Journal of Computer Science 6 (10): 1065-1069, 2010
ISSN 1549-3636
© 2010 Science Publications

Classical Flexible Lip Model Based Relative Weight Finder for Better Lip Reading Utilizing Multi Aspect Lip Geometry

[1]B. Sujatha and [2]T. Santhanam
[1]Department of Computer Science, Meenakshi College for Women, Chennai, India
[2]Department of Computer Science, DG Vaishnav College, Chennai, India

Abstract: Problem statement: Deaf and dumb needs assistance from a technical box that takes movements of lips to identify the words. This technical article provided appropriate model implementation of flexible lip model for better visual lip reading system. **Approach:** From the frame sequence of words, Active Shape Model (ASM) based lip model provided local tracking and extraction of geometric lip-feature. Two geometric criteria define required geometric features and its variations in the sequence. **Results:** The feature established machine classification using Analytic Hierarchy Process (AHP), a relative weight finder. AHP presents weight vector to fuzzy classifier to decide the video frame sequence belonging to a respective word. **Conclusion:** The suggested model tested on a total of 5 different sample databases results in 83.2% accuracy over the other combinational algorithms.

Key words: Active shape model, analytic hierarchy process, fuzzy classifier

INTRODUCTION

Identification of a word is from a video frame-sequence with the help of a statistical classifier using a snapshot database has been attempted in this study. Compared to the requirements of speech signal, environmental noise problems are not difficult to deal visually. This feature reduces the burden of pre-processing and makes the model to run fast. In a sequence of frames, lip tracking and segmentation can be done in every frame using traditional image based techniques. Here, lip-tracking becomes stand alone issue. In addition, any image based technique makes it difficult to find geometrics in glossy environment. Regarding local tracking and segmentation of lips, model based techniques called active shape model solves both issues in single step by minimal-set features effectively, which are consistent over frame constraints (Mattews *et al*., 2002).

ASM is used in lipreading but the feature extraction method supported in this study is different from the technique proposed in (Faruquie *et al*., 2000; Mok *et al*., 2004; Sum *et al*., 2001). ASM is widely used in facial recognition and most of the researchers use this model only to extract the lip region and branch towards other areas, not focusing on the lip reading. Geometrics of outer and inner lips are found from ASM with optimal-point lip model. Inner lip geometry is a sub-set to provide additional information. Usual training and testing of classifiers like HMM and other network models become time consuming and more erratic when database size increases with more words. These classifiers are loaded heavily in sequence comparison of individual frame features belonging to a word, while training it. By introducing AHP, a relative weight finder wherein the relationship of each feature to a shape has been characterized on a numerical scale and its weight is defined, to minimize erratic decisions taken by classifier. Once the burden is shared by AHP, simple fuzzy classifier makes decision about a particular word, belonging to the test frame-sequence.

The objective of this study is to deploy the above techniques in lip reading to classify the words from ‘one’ to ‘nine’, in a customized database created for this purpose. The organization of this technical article is as follows: Under active shape model defines local tracking and segmentation lip images. Feature selection defines about feature extraction based on length and area information. The relative weight finder includes: the subsection AHP classifier algorithm and the next subsection as fuzzy decision maker. Finally the results and discussion followed by the conclusion.

MATERIALS AND METHODS

Active shape model: ASM is a shape constrained iterative fitting algorithm (Koschan *et al*., 2003; Cootes, 2000; Cootes *et al*., 1993; 1994; Fieguth and

Corresponding Author: T. Santhanam, Department of Computer Science, DG Vaishnav College, Chennai, India
Tel: +91-94441690901

Terzopoulos, 1997) where the Point Distribution Model (PDM) called statistical shape model, characterizes the shape constraint. ASM represents the shape as:

$$A = [(X_{1,}Y_1),....(X_{16},Y_{16});(a_1,b_1)....(a_4,b_4)] \quad (1)$$

Where:

X_i, Y_i ($1\leq i \geq 16$) = The coordinate of the ith Point of outer lip contour

a_j, b_j ($1\leq j \geq 4$) = The coordinate of the jth Point of inner lip contour

Initial model points x_k in the shape represented by the function:

$$\text{Transformation of } (X' + \lambda P) \text{ by} [x_k, y_k, s, \theta] = X_k \quad (2)$$

When initial shape obtained from training of it based on sample of shape of a person, then tracking and deformation of the shape to frame lip shape does not need multi-resolution strategy (Cootes *et al*., 1993), further it achieves fast convergence. Figure 1a is the histogram equalized gray scale test frame for the word 'seven'. ASM segmentation of that lip image is displayed in Fig. 1b from a video frame of that word. Figure 1c represents outer and inner lip contours meanwhile to realize the dimensions.

