Optik 125 (2014) 5181–5186

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

A novel lip reading algorithm by using localized ACM and HMM: Tested for digit recognition

Sunil S. Morade[a,*], Suprava Patnaik[b]

[a] Department of Electronics Engineering, SVNIT, Surat, India
[b] Department of E and TC Engineering, Xavier Institute of Engineering, Mumbai, India

ARTICLE INFO

Article history:
Received 5 September 2013
Accepted 1 May 2014

Keywords:
Geometrical parameter
Hidden markov model
Lip reading
Lip tracking
Localized active contour model

ABSTRACT

Lip contour tracking is an integral part of lip reading application. Fast and accurate lip tracking is an important step in lip reading. This paper uses a novel active contour model for lip tracking and proposes geometrical feature extraction approach for lip reading. Effect of individual features are compared and a joint feature model is obtained by combining weighted decision obtained by a feature vector of difference in inner area, height and width of lip. Ergodic hidden markov model (HMM) is used as a classifier. For each digit Markov Model is tested for 3 states and 5 states. Videos of English digit from 0 to 9 have been recorded for recognition test. Cuave database is used for comparison along with an in-house database. While doing computation of feature vectors, only significant frames are used to reduce the computation complexity. Results of experimentations on digit utterances are given to show that the maximum recognized digit can be used for important programming command of computerized numerical control machines.

1. Introduction

Lip reading means seeing the movement of lips which is used to collect speech information. Efficient speech recognition is useful for physically handicap person, hearing impaired persons and operating machines in noisy area with speech command. McGurk and MacDonald [1] in his experiment concluded that video information has impact on speech recognition. Due to mismatch utterance of audio and video, recognized word will be different. Best lip reader can understand the speech by lip movement. Lip reading is speech reading process from visual information of lip. From long days it is well known that visual speech information is important for speech recognition in noisy area. Audio information is affected by acoustic noise and crosstalk noise among speakers. So in noisy area it will be beneficial to use visual information along with audio. Most of the speech recognition systems are based on digit utterance. Many researchers have tested their model on digit as database. So in this paper we have used 0–9 digit as database. Visual digit recognition is also useful while dialing number from noisy area and in user interactive system.

Basically human provide very less visual information related to speech which is one of the challenges of lip reading. Other challenges of lip reading are frame rate while recording, light conditions, video quality, few visemes and visual appearance differences between individuals. Identical visual information is obtained for different words, which is also a major concern.

2. Related work

For lip reading mouth area of the face is important. Visual features are classified into two categories. In category one is geometric features are considered in spatial domain. Second category is image transform domain which includes spectrum analysis. Lip parameterization may be geometric based or image transform based. Geometric features are less complex and easy to extract, as compared to image transform features. To our knowledge of literature survey, first geometrical based feature system was developed by Petajan [2]. In their system simple thresholding of the mouth image is used to highlight the lip area, and then measurements of mouth height, width, and area were taken. Since then, many approaches have been developed which exploit our knowledge of the shape of a human mouth to fit more complex models to speakers' mouth. Kaynak et al. used width, height, area and angle parameter of a lip as geometric parameter and concluded that recognition rate is higher for combination of these parameters [3]. Wang et al. used the set of visual features of width, height along with shape and

* Corresponding author at: Department of E and TC Engineering, K.K. Wagh Institute of Engineering Education & Research, Nashik 422003, India.
Tel.: +91 253 2513692.
E-mail addresses: ssm.eltx@gmail.com (S.S. Morade), suprava_patnaik@yahoo.com (S. Patnaik).

http://dx.doi.org/10.1016/j.ijleo.2014.05.011
0030-4026/

Table 1
Human lip reading results [5].

Subject	% Video word speech information
Female1	61
Male1	50
Male 2	37
Female 2	63

inner mouth parameter which resulted increase in recognition rate [4]. Hassant and Jassim used both geometric and image transform features. They used width, height and dynamic information of lip as a feature vector [5].

