Nasolabial Wrinkle Classification

David Lluncor

University of California, San Diego

9450 Gilman Drive, La Jolla, CA, 92092

dlluncor@ucsd.edu

Abstract

Categorization of facial wrinkles is used in dermatology studies to determine the efficacy of a certain treatment over a certain period of time. These studies, however, are limited by the fact that when dermatologists assign these categorizations, it is done so through qualitative, in-vivo observation. The objective of this study was to train a computer to categorize nasolabial wrinkles into one of four categories based on one image of the patients face. Preliminary results suggest that the computer will be able to correctly categorize category 1, 2, 3, and 4 wrinkles with 88.2%, 57.3%, 26.2%, and 80% success, respectively.

1. Introduction

As the fields of cosmetic surgery, dermatology, and plastic surgery continue to grow, more and more products are developed to help eliminate blemishes and facial wrinkles. Dermatologists have developed several methods [1] to categorize wrinkles into distinct categories depending on wrinkle length, depth, and overall how it looks. There are several methods which take into account different aspects of a wrinkle, but all in all it is up to the subjective observation of a dermatologist to assign these wrinkle categories. While there is a fairly strong correlation among dermatologists who assess the same patients [2], there is little quantitative evidence that is used in the process of determining wrinkle severity. Companies who develop aesthetic products use these scores to determine how well a treatment has helped to alleviate wrinkles over a period of time. Some have quantified wrinkles using plastic molds [2] to measure the actual depth of a wrinkle and using its depth to categorize the wrinkle. This technique is quite invasive and our goal is to create a method which is cheap, effective, and has a quick turnaround time.

In this paper I develop an approach to categorize wrinkles based on various light intensities that emit from the skin, where a more pronounced wrinkle is likely to have less light intensity emission at the point of the wrinkle. The preliminary results (n = 23) show that indeed the “intensity profile”, that is the pattern of light emission, from wrinkles of category 1-4 exhibit distinguishable patterns.

2. Approach

It is presumed that from measuring the light intensity values across a part of the face which contains the wrinkle we can extrapolate the physical characteristic of how deep a wrinkle is. While we are not actually measuring the physical depth of the wrinkle itself, we are using the fact that a more pronounced wrinkle produces a greater shadow to leverage taking solely the intensity profiles from an image. The images were taken using a Visia Camera. The method for extracting the intensity profile from an image can be broken into three simple steps.

2.1 Visia Scanner

Fig 1. The Visia scanner was used to taken images from 45°, 0°, and -45° respective to the front of the face in seven different illumination settings.

The Visia skin analysis machine is a standard in skin complexion analysis [3]. The scanner can take pictures using up to seven different illuminations, however, there was only one illumination used from the 0° perspective to analyze the data. In future studies, I may consider using different perspectives and illuminations to analyze the data.

2.2 Relevant Pixel Markers

A standard approach was used to extract four points on the face: the right corner of the lip, the left corner of the lip, the right corner of the nose, and the left corner of the nose.

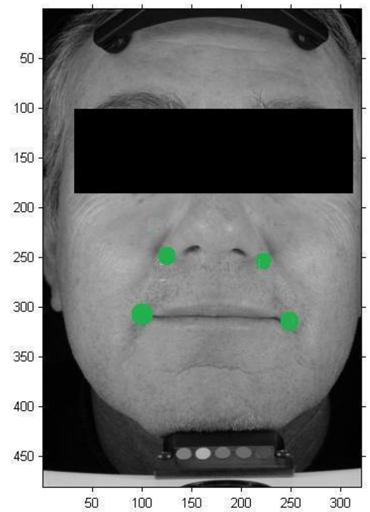

Fig 2. Example of where the four points that were selected from each patient.

These four markers were selected because they can easily be detected by certain face recognition algorithms [4]. In future works the process will be automated such that rather than a user inputting where to select relevant pixels, a computer will ideally locate where the corners of the mouth and nose are.

2.3 Intensity Profile Lines

Using the four point markers, a line was then extracted from the black and white image of the light intensity values. Two different approaches were used to take the intensity profile: averaging values and the worst-case value.