They represent texture model with edge constraints to have better lip contour introduced with 16 point references. Inner lip contour gets only 4 point references with reduced effort on it. These point model is compact representation of lip shape and it can be obtained from sample lip image of a person belongs to a particular database called training image.

Use the Constraint-Based Model m Matching technique for image search, to match frame lip shape. Gaussian satisfied, normalized profile defines guidelines for gradient matching. The degree of reaching frame lip shape from sample is given by:

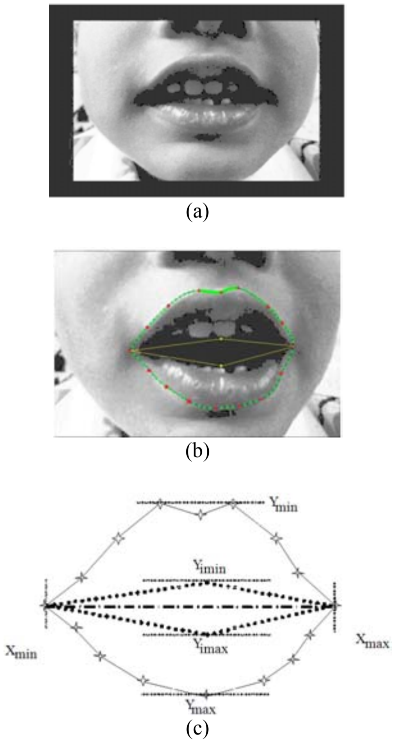
$$f(m_s) = (m_s - m')^T S_m^{-1} (m_s\text{-}m') \quad (3)$$

Where:

m' = The mean of normalized first derivative of gradient profile

$\{m_i\}$ ($1\leq i \geq 16$) and S_m = The covariance

Convergence minimizes f (m_s) and its value in the probability of m_s in the distribution. Strong edge of lip shape is identified by least probability of $m_{s,}$ this may be the best choice but could not be the optimal choice.


(a)

(b)

(c)

Fig. 1: (a) histogram equalized Gray scale test frame. (b) ASM segmentation of lip from a video frame of the word "seven". (c) It represents outer and inner lip contours to realize the dimensions

Edge constraint keeps all the 16 points to lie on strong gradient edge of outer lip. Improved mahalanobis distance function with Sobel edge intensity can increase the probability of finding strong edge of lip shape. For the Inner lip features, finding 4 point inner lip shape is an integrated one in these sample models.

Feature selection: Selection of features, belongs to a two-dimensional lip shape with outer and inner contour, is based on length and area-based criteria. The length based feature is being further classified into outer lip length-width ratio, vertical up-distance ratio and inner lip length-width ratio.

The Outer Lip Length-Width Ratio (OLWR) can be obtained by the expression:

$$OLWR = (X_{max} - X_{min})/(Y_{max} - Y_{min}) \quad (4)$$

pixel positions are key points to calculate OLWR. In both X_{min} and X_{max}, change in row positions will never affect conceptual analysis, so that the OLWR is calculated as 1.3065. When the pronunciation of the word happens, wide variation happens in OLWR. It becomes vital feature in feature domain.

The Vertical Lip Distance Ratio (VLDR) is the second length based feature. From the centre of $X_{min} – X_{max}$ line, the centre pixel position of the line, the distance to the top edge of outer upper lip contour is called L_1, is of 133 pixels and the distance to the bottom edge of outer lower lip contour is called L_2, is of 129 pixels. The ratio between L_1 and L_2 is called vertical lip distance ratio and its value is 1.0310. This geometry is one of the length based representation of openness.

Third length based feature is Inner Lip Width-Length Ratio (IWLR). IWLR is similar to outer lip length-width ratio.

$$IWLR = (Y_{imax} – Y_{imin})/(X_{max} – X_{min}) \quad (5)$$

Where:
IWLR = 0.1731
Y_{imax} = Inner lip contour's bottom edge
Y_{imin} = Inner lip contour's top edge

The area based feature constitutes the lip-openness area ratio. This requires the area occupied by outer lip contour as a whole (A_1) and the area occupied by inner lip contour (A_2), are obtained by applying a seed-filling algorithm. Lip-Openness Area Ratio (LOAR) is the first feature under area-based features that needs both A_1 and A_2.

$$LOAR = A_2/A_1 = 0.1676 \quad (6)$$

LOAR becomes pure indication of openness. LOAR accounts inner lip length and width indirectly.