Efficient classification of feature vectors is useful because final result after feature extraction will depends on classifier. For lip reading KNN, Neural network, SVM and HMM classifiers are mostly used. HMM model has been widely used for speech signal processing. As it handles temporal effect of visual speech, HMM model is suitable to understand speaker specific feature and it gives better results, therefore we have selected this model. Kaynak et al. [3] used the effect of the modeling parameters of HMM for geometrical lip features recognition. Wang et al. [4] used HMM with 6 states model and Regularized Discriminant Analysis (RDA) for classification. Chiou and Hwang [6] built integrated lip reading system based on snakes, Karunen loeve transform (KLT) and HMM as a classifier. Matthews et al. [7] compared different transform techniques for large vocabulary continuous speech recognition (LVSCR) using HMM classifier and they found that word error rate is more for LVSCR. Potamianos et al. [8] compared PCA, DWT and DCT transform techniques for digit recognition using HMM with 6 states and found that result of DCT is more accurate as compared to other techniques. Lee et al. [9] have used new training method of HMM parameters based on a variant of the Simulated Annealing (SA) algorithm to overcome the limitation of local optimization by the conventional method. Seymour used 10 states HMM and Puviarasan et al. have used HMM with 6 states of 33 words for hearing impaired persons [10,11].

Hassant and Jassim [5] while experimenting on only visual information concluded that different people generate different visual information. As described in their paper and shown in Table 1, some videos were easier to read than others (37–63%). Different people have different abilities to perceive speech from visual cues only. Performance of human lip reader varies from 23% to 73%. Moreover human lip readers benefit from visual cues like eye, muscle information of face detected outside mouth area. Authors have concluded that the recognition rate for person independent model is low. Lip reading still has scope for improvement and to generate significant set of features in recognition paradigm will be very useful.

The objective of this paper is to recognize different digits using visual features such as geometrical shape of the lip. Geometrical parameters such as change in area, height and width are used as a feature vector. The combined effect of these parameters is tested. Testing is performed on Indian English accent data base and Cuave database.

This paper is organized as follows. Section 2 describes LACM, geometrical feature extraction and HMM with 3 state and 5 state. Section 3 describes database, contour result, HMM model design. Section 4 explains geometrical model experimental results for both database and results of testing HMM model for different states. Section 5 discusses conclusions on work performed.

3. Proposed method

There are large inter and intra subject variations due to the variation in speed of utterance and this results difference in the number

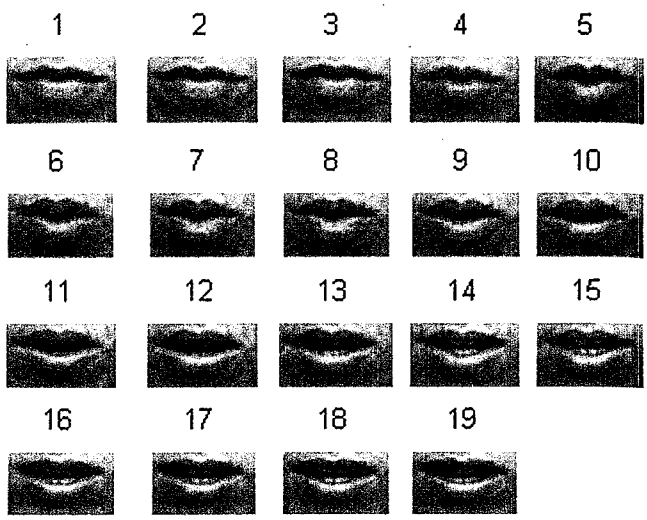

Fig. 1. Total number of frames for utterance of four.

of frames for each utterances. As a first step, we converted the color images into gray images and then mean squared difference (σ_i) between the consecutive frames are calculated to get active and prominent values by using Eq. (1). Based on the higher values of σ_i, ten significant frames are selected. The number of frames for each utterance is made same such that the feature vectors size is same for each utterance. I_i is ith frame, I_{i+1} is i+1 frame. Frame size is $M \times N$ pixels. σ_i is the mean squared difference between consecutive frames. i varies from 1 to size of frames required to utter a digit. Fig. 1 shows total 19 frames required for utterance of digit 4 and Fig. 2 shows 10 significant frames.

