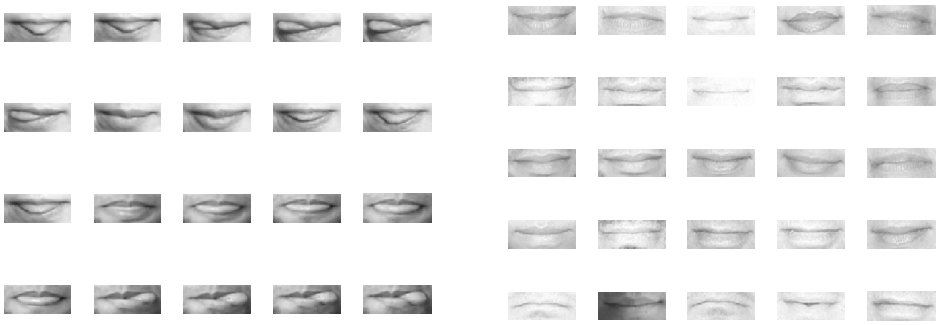
3P-3

FRONTAL VIEW TONGUE-POSITION DETECTION USING A WEBCAMERA

Luis SAPAICO[1] Suguro Saito[1] and Masayuki Nakajima[1 2]

[1]Graduate School of Information Science and Engineering, Tokyo Institute of Technology
[2] National Institute of Informatics, Japan



1. Introduction

Sometimes we continue doing the things we are accustomed to in the way we have always carried them out, but other times curiosity takes us and we wonder how it would be if we try to change. Computer environments are not far from this situation, on the contrary, it is constantly changing and improving; giving us more alternatives in order to enjoy the time we spent in front of the screen. Web-cameras are part of this "revolution", the first one was created some 15 years ago ([1]), and now it seems illogical to buy a computer without one. Nevertheless its primary usage (communicating with friends or family through the Internet), in this thesis we try to find it a new utilization: we use a common non-expensive web-camera for detecting and extracting facial information, the tongue, such that the system detects the tongue on the right or left side of the mouth; or in the middle of it, always between the lips. The applications for such a environment range from a mouse-clicking system, very useful for people with physical disabilities, to more complex tracking or gesture recognition systems.

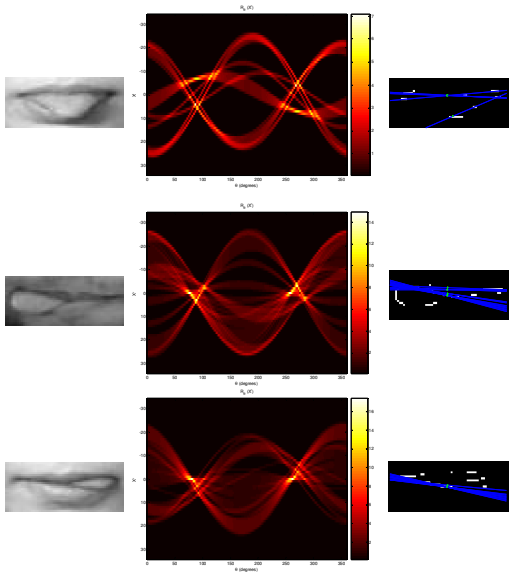
2. Method Overview

For creating such a system, we decide to create and train a learning algorithm, such as it is trained during some offline period of time, and when the user starts using it, the system will just detect the user's facial characteristics and evaluate whether the tongue is present or not in the image. On top of that, after knowing if the tongue is present, the algorithm will find the location of the tongue within the mouth region.
For that purpose, we first need to input data for the system to learn the environment under which it will work. Unfortunately, because this type of research is completely new, there is no available database for its training. Therefore, we needed to create our own database, consisting in "mouth-region" images extracted from video shot with the webcam. For this, we needed to extract manually those regions and then we had to normalize their dimensions, because due to distance of the user from the PC, or due to other factors, the extracted mouth region was not uniform in size. We work in the HSV Color Space, because of the redness of the lips ([2]), which makes it easy to find them. Hence, we created a database with inputs of 25x60 pixels, which also will give us some benefits when learning because is dimensions are not exaggerate. Some results of the "Tongue Position Detection" database are shown in the Figure on top of this page.

The next step in the solution of the problem was to properly train the learning algorithm. Since we have images where the tongue is present and images where it is not, we can roughly classify our data into "Tongue" vs. "Not-tongue"; hence, we decided to use Support Vector Machines due to its facility to deal with 2-class problems. After evaluating with different parameters, we found out that the best classification was obtained when using a Gaussian Kernel of sigma=32. And, for minimizing the Generalization error we used 10-fold Cross Validation. Our training data consisted of 135

images (69 ‘tongue’ vs 66 ‘not-tongue’); with some images taken also from the CALTECH Frontal Face Dataset. We obtained a successful 5.5% percent of training error. Curiously, when we tested the system, the error rate went even lower, close to 0.01%.

Once we know the tongue is somehow present in the image, we have to detect the position it is located in the surface of the mouth.
For achieving this, we use the Hough Transform, since we realized that in all the images in which the tongue is present, there is always lines product of the edges between the tongue and the upper or lower lips. Therefore, we tried to find the position by calculating in the Hough space, the points with the most energy, i.e., lines in our gray-scale mouth region image.
Some results are shown in the following figure.



As we were expecting, we obtained lines corresponding to the edges in our intensity image. Locating the position of the tongue now is only matter of analyzing the results: apparently, there is no problem about locating the tongue in the middle of the mouth; however, the results from the left and right side are somehow similar. Then, all we have to do is to analyze the image in its original form, and notice that the line present there is, in the case of the left-tongue image, located greatly over the right sector; and vice versa, the line in the case of the right-tongue image is located greatly over the left-sector.

3. Discussion on the Results

Although some results were presented in the Section 2, we will now analyze their meaning.
For instance, when testing the learning algorithm, we got an incredible 0.01% of error rate. Indeed, this rate is likely to be affected by a bias in the learning, since most of the images that considered the tongue belonged to a very limited number of people, in comparison to the variety of subjects we found for the case of training sets in which the tongue is not present. However, the approximation to the Generalization Error by using the 10-fold Cross Validation, was more “real”, indicating that it is highly likely that this system will perform well in normal conditions.
Regarding to the Hough Transform results, as we mentioned, it was clear when the tongue was located in the middle, but it was not when it was to one of the sides. We believe that this is, in part. Because of the symmetry present in those two cases. However, by including an extra step the solution is obtained.

4. Conclusions and Future Work

We have introduced a new pattern recognition problem and we were successful in solving it. However, there is indeed the necessity of expanding this database in order to make it more trustable, e.g., dark skin-color were not included in the experiments or in the testing.
The results obtained make us believe that this is, in fact, a ‘researchable’ topic; with several applications. As an example, we can imagine the utilization of this system together with a head-tracking module in order to give physical disabled people the opportunity to control the computer and stay up-to-date like most of us does.

5. References

[1] Quentin Stafford-Fraser, The Life and Times of the First Web Cam: When convenience was the mother of invention, *Communications of the ACM*, Vol.44, No.7 pp. 25-26, July 2001.
[2] Zhang, X., Mersereau, R.M., Lip feature extraction towards an automatic speechreading system, ICIP, Vol. 3, pp. 226-229, 2000