By analyzing these features proportional decremented changes in length with incremented change in width may keep A_2 as approximately constant but ILWR increases rapidly. These two parameters play a major role to decide different kind of openness. Feature matrix of considered frame is:

$$[1.3065\ 1.0310\ 0.1731\ 0.1676]$$

represents OLWR, VLDR, IWLR and LOAR features in the sequence.

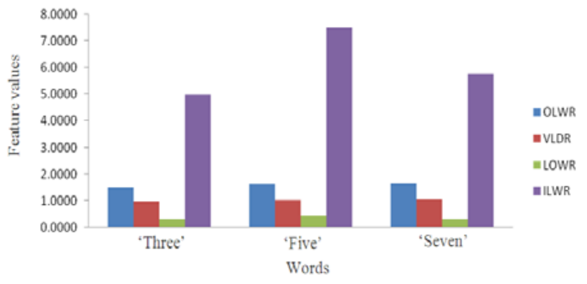


Fig. 2: Comparison of all four features for words Three', 'Five' and 'Seven'

Table 1, lists all the essential attribute values of normalized frame 1, 6, 11 and 16 sequence belongs to the words 'Three', 'Five' and 'Seven'. Then, the comparison of individual features for three different words is shown in Fig. 2. From Fig. 2 the ILWR for all these words are more and among this the ILWR for 'Five' is more than 'Three' and 'Seven'.

Relative weight finder: When a database is applied to feature extraction after both normalization and selection of p number of test frames, its feature matrices are stored. Input video frames are also applied to the normalization of requirement and selection of same p number of test frames. Finally input video's feature matrix is obtained. The four features of lip shape, is used to find individual feature matrices as follows:

$$B = [f_1 \quad f_2 \quad \dots \quad \dots \quad f_p] \quad (7)$$

Table 1: Normalized feature set for the words 'three', 'five' and 'seven'

Words	Frame Name	OLWR	VLDR	LOWR	ILWR
'Three'	frame3_01	1.4246	0.9385	0.3379	3.7789
	frame3_06	1.7073	1.0297	0.4314	4.1667
	frame3_11	1.9492	1.1071	0.5100	5.7500
	frame3_16	1.5000	0.9832	0.3249	4.9859
'Five'	frame5_01	1.3745	0.9328	0.2618	5.0141
	frame5_06	1.6875	1.0392	0.4596	9.7500
	frame5_11	1.6193	1.0000	0.2733	7.3542
	frame5_16	1.6296	1.0187	0.4424	7.4894
'Seven'	frame7_01	1.6355	1.0189	0.2904	38.8889
	frame7_06	2.3267	1.3438	0.4157	18.3684
	frame7_11	1.6509	1.0190	0.5178	7.4468
	frame7_16	1.6449	1.0381	0.3227	5.7705

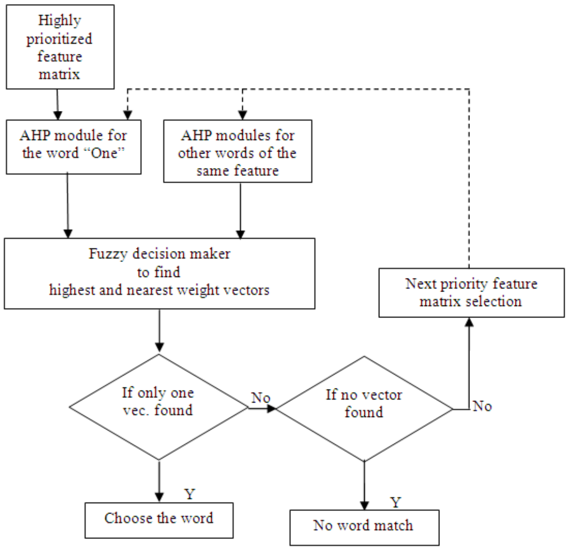

Highly prioritized feature matrix
AHP module for the word "One"
AHP modules for other words of the same feature
Fuzzy decision maker to find highest and nearest weight vectors
Next priority feature matrix selection
If only one vec. found
No
If no vector found
No
Y
Choose the word
Y
No word match

Fig. 3: Hierarchy structure of classifier

where, f is a particular feature collected from p number of test frames of a word. The individual feature matrices are sequenced according to the priorities decided by ourselves based on geometric importance. AHP's structure is decided as shown in Fig. 3.