$$\sigma_i = \frac{1}{M \times N} \sum_{x=0}^{x=M-1} \sum_{y=0}^{y=N-1} [I_i(x,y) - I_{i+1}(x,y)]^2 \quad (1)$$

3.1. Localized active contour model

We have used localized active contour model (LACM), proposed by Lankton and Tannenbaum. In our previous work [12] comparison LACM with other active contour model for contour detection was done. It was found that LACM works better as compare to other model. LACM is a natural framework that allows any region-based segmentation energy to be re-formulated in a local way. This model is also useful for biomedical images. For biomedical images contrast is low as well as color variation is also insignificant. As similar characteristics are observed for lip and outside the lip portion, therefore we have used this method for lip tracking. It

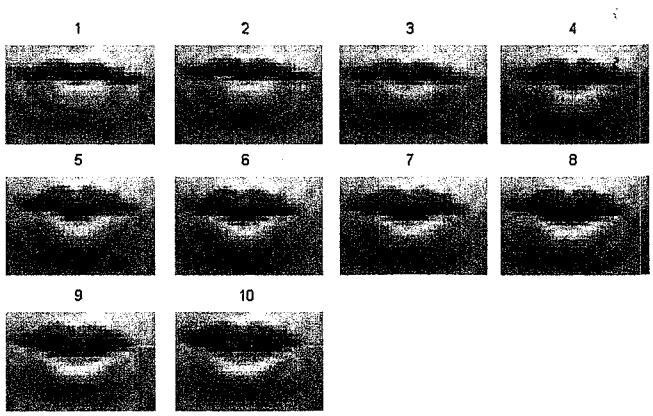

Fig. 2. Significant frames for utterance of four.

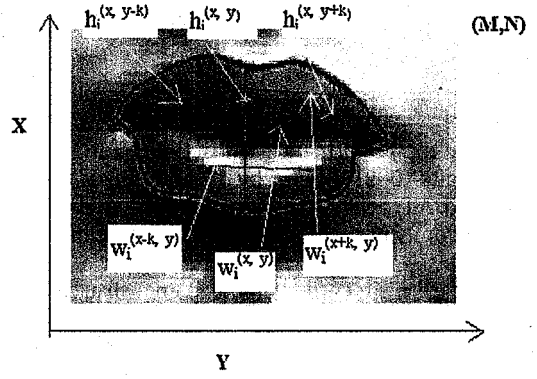

Fig. 3. Lip parameters w and h from contour C.

is also possible to detect contour accurately and to perform robust segmentation using LACM.

Lankton and Tannenbaum [13] considered local rather than global image statistics and evolved a contour based on local information. Localized contours are capable of segmenting objects with heterogeneous feature profiles that would be difficult to capture correctly using a standard global method. This method is used for lip tracking and equations used are as follows.

$$\frac{d\varnothing}{dt}(x) = \delta\varnothing(x)\int_{\Omega_y} D(x,y)\cdot\nabla(\varnothing(y))(F(I(y),\varnothing(y))dy + K(\varnothing(x)) \quad (2)$$

$$K(\varnothing(x)) = \propto \delta(\varnothing(x))\mathrm{div}\left(\frac{(\nabla(\varnothing(x)))}{(|\nabla(\varnothing(x))|)}\right) \quad (3)$$

$$\nabla_{\varnothing(y)}F = (\delta\varnothing(y)((I(y)-u_x)^2 - (I(y)-v_x)^2) \quad (4)$$

In Eq. (2) x and y are independent variable each representing single point in domain Ω of image. 'I' represent image from frames. F is generic internal energy measure used to represent local adherence. In the energy models the foreground and background as constant intensities represented by their means of u and v. $\delta(\varnothing)$ is a smooth version dirac delta. α is the weight of smoothing term. F is the energy function of uniform modeling energy. Energy function is given by Eq. (3). $D(x, y)$ is a mask local region whose value is one inside radius r and is zero outside r.

