Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 9, September 2014, pg.841 – 843

RESEARCH ARTICLE

Lip-Reading using Neural Networks

Priti Yadav[1], Priyanka Yadav[2], Vishal Sharma[3]

yadav.preeti44@yahoo.com, yadav.priyanka2601@gmail.com, 02vishalsharma@gmail.com

[1]Dronacharya College of Engineering (Computer Science Department)

[2]Dronacharya College of Engineering (Computer Science Department)

[3]Dronacharya College of Engineering (Computer Science Department)

Abstract

Lip-Reading has been practiced over centuries for teaching deaf and dumb to speak and communicate effectively with the other people. In this study, the use of neural networks in lip reading is explored. We convert the video of the subject speaking different words into images and then images are further selected manually for processing. As per the research the horizontal and the vertical distance between the lips varies for each and every word considering the close proximities of similar sounding words.

Based on this research we can create the database of commonly used words and our neural network model can form clusters of words based on its intelligent approach. This approach can be associated with various voice recognition softwares and help in increasing their efficiency readily even in a noisy environment, and creating new dimensions for human computer interaction.

I. INTRODUCTION

Lip-Reading is an area which till date is only practiced by elementary student-teacher method, but in this approach lip reading has been implemented through e-learning or virtual learning technique which is a difficult and a challenging problem. Also, in a noisy environment where voice recognition softwares tend to underperform, using this approach increases the efficiency remarkably.

Despite the difficulty associated with the unpredictable nature of problem domain, several researchers have attempted to develop models for the same problem but with different approaches. Most of them have tried to compare the two images by classic image comparison techniques. However the results obtained were of lower efficiency considering the efficiency downgrading constraints of image comparison technique used in the approach. In our study, we explore the use of neural networks in lip reading techniques. We convert this problem into a clustering one, where we calculate the four different distances: horizontal and vertical viz. H-1 (horizontal distance between inner lip points), H-2(horizontal distance between outer lip points), V-1 (vertical distance between inner lip points) and V-2 (vertical distance between outer lip points) and cluster them. The remainder of this paper is organized as follows. Section 2 briefly reviews the basics of ANN and also presents the comparison between the statistical techniques and our ANN approach. Section 3 gives the details of our methodology by specifically talking about the data, the neural network model, the experiment methodology and the performance measures used in this study. Finally, the Section 4 of the paper discusses the overall contribution of this study, along with its limitations and further research directions. Lip reading involves the extraction of visual speech features. The most visual speech information is contained in the inner and outer lip contour, it has also been shown that information about the visibility of teeth and tongue provide important speech cues. Particularly for fricatives, the place of articulation can often be determined visually, i.e. for labiodentals (upper teeth on lower lip), interdentally (tongue behind front teeth) and alveolar (tongue touching gum ridge) place. Other speech information might be contained in the protrusion and wrinkling of lips.

Lip reading approaches can be classified into:

Image-based systems.

Model-based systems.

Image-based systems use grey level information from an image region containing the lips either directly or after some processing as speech features. Most image information is therefore retained, but it is left to the recognition system to discriminate speech information from linguistic variability and illumination variability.

Model-based systems usually represent the lips by geometric measures, like the height or width of the outer or inner lip boundary or by a parametric contour model which represents the lip boundaries. The extracted features are of low dimension and invariant to illumination. Model-based systems depend on the definition of speech related features by the user. The definition may therefore not include all speech relevant information and features like the visibility of teeth and tongue which are difficult to represent.
The early systems performed well for a speaker independent recognition task, but it did not contain any intensity information which might provide additional speech information. Here we extend this system by augmenting the feature vector with intensity information extracted from the mouth region. We evaluate the contribution of intensity information separately and in combination with shape features.

II. SHAPE MODELLING

For modelling the shape variability of lips, we use an approach based on active shape models. These are statistically based deformable models which represent a contour by a set of points. Patterns of characteristic shape variability are learned from a training set, using principal component analysis (PCA). The main modes of shape variation captured in the training set can therefore be described by a small number of parameters.
The main advantage of this modelling technique is that heuristic assumptions about legal shape deformation are avoided. Instead, the model is only allowed to deform to shapes similar to the ones seen in the training set. Any shape x representing the co- ordinates of the contour points can be approximated by

$$x=x' + Pb$$

Where x' is the mean shape, P the matrix of eigenvectors of the Covariance matrix and b, a vector containing the weights for each eigenvector. Only the first few eigenvectors corresponding to the largest eigenvalues are needed to describe the main shape variability.
Shape model for the inner and outer lip contour with profile vectors, perpendicular to the lip contours.
Lip model with mean shape and mean intensity
We built and tested two models of the lips: Model 1, which represents the outer lip boundary only and Model 2, which represents the outer and inner lip boundary. The models are used to locate, track and parameterize lip movements in image sequences. The weights for the shape modes are recovered from the tracking results and serve as features for the recognition system