AHP classifier:

Step 1: Calculating relative weights: The relative weights in the feature level are expressed as pair-wise comparisons (Saaty, 1980; 2008) as the ratios of relative importance between pairs of same feature from database and test input. The weight of every feature is represented as an integer between 1 and 9 and the relative weights of features are expressed as a ratio. These values define only the weight:

$$w = (w_1, w_2 \dots w_p) \tag{8}$$

It is therefore always possible to normalize the vector w by imposing:

$$\sum_i w_i = 1 \tag{9}$$

The ratio w_i / w_j would then be the relative weight of the ith to the jth element. By combining all possible pair-wise relative weights into a preference matrix C as follows:

$$C = \begin{bmatrix} w1/w1 & w1/w2 & w1/w3 & \dots & w1/wp \\ & & & & . \\ w2/w1 & \dots & \dots & & . \\ & & & & . \\ wp/w1 & \dots & \dots & & wp/wp \end{bmatrix} \tag{10}$$

The elements of matrix A have the special property $w_i / w_j = a_{ij} = 1/a_{ij} = 1/(w_j/w_i)$ for all i and j.

Step 2: calculating global weights: The resultant vector with a heavier global weight is the module's classification decision. The ten AHP modules of pronunciation of numeric's gave us ten global weights. AHP normalizes the weight of the ith row as follows:

$$W_i = \frac{1}{p}\sum_{j=1}^{p}\frac{a_{ij}}{\sum_{k=1}^{p} a_{kj}} \tag{11}$$

Let $w^{(k-1)} = (w_1{}^{\{k-}$ and then weight is calculated from top to ${}^{1)}$, $w_2{}^{\{k-1)}$,......, $w_p{}^{\{k-1)}$) be the weight for the given number of elements in the (k-1)th level and $q_j{}^{(k)} = (q_{1j}{}^{(k)}, q_{2j}{}^{(k)}, \dots\dots, q_{pj}{}^{(k)})^T$ is the weight for the next level of p elements in the kth level bottom as:

$$w_i^{(k)} = \sum_{j=1}^{m} q_{ij}^{(k)} w_j^{(k-1)},\ i = 1,2,3,\dots\dots,p \tag{12}$$

This weight is called as global weight of every AHP module. Consistency of the preference matrix (Forman and Peniwati, 1998; Barzilai and Golany, 1990; Peniwati, 1996) is verified by consistency Index.

Fuzzy decision maker: The study is focused on classification and not recognition. Although AHP converges well in terms of weight vector, each AHP module can only distinguish one word's lip movement from other word's lip movement. Secondly, as the lip movement happens, it need not belong definitively to one word or another in the database.

The fuzzy membership theory (Mikhailov and Tsvetinov, 2004; Pedrycz and Gomide, 1998) is premised on the observation that many phenomenal lip movements cannot be discreetly categorized as members of one word or another, but rather, share features with other word so that they may be said to belong to one or more word and only to some degree. These factors are imposing the necessity of using fuzzy final decision maker.

RESULTS AND DISCUSSION

Well identified feature and its variations define better classification process. OLWR is identified as primary feature and its values of the word "FIVE" and "seven" are compared as given in Fig. 4 from Table 1 values. With reference to the comparison given, it is proved that feature definition is more vital in lip reading process.

A multi-feature lip model derivation increases the consistency as well as accuracy. A database consisting of five different speakers with all 10 digit video frame sequence is created. The image used for the experiment is of person-3 created by recording Audio Video Interface (AVI) files and it is converted to frames for further processing.

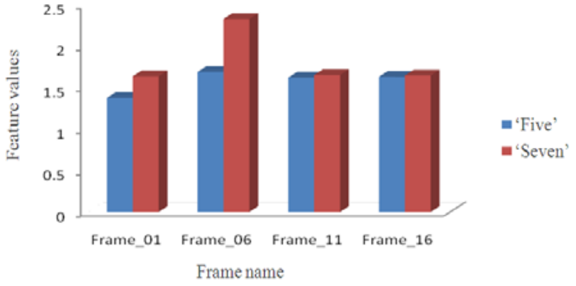


Fig. 4: Comparison of OLWR feature values ('Five' and 'Seven')

CONCLUSION

The environment from where the inputs were taken is not optimized. This improves applicable nature of our study. But the distance between speaker and video camera is to be analyzed in better way. Combination of ASM and AHP with Fuzzy decision maker increases the identification level and resulted in 83.2% accuracy as compared with (Sujatha and Santhanam, 2010) the image based and shape based approaches that yielded 66.8% accuracy, proving its supremacy.