3.2. Geometric parameter extraction

The holistic approach i.e. visual digit is used instead vismeic approach to automatic lip reading. In the geometric feature approach we first appropriately normalize and rotate the outer lip contours, in order to compensate for relative location variations between the subject and camera. Rotation of contour is achieved by correctly aligning and registering first lip frame and then comparing next frames with registered frame. Geometric features are extracted from the contour C. Features, namely lip height (h), width (w) and area (a) are the most informative for speech-reading and LACM method is used to find contour. The parameters area, height and width are calculated as follows. Function $f(x, y)$ represent the lip image,

$$f(x,y) = \begin{cases} 1 & \text{if } (x,y)\in C_{\text{inside}} \\ 0 & \text{otherwise.} \end{cases}$$

where C is the lip contour, x and y are pixel position as shown in Fig. 3.

Then the three features of interest are defined as follows

$$w_i = \max_y \sum_x f(x,y) \quad (5)$$

$$h_i = \max_x \sum_y f(x,y) \quad (6)$$

$$a_i = \sum_x \sum_y f(x,y) \quad (7)$$

where w_i indicates lip width of current frame; h_i indicates lip height of current frame; $\max_y$ indicates y to produce maximum width; $\max_x$ indicates x to produce maximum height and a_i denotes area of lip contour.

Parameters change in area (Δa), change in height (Δh) and change in width (Δw) are calculated by taking difference of parameter between frames. Change in features are calculated as follows and given as input to HMM. Δw_i measures difference in width of contour of current frame and next frame. Δh_i measures difference in height of current frame and next frame. Δa_i measures difference in area of current frame and next frame. W, H and A represents matrix of dimension $1 \times n$. For our experimentation n is equal to 9.

$$\Delta w_i = w_{i+1} - w_i \quad (8)$$

$$\Delta h_i = h_{i+1} - h_i \quad (9)$$

$$\Delta a_i = a_{i+1} - a_i \quad (10)$$

$$W = [\Delta w_1, \Delta w_2, \ldots \Delta w_9]$$
$$H = [\Delta h_1, \Delta h_2, \ldots \Delta h_9] \quad (11)$$
$$A = [\Delta a_1, \Delta a_2, \ldots \Delta a_9]$$

3.3. Classification by HMM model

A classifier tool WEKA is used for this database with geometrical parameter. For this database, recognition rate of classifier K-Nearest Neighborhood (KNN), Support Vector Machine (SVM) and neural net with back propagation are not satisfactory. Hence we proposed Ergodic HMM as a classifier. For sequence $\{q_1, q_2, \ldots, q_n\}$: $P(q_n|q_{n-1}, q_{n-2}, \ldots q_1) = P(q_n|q_{n-1})$. This is called a first-order Markov assumption: we say that the probability of a certain observation at time n only depends on the observation q_{n-1} at time $n-1$. We express the probability of a certain sequence $\{q_1, q_2, q_3, \ldots, q_n\}$, the joint probability of certain past and current observations using the Markov assumption. Rabiner [14] tutorial is used for developing HMM.

$$P(q_1, \ldots, q_n) = \prod_{i=1}^{n} P(q_i)|q_{i-1} \quad (12)$$

The operation of HMM is characterized by (hidden) state sequence $Q = \{q_1, q_2, \ldots, q_N\}$. The observation sequence. $X = \{x_1, x_2 \ldots, x_n\}$ A HMM model is specified by the set of states $S = \{s_1, s_2, \ldots, s_N\}$, and set of parameters $\theta = \{\Pi, A, B\}$.

i) **Prior probabilities** $\Pi_i = P(q_1 = s_i)$ are the probabilities of s_i being the first state of sequence. Prior probabilities are collected in a vector Π. The prior probabilities can be assumed equiprobable and

$$\sum_j \Pi_j = 1 \quad (13)$$

ii) **Transition probabilities** are the probabilities to go from state i to j $a_{ij} = P(q_{n+1} = s_j \,|q_n = s_i)$. They are collected in the matrix **A**. A transition probability matrix is as below A: $s \times s \rightarrow R$ is a state transition probability function where R is a real number set.

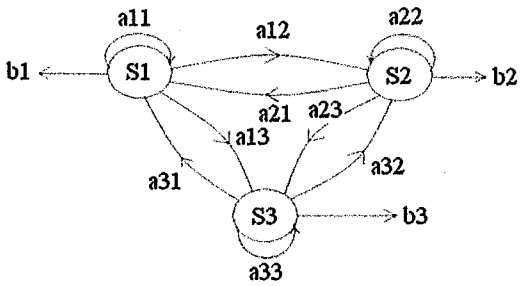

Fig. 4. State transition diagram for height.