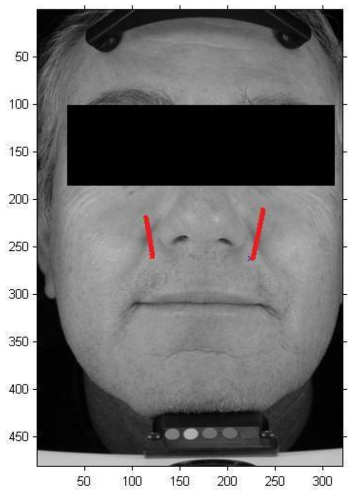

Fig 3. Example of where the intensity values were taken from when analyzing the worst-case.

The worst-case line, selected among 10 parallel lines, is defined as the line which had the greatest difference in lowest and highest intensity value within the line consisting of approximately 80 pixels.

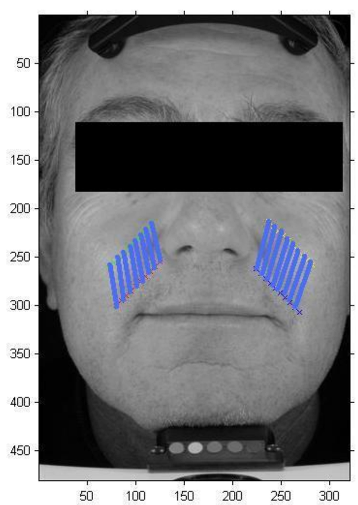

Fig 4. Example of where the intensity values were taken from when the average-case.

In this approach, all lines were treated equally and the intensity values were averaged at concurrent locations in each of the lines.

3. Training and Testing

The k-nearest neighbor approach was used to categorize the wrinkle based on two input parameters: the the fit slope (m) and fit b-intercept (b) of the intensity values to the line $y = |x|$ for a particular wrinkle. Majority voting was used when $k = 3$ and $k = 5$ to determine the class, while ties were broken based on the class of the nearest neighbor. Neighbors were determined using the Euclidian distance between the two input parameters.

There were 17 intensity profile examples from category one, 11 intensity profile examples from category two, 12 intensity profile examples from category 3, and 6 intensity profile examples from category 4.

To use cross validation, one intensity profile was omitted from each of the four categories and was used as the test data while the rest of the examples were used as training data.

To fully support the use of five-fold cross validation, the number of examples from category 4 will need to increase to match the number of examples provided by categories 1, 2, and 3. Also, less data will need to be used as training data in each configuration.

As a result of such a method for cross validation, there were 11,220 different configurations for which the test data and the training data was chosen.

4. Data and Results

There were 23 patients whose nasolabial wrinkles were analyzed. There were 26 patients in total whose intensity profiles were collected, but because of mustaches and other visual distractions that would hinder the use of the algorithm described in this paper, those 3 patients were omitted.

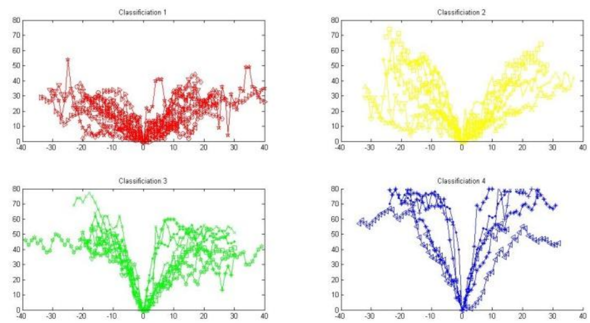

Fig 5. Graph of the intensity values across 23 patients taken from the worst-case.

Based on the ground truth information established by one dermatologist who analyzed the patient's wrinkle severity *in vivo*, the intensity profiles were mapped to their appropriate graph. The red represents category 1, the yellow category 2, the green category 3, and the blue category 4. It is clear that there is a distinguishable disparity in slope across the four categories. The approach to be taken is to perform parametric and non-parametric data analysis. The parametric method would involve fitting the data to a graph such as y=|x|.