III. INTENSITY MODELLING

Several approaches for speech reading, based on intensity information have been developed. Our approach for extracting intensity information is based on principal component analysis and is related to the exigent lips. This approach placed a window around the mouth area on which PCA was performed. Since the window does not deform with the lips, the eigenvectors of the PCA mainly account for intensity variation due to different lip shape and mouth opening. We already obtain detailed information of the lip shape from our shape model by a small number of parameters and are therefore mainly interested in intensity information which is independent of lip shape.
We follow an approach, where one dimensional profile is sampled perpendicular to the contour at each model point as shown in Figure 1. But instead of using local grey level models we construct a global grey-level model by concatenating the vectors of all model points to form a global intensity vector h. We then estimate the covariance matrix of the global profile vectors over the training set and perform PCA to obtain the principal modes of profile variation. Any profile h can now be approximated by where is the mean profile, Pg the matrix of the first column eigenvectors, corresponding to the largest eigen values and bg , a vector containing the weights for each eigen vector.
Example images of a person saying the word "three" with tracking results

IV. LIP TRACKING: MATCHING THE INTENSITY MODEL TO THE IMAGE

The profile model was initially designed and tailored to enable robust tracking of the lips rather than to extract speech information from the profile vectors. The profile model is used to describe the fit between the image and the model. During image search the model is aligned to the image as closely as possible by calculating the optimal weights for the first few eigenvectors. The mean square error (MSE) between the aligned profile and the image is used as cost and a minimization algorithm deforms the shape model to find a minimum cost. The profile weight vector for aligning the model is found using. The profile vectors deform with the shape model and therefore always represent the same object features. The weight vector bg provides information about the principal modes needed to align to the image. We recover the weights from the tracking results and use them as speech features.

V. SPEECH MODELLING

The weights for the shape model and the intensity model are extracted at each image frame to form frame dependent feature vectors for the recognition system. We use either the shape parameters or the intensity parameters or both parameter sets as feature vector for the recognition system. Assuming accurate tracking performance, the shape and intensity parameters are invariant to translation, rotation and scale. The intensity modes account for both, illumination differences and differences due to the visibility of teeth and tongue and protrusion.
Dynamic speech information is important and often less sensitive to inter speaker variability, i.e. intensity values of the lips will remain fairly constant during speech while intensity values of the mouth opening will change during speech. The intensity values of the lips will vary between speakers but the temporal changes of intensity might be similar for different speakers. Dynamic features will therefore be more robust to different illumination and different speakers.

VI. CONCLUSION

The world of computing has a lot to gain from neural networks. Their ability to learn by example makes them very flexible and powerful. Neural networks also contribute to other areas of research such as neurology and psychology.
We have described lip reading system that uses both, shape and intensity information. An important property of the intensity model is that it deforms with the lip contour model in order to represent the same object features after lip movements. Recognition tests using only intensity parameters indicate that much visual speech information is contained in grey level information which might account for protrusion or visibility of teeth and tongue. Recognition performance was slightly higher for intensity features than for shape features and their combined use outperformed both feature sets.
This excellent application in lip reading is under research and expected to give out lot of fruitful outcomes. Its wide usage for the impaired adds more importance to this application.

References

[1] Advanced X Video Converter, Version 5.0.3. The World Wide Web address is www.aoamedia.com.
[2] Cubic Spline Interpolation, Sky McKinley and Megan Levine, Math 45: Linear Algebra.
[3] Matlab,The MathWorks, Inc. (Copyright 1984-2005) Version 7.1.0.246 (R14) Service Pack 3. The World Wide Web address is www.mathworks.com.
[4] Neural Network Clustering Based on Distances between Objects, Leonid B. Litinskii, Dmitry E. Romanov, Institute of Optical-Neural Technologies Russian Academy of Sciences, Moscow.
[5] Neuro Dimension, Inc. (2004). Developers of Neuro Solutions v4.01: Neural Network Simulator. The World Wide Web address is www.nd.com Gainesville, FL.
[6] Predicting Admission Counselling Triumph of Colleges Using Neural Networks, Maitrei Kohli, Priti Puri, 7th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Databases (AIKED'08), University of Cambridge, UK, Feb 20-22, 2008.
[7]http://paper.ijcsns.org
[8]http://www.ukessays.com/essays/computer-science/lip-reading-using-neural-networks-computer-science-essay.php