The N is number of states. The three state transition matrix A is given by Eq. (14)

$$a_{ij} \geq 0$$

$$\sum_{j=1}^{N} a_{ij} = 1$$

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{14}$$

iii) **Emission probabilities** characterize the likelihood of a certain observation x, if the model is in state s_i. Depending on the kind of observation x, B represent emission matrix or observation matrix. M is number of observation. Hidden markov model is used to represent these states. Ergodic HMM state diagram is shown in Fig. 4.

$$\sum_{k=1}^{M} b_{ij}(k) = 1$$

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} & \dots & b_{19} \\ b_{21} & b_{22} & b_{23} & \dots & b_{29} \\ b_{31} & b_{32} & b_{33} & \dots & b_{39} \end{bmatrix} \tag{15}$$

Eq. (18) is used to calculate most likelihood probabilities. Then minimum value of log likelihood probabilities is calculated to recognize the output of HMM. These HMM results of width, height and area are combined. In testing phase P Probability of state sequence through HMM. $q(t)$ is test state sequence of data at time t. T is a length of test sequence.

$$P = \Pi(q(1))B(q(1), 1) \tag{16}$$

$$Q = \prod_{2}^{T} A((q(t-1), q(t))B(q(t), t)) \tag{17}$$

$$L = \log(PQ) \tag{18}$$

Eqs. (19) and (20) L_t is total likelihood probability and its minimum value indicates recognized digit. L_a, L_w and L_h are likelihood probabilities of area, width and height respectively.

$$L_t = L_a + L_w + L_h \tag{19}$$

$$L_t = 0.4L_a + 0.2L_w + 0.4L_h \tag{20}$$

Eq. (20) is empirically derived from results of Table 5 of individual parameter.

4. Experimentation

4.1. Database

4.1.1. In-house database

To evaluate the performance of feature vector a database consists of 10 digits (0–9) uttered by 16 individuals (8 male and 8 female) is used. All speakers were asked to repeat the digit sequentially and randomly under natural lighting condition. Frame size is 720×480 pixels.

Video Data base is recorded in Indian English accent. Recording distance is kept constant. No head movement is allowed. Background used is blue. Video was recorded for digit 0–9 number. Data base having frame rate of 25frames/s and audio sampling frequency 48,000 is used. For each person six videos are recorded in sequence of number 0–9 digit and random manner. Videos are recorded in normal light. Volunteers are selected from the age group of 22–25 years.

4.1.2. Cuave database

Cuave database is standard database and its video frame rate is 30frames/s. It contains mixture of white and black features. Database digits are continuous and with pause also. Data is recorded with sequential and random manner. Some videos are taken from side view. Total 36 videos are in data base, out of which, 19 are male and 17 are female. Disruptive mistakes were removed, but occasional vocalized pauses and mistakes in speech were kept for realistic test purposes. The data was then compressed into individual MPEG-2 files for each speaker and group. It has been shown that this does not significantly affect the collection of visual features for lip reading [15].

Each individual speaker was asked to move side-to-side, back-and-forth, or tilt the head while speaking 30 isolated digits. In addition to this isolated section, there is also a connected-digit section with movement as well. So far, much research has been limited to low resolution, pre-segmented video of only the lip region.

4.2. Video data separation

From audio analysis, numbers of frames are decided. Pratt software is used for audio analysis which is used to separate important video frames of a digit. On an average 16 frames are sufficient for a digit. Frames are captured before 0.12 s from starting of each digit. Out of 16 frames we have used 10 significant numbers of frames by using Eq. (1). Only lip portion is used for obtaining contour so that the time required to capture contour is reduced. Lip portion is cropped with size of 64×40 manually. Color lip portion is converted into gray scale and then LACM is applied. Complexity of computation is reduced by using less number of frames.