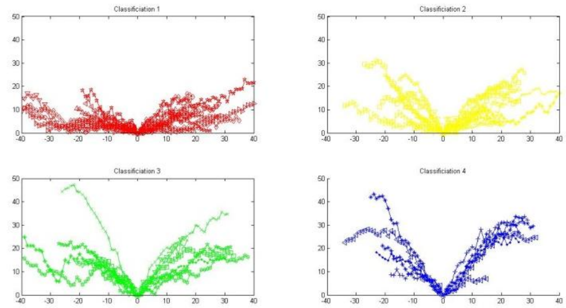

Fig 6. Example of the intensity values across 23 patients taken from the average-case.

A similar graph was generated as to the average case. There is less noise in the data which is to be expected, however, there is less of a noticeable difference in the shapes of the different data plots.

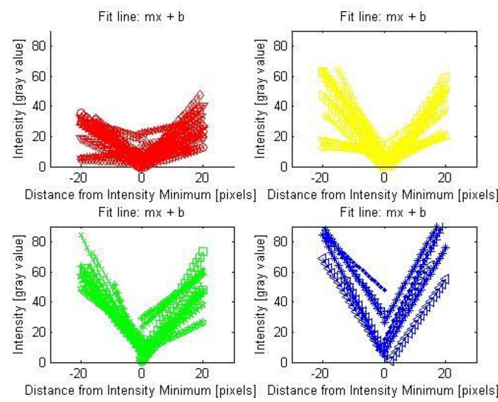

Fig 7. Graph of the intensity values fit to y = |x| for 23 patients taken from the worst-case.

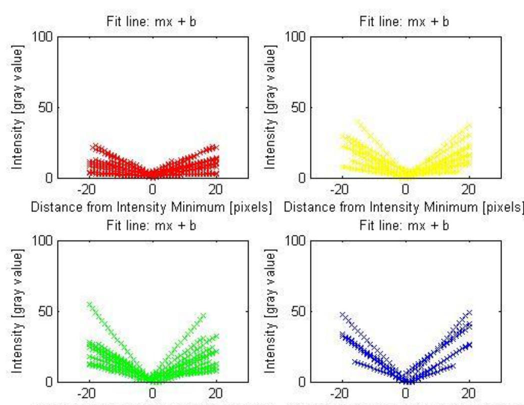

Fig 8. Graph of the intensity values fit to y = |x| for 23 patients taken from the average-case.

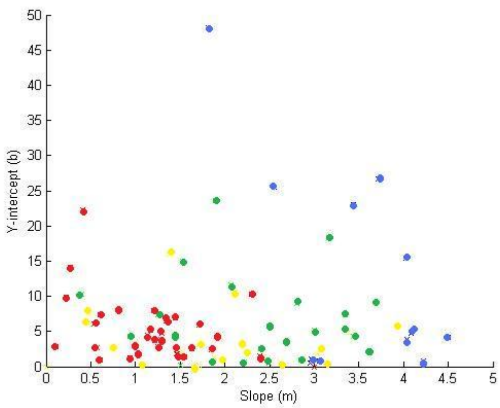

Fig 9. Graph of the fit slope (m) on x-axis to fit b-intercept (b) to the line y = |x| for 23 patients taken from the worst-case.

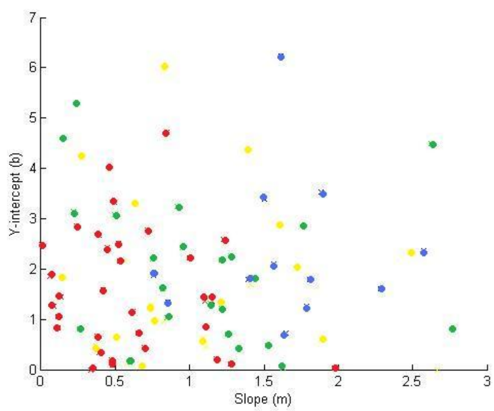

Fig 10. Graph of the fit slope (m) on x-axis to fit b-intercept (b) to the line y = |x| for 23 patients taken from the average-case.

