(1992). https://doi.org/10.1109/IJCNN.1992.226994

[2] Z. Bin , F. Yasuhiro , Research on an Automated Speech Pattern Recognition System Based on Lip Movement, Int. Conf. Proc. IEEE. Eng. Med. Biol. Soc., (1996). https://doi.org/10.1109/IEMBS.1996.647537

[3] S.L. Wang, W.H. Lau , S.H. Leung, A New Real-Time Lip Control Extraction Algorithm, IEEE Int. Conf. Acoust. Speech. Signal. Proc., 3 (2003) 217-20. https://doi.org/10.1109/ICASSP.2003.1199146

[4] G. S. Luciiana, F. Jacques , L. B. Dbio ,Visual Speech Recognition: a solution from feature extraction to words classification, XVI Brazilian Symposium on, Comp. Graph. Image. Proc. (2003).

[5] Y. Zhang, L .Yuan, Hybrid Lip Shape Feature Extraction and Recognition for Human-Machine Interaction, Int. J. Model. Identif., 18 (2013) 191-198. https://doi.org/10.1504/IJMIC.2013.052812

[6] B. Sujatha , T. Santhanam, Classical Flexible Lip Model Based Relative Weight Finder for Better Lip Reading Utilizing Multi Aspect Lip Geometry, J. Comp. Sci., 6 (2010) 1065-1069 . https://doi.org/10.3844/jcssp.2010.1065.1069

[7] X. Zhang , R.M. Mersereau, Lip Feature Extraction towards an Automatic Speech reading System, Int. Conf. Image. Proc., 3 (2000). https://doi.org/10.1109/ICIP.2000.899336

[8] M. Iain, F. C. Timothy, B. Andrew , C. Stephen , H. Richard , Extraction of Visual Features for Lipreading, IEEE Trans. Pattern. Analy. Pattern. Anal. Mach. Intell., 24 (2002).

[9] T. Wark, S. Sridharan, A Syntactic Approach to Automatic Lip Feature Extraction for Speaker Identification, IEEE EURASIP. J. Adv. Signal. Process., 6 (1998).

[10] D. Philippe , D. Paul, Statistical Lip-Appearance Models Trained Automatically Using Audio Information, EURASIP J. on Applied Signal Processing:11(2002)1202–1212 Hindaw i Publishing Corporation. https://doi.org/10.1155/S1110865702206186

[11] E. D. Petajan, Automatic lipreading to enhance speech recognition Proceedings of the IEEE Communication Society Global Telecommunications Corgference, November. (1984) 26-29. Atlanta, Georgia

[12] A. Pentland , K. Mase, Lip reading: Automatic visual recognition of spoken words, Proc. Image Understanding and Machim Vision, Optical Society of America. (1989)12-14 .

[13] L. Zhifeng , T. Xiaoou, Bayesian face recognition using support vector machine and face clustering, IEEE Proc. Comp. Soc. Conf. Comp. Vision. Pattern. Reco., 2004 (CVPR'04). https://doi.org/10.1109/CVPR.2004.1315188

[14] K. Eamonn ,Naïve Bayes Classifier" Pattern Recognition and Machine Learning,Christopher Bishop, Springer-Verlag, 2006.

[15] B. M. Amel , I. K. Mohamed , A. Mohamed , Naive Bayesian Fusion for Action Recognition from Kinect, Int.Conf. Comp.Net. Data. Comm., (2017).

[16] M. Bansal, M. Kumar, M. Kumar, 2D Object Recognition Techniques: State-of-the-Art Work. Arch. Comput. Methods . Eng., 28 (2020) 1147–1161. https://doi.org/10.1007/s11831-020-09409-1