4.3. LACM

Results of contour depend on number of iteration, localization radius (in pixels) and alpha weight of smoothing term. For radius equal to 3 and 4, results are better. It is considered that smaller the radius, more the local energy and vice versa. Higher the value of alpha smoother is contour. From the contour different geometrical features are extracted.

Localization radius of 3 and 4 is used for finding contour. Computation is divided into two parts. First for radius equal to 3 and iterations are 200 then for radius equal to 4, 150 iterations are used. Alpha (α) is 0.5. Contour information is stored in matrix. For Cuave database localization radius of 2 and 3 is used for finding contour. Figs. 5 and 6 are the examples of lip contour extraction of outer portion of lip of in-house database and Cuave database.

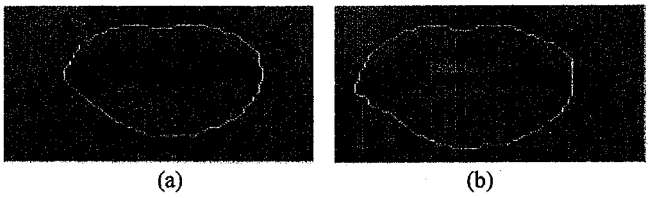

Fig. 5. Lip contour of utterance (a) zero and (b) three for in-house database.

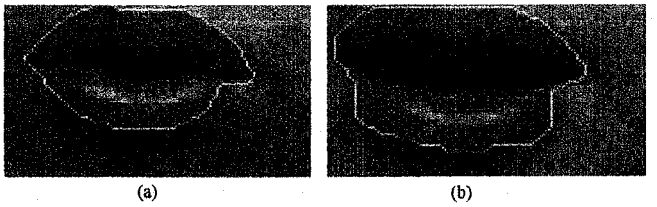

Fig. 6. Lip contour of utterance (a) four and (b) nine for Cuave database.

4.4. Training of HMM for digit 0–9

For most of the cases, left to right HMMs are used for visual speech recognition. We have used Ergodic HMM. Only some of the frames having visual speech information, so selected frames are used. For example rounding shape, wide opening, small opening of mouth are the characteristic of some digit. This is similar to viseme. Particular viseme plays a role in digit recognition. Last frames of lip shape changes are affected due to continuous utterance of digit and will also depend upon next utterance.

Lip height, width, area are calculated from contour. Three states are considered. Change in height of lip is increasing (state 2), decreasing (state 1) or unchanged (state 3). These changes in states are calculated for width and area also. In five state model, threshold is 0.12 and 0.4, so two more states are added. This indicates amount of change in lip movement. The change in states is calculated for width, area and height. Hidden markov model is used to represent 5 state model. Individual model is used for height, width and area. HMM model for each digit is as shown in Fig. 7. In classification, result of each HMM is compared and maximum probability HMM digit is selected as result.

5. Experimental results

Cross validation technique with 10 fold (90% data for training and 10% data for testing) is used for experiment. Figs. 8 and 9 indicate feature vectors A is matrix of change in area, H is matrix change in vertical position of lip and W is matrix of change in horizontal position of lip. Bar graph depicts the performance of individual parameter as well as combines parameter. Bar graph in Figs. 8 and 9 show that result of recognition rate of vertical movement of lip gives more information as compare to horizontal movement. Therefore the results obtained using Cuave database are

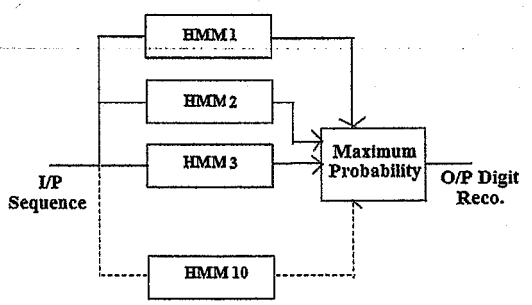

Fig. 7. Ten different HMMs used for 10 digit recognition.