From these graphs, it appears as if there is some sort of gradient in which the steeper the slope and the larger the b-intercept then the more severe the wrinkle. These two parameters were as parameters when calculating the Euclidian distance between feature vectors (intensity profiles) for training and testing the data.

When using worst-case analysis, for k = 5, category 1, 2, 3, and 4 wrinkles were predicted with 88.2%, 3.6%, 26.2%, and 80% success, respectively. When k = 3, category 1, 2, 3, and 4 wrinkles were predicted with 73.7%, 5.2%, 30.9%, and 80% success, respectively. And when k = 1, category 1, 2, 3, and 4 wrinkles were predicted with 55.5%, 3.6%, 20.6%, and 80% success, respectively.

When using average-case analysis, for k = 5, category 1, 2, 3, and 4 wrinkles were predicted with 68.5%, 56.1%, 10.2%, and 23.3% success, respectively. When k = 3, category 1, 2, 3, and 4 wrinkles were predicted with 84.3%, 57.3%, 10.2%, and 23.6% success, respectively. And when k = 1, category 1, 2, 3, and 4 wrinkles were predicted with 60.9%, 56.1%, 10.1%, and 23.3% success, respectively.

All the previous information is summarized below.

% Success of Classification Given Selecting Method

	Class1	Class2	Class3	Class4
Worst-case				
k = 1	55.5%	3.6%,	20.6%,	80.0%
k = 3	73.7%,	5.2%	30.9%	80.1%
k = 5	88.2%	3.6%,	26.2%	80.4%
Average-case				
k = 1	60.9%	56.1%	10.1%	23.3%
k = 3	84.3%	57.3%	10.2%	23.6%
k = 5	68.5%	56.1%	10.2%	23.3%

It appears as though using worst-case analysis is most beneficial when discriminating between the most severe and least severe wrinkles. However, the category 2 and 3 results are dismal, as well as for the average-case scenario as well.

There are numerous parameters that can be tweaked in this model, such as the location of the relevant pixel markers, the number of images used from one particular patient, the number of lines averaged over or taken the worst-case of for a particular intensity profile of a patient, the number of pixels taken from one side of the wrinkle to the other, the model to fit the data to (other than y = |x|) and the list continues. Therefore, there needs to be some fine tuning of parameters I believe to make this classification as robust as it can be given the raw data.

5. Conclusions

Currently it is not clear as to what is the best approach in terms of how to best analyze the data. Two approaches were attempted, and it seems that the worst-case scenario might be more representative of how wrinkles are categorized by a dermatologist. There seems to be more success in the prediction of category 1 and 4 wrinkles given the worst-case scenario, with slightly better results in category 2 for the average-case scenario.

However, it is notable that these graphs where the four wrinkle categories have distinguishable shapes seems to suggest that taking one image from the frontal perspective using standard lighting conditions produces promising results. In future studies it would be worthwhile to note whether taking an image using a handheld camera as opposed to a more expensive machine such as the Visia Scanner could produce similar results. Also to improve the success rate of predicting a particular wrinkle, it might be worthwhile to use images given several lighting schemes from several different angles.

Another important feature to be analyzed is whether automatically detecting the feature points of the corner of the mouth and nose can generate similar results as in the case of manually selecting points.

Limitations and future work need to account for the presence of moustaches which may severely skew the intensity values that are pulled from the face.

6. Acknowledgements

Thanks to Serge Belongie and the Winter 2010 CSE 190 course for the close guidance and suggestions in moving forward with this project. Also, very big thanks to Vera Morhenn who used the Visia Scanner at Clinical Therapeutics Research to acquire the data from the patients, that was a big help!

7. References

1. Hamilton, D. A classification of the aging face and its relationship to remedies. *J. Clin. Dermatol.* Summer 1998: 35, 1998.

2. Lemperle, G., et al. A classification of facial wrinkles. *Plastic and Reconstructive Surgery*. Nov. 2001: Vol 108, Issue 6.

3. Yuen, C. et al. Automatic detection of face and facial features. Pgs. 230-234, 2008.