REFERENCES

Barzilai, J. and B. Golany, 1990. Deriving weights from pairwise comparison matrices: The additive case. Operat. Res. Lett., 9: 407-410. DOI: 10.1016/0167-6377(90)90062-A

Cootes, T., J. Taylor and T.F. Cootes, 1993. Active shape model search using local grey-level models: A quantitative evaluation. Proceeding of the 4th British Machine Vision Conference, (BMVC'93), BMVA Press, pp: 639-648.

Cootes, T.E., C.J. Taylor and A.L. Anitis, 1994. Active shape models: Evaluation of a multi-resolution method for improving image search. Proceeding of the British Machine Vision Conference, (BMVC'94), BMVA Press, UK., pp: 327-336.

Cootes, T., 2000. An Introduction to Active Shape Models, Model-Based Methods in Analysis of Biomedical Images. In: Image Processing and Analysis: A Practical Approach, Baldock, R. and J. Graham (Eds.). Oxford University Press, Oxford, pp: 223-248.

Faruquie, T.A., A. Majumdar, N. Rajput and L.V. Subramaniam, 2000. Large vocabulary audio-visual speech recognition using active shape models. Proceeding of the 15th IEEE International Conference on Pattern Recognition, Sept. 3-8, IEEE Computer Society, Washington DC., USA., pp: 3110-3113. DOI: 10.1109/ICPR.2000.903496

Fieguth, P. and D. Terzopoulos, 1997. Color-based tracking of heads and other mobile objects at video frame rates. Proceeding of the International Conference on Computer Vision and Pattern Recognition, July 17-19, IEEE Computer Society, Washington DC., USA., pp: 21-27.

Forman, E. and K. Peniwati, 1998. Aggregating individual judgments and priorities with the analytic hierarchy process. Eur. J. Operat. Res., 108: 165-169. DOI: 10.1016/S0377-2217(97)00244-0

Koschan, A., S. Kang, J. Paik, B. Abidi and M. Abidi, 2003.Color active shape models for tracking non-rigid objects. Patt. Recog. Lett., 24: 1751-1765. DOI: 10.1016/S0167-8655(02)00330-6

Mattews, I., T.F. Cootes, J.A. Bangham, S. Cox and R. Harvey, 2002. Extraction of visual features for lip reading. IEEE Trans. Patt. Anal. Mach. Intell., 24: 198-213. DOI: 10.1109/34.982900

Mikhailov, L. and P. Tsvetinov, 2004, Evaluation of services using a fuzzy analytic hierarchy process. Applied Soft Comput., 5: 23-33. DOI: 10.1016/j.asoc.2004.04.001

Mok, L.L., W.H. Lau, S.H. Leung, S.L. Wang and H. Yan, 2004. Lip features selection with application to person authentication. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Proceedings, May 21-21, IEEE XPlore Press, China, pp: 397-400.

Peniwati, K., 1996. The analytic hierarchy process: The possibility theorem for group decision making. Proceeding of the 4th International Symposium on the Analytic Hierarchy Process, July 12-15, Vancouver, Canada, pp: 202-214.

Pedrycz, W. and F. Gomide, 1998. Introduction to Fuzzy Sets: Analysis and Design. Illustrated Edn., MIT Press, Cambridge, MA., ISBN: 10: 0262161710, pp: 465.

Saaty, T.L., 1980. The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation. 1st Edn., McGraw-Hill, New York, ISBN: 13: 9780070543713, pp: 437.

Saaty, T.L., 2008. Relative measurement and its generalization in decision making why pair-wise comparisons are central in mathematics for the measurement of intangible factors the analytic hierarchy/network process. Rev. R. Acad. Sci. Ser. A. Math., 102: 251-318.

Sujatha, B. and T. Santhanam, 2010. A novel approach integrating geometric and Gabor wavelet approaches to improvise visual lip-reading. Int. J. Soft Comput., 5: 13-18. DOI: 10.3923/ijscomp.2010.13.18

Sum, K.L., W.H. Lau, S.H. Leung, A.W.C Liew and K.W. Tse, 2001. A new optimization procedure for extracting the point based lip contour using active shape model. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, May 7-11, IEEE Computer Society, Washington DC., USA., 3: 1485-1488. DOI: 10.1109/ICASSP.2001.941212