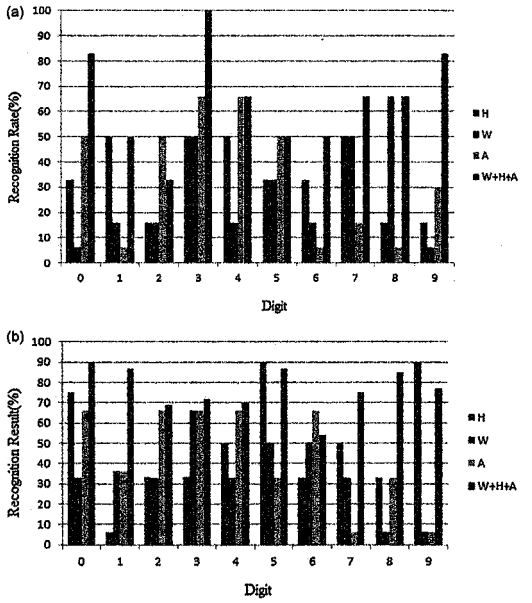

Fig. 8. (a) Recognition rate (%) of each digit with different parameters in-house database (3 states). (b) Recognition rate (%) of each digit with different parameters in-house database (5 states).

comparatively better than in-house database as in Cuave database for many digit utterances mouth opening in vertical direction is more frequent. Change in vertical position of lip with respect to time is important parameter. Performance of the output increases from 1% to 2% by using Eq. (20).

For different digits, recognition rate with combined parameters are as shown in Table 2. For Machine learning process this digit recognition rate is important. Zero, one, five and eight are found to be most recognized digits for the in-house database. These digits can be used for machine learning and the combination of these digits can be used for important programming commands of CNC machines. In-house database maximum recognized digit is zero and least recognized digit is six.

Table 3 shows the performance of three feature vectors (*A*, *W* and *H*) and comparison of recognition rate for different number of

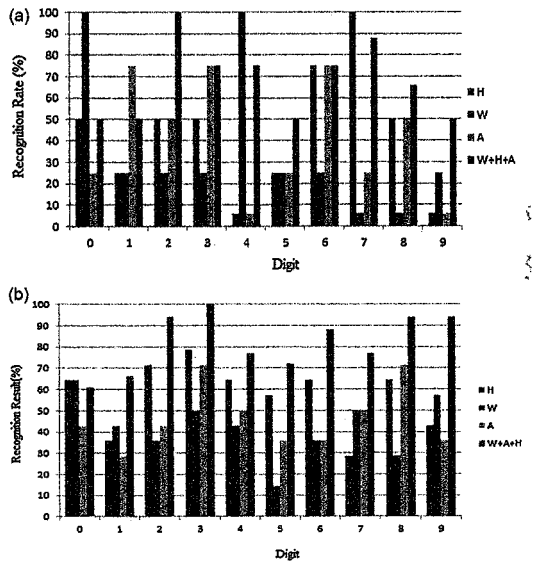

Fig. 9. (a) Recognition rate (%) of each digit with different parameters Cuave database (3 states). (b) Recognition rate (%) of each digit with different parameters in Cuave database (5 states).

Table 2
Recognition rate (R.R) for different digits with 5 states HMM in-house database.

Database	States	Average recognition rate (%) for 0–9 digits									
		0	1	2	3	4	5	6	7	8	9
In-house	3	83	50	33	100	66	50	50	66	66	83
	5	90	87	69	72	70	87	54	75	85	77
Cuave	3	61	66	100	75	75	50	75	88	66	50
	5	61	66	94	100	77	72	88	77	94	94

Table 3
Comparison of R.R. individual feature and combining features for own database and Cuave database using 3 and 5 states HMM model.

Database	Feature set	3 state HMM R.R. (%)	5 state HMM R.R. (%)
Cuave	H	44	58
	W	33	37
	A	40	45
	$W+H+A$	66.3	78.33
In-house	H	35	49
	W	26	33
	A	33	43
	$W+H+A$	64.7	76.6

Table 4
Recognition rate (%) of male and female candidate for in-house database.

Male candidate	Reco. rate %	Female candidate	Reco. rate (%)
1	79.3	1	69.6
2	80.3	2	80
3	70	3	71.3
4	70.6	4	80
5	69.3	5	81
6	82	6	89.3
Average	75.25	Average	78.33

Table 5
Recognition rate (%) of male and female candidate for Cuave database.

Male candidate	Reco. rate %	Female candidate	Reco. rate %
1	80	1	82.5
2	82.3	2	84
3	84	3	70
4	79.6	4	68
5	68	5	86.7
6	73	6	86.7
Average	77.8	Average	79.6

states with HMM for both database. Recognition rate is increased by using more training data and increasing number of states. Combination of three parameter result gives $(A+W+H)$ better recognition rate instead of individual parameter. Performance of HMM model with Cuave database is better as compare to In-house database. Tables 4 and 5 show percentage recognition rate of female candidate is more as compare to male candidate for both database.

6. Conclusion

In this paper, we have proposed and verified the usefulness of LACM for contour extraction which is used to calculate different geometrical parameters such as area, height and width of the lip during lip reading process. Changes in area, height and width of lip are used as a feature vector, so that dynamic information is captured. Individual features are compared and it is found that changes in vertical direction of lip have significant impact on recognition rate. The combination of these parameters is important as it drastically improves recognition result. Experimental results demonstrated that the proposed model of HMM with 3 states provides acceptable recognition result with less complexity. Five states HMM recognition rate is more as compared to three states HMM. It is also found that recognition results of experiments performed on Cuave database are much better than that of in-house database. Zero, one, five and eight are most recognized digits from the experiment performed on in-house database, therefore they can be used for most important commands used for programing of CNC machine.

References

[1] H. McGurk, J. MacDonald, Hearing lips and seeing voices, Nature 264 (1976) 746–748.
[2] E. Petajan, B. Bischoff, D. Bodoff, An improved automatic lip reading system to enhance speech recognition, CHI' 88 (1988) 19–23.
[3] M.N. Kaynak, Q. Zhi, A.D. Cheok, K. Sengupta, Z. Jian, K. Chi Chung, Lip geometric features for human–computer interaction using bimodal speech recognition: comparison and analysis, J. Speech Commun. 43 (2004) 1–16.
[4] S.L. Wang, A.W.C. Liew, W.H. Lau, S.H. Leung, An automatic lip reading system for spoken digits with limited training data, IEEE Trans. Circuits Syst. Video Technol. 18 (2008) 1760–1764.
[5] A.B. Hassant, S. Jassim, Visual words for lip reading, in: Proc. SPIE 7708, Mobile Multimedia/Image Processing, Security and Applications, 2010, p. 7708.
[6] G. Chiou, J.-N. Hwang, Lipreading from color video, IEEE Trans. Image Process. 6 (1997) 1192–1195.
[7] I. Matthews, T. Cootes, J. Bangham, Extraction of visual features for lipreading, in: IEEE Trans. on Pattern Analysis and Machine Vision, 2002, pp. 198–213.
[8] G. Potamianos, H. Graf, E. Cosatto, An image transform approach for HMM based automatic lip reading, in: International Conference on Image Processing, 1998, pp. 173–177.
[9] J. Lee, C. Park, Training hidden markev model by hybrid simulated annealing for visual speech recognition, in: IEEE Int. Conference Systems Man and Cybernetics, 2006, pp. 8–11.
[10] R. Seymour, D. Stewart, J. Ming, Comparison of image transform-based features for visual speech recognition in clean and corrupted videos, EURASIP J. Video Process. 2008 (2008) 1–9.
[11] N. Puviarasan, S. Palanivel, Lip reading of hearing impaired persons using HMM, J. Exp. Syst. Appl. 38 (2010) 1–5.
[12] S. Morade, S. Patnaik, Automatic lip tracking and extraction of lip geometric features for lip reading, IJMLC 2013 (3) (2013) 168–171.
[13] S. Lankton, A. Tanenbaum, Localizing region based active contours, IEEE Trans. Image Process. 17 (2008) 2029–2039.
[14] L. Rabiner, A tutorial on hidden markov models and selected applications in speech recognition, Proc. IEEE 77 (1989) 257–286.
[15] E. Patterson, S. Gurbuz, Z. Tufekci, J. Gowdy, CUAVE: a new audio-visual database for multimodal human computer-interface research, Proc. IEEE Int. Conf. Acoust. Speech Signal Process. 2 (2002) 2017–2020.