Age Estimation from Facial Images

Carl Olsson (carl_olsson@brown.edu)
Department of Engineering, 69 Brown St
Providence, RI 02912 USA

Abstract

This research investigates computational age estimation from facial images. The goal was to approach human performance. Using two novel features as well as the features from the Kwon-Lobo paper, binary classification accuracy improved (97.84% vs. 92.94% AUC). Incorporating wrinkle analysis improved results, with about 96% total accuracy and 98% AUC. Binary classification also agreed with evidence that humans estimate age by fixating on the center parts of the face. Regression almost approached human age estimation accuracy, with mean absolute errors under 10 years for ages between 0 and 21. However, adding wrinkle analysis worsened regression accuracy (11.75 vs. 9.37 years MAE).

Keywords: Computer vision, machine learning, age estimation

Introduction

Estimating someone's age is very intuitive for humans. On the other hand, estimating a person's age computationally is difficult. This research addresses this problem and aims to approach human performance in age estimation.

Age estimation is a difficult problem because different people age differently. For example, sun exposure and other environmental factors can lend the appearance of older age. Moreover, people of different genders and ethnicities may age differently.

A successful technique for age estimation would be useful in many scenarios such as biometrics, security, and marketing.

Background

Typically, aging from infancy to puberty consists of craniofacial growth and changes in head size and shape. Past puberty, aging consists of subtle changes in facial geometry, aging of the skin, and changes in hair color and hair growth. This suggests that important features to consider for age estimation are face shape and size, facial geometry, wrinkles, and hair color and growth.

However, a study of gaze behavior during age estimation showed that participants' eyes remained fixated within an area which covers the eyes, the nose and the middle upper part of the mouth. This suggests that external features such as hair color and growth aren't important in biological age estimation. Therefore, facial feature geometry and wrinkles were used to estimate age.

This research relied on the publicly available FG-NET aging database which contains 1,002 color and grayscale images of people from 0 to 69 years of age. Each image had 68 manually annotated points marking facial feature locations.

Method

All computation was done with the MATLAB software package. A brief outline of the method used is presented next:

A. Locate facial features
   a. Find center of pupils, mouth, nose, outline of jaw
B. Compute facial feature ratios
C. Wrinkle analysis
   a. Mask wrinkle-prone areas
   b. Filter areas with bank of Gabors
   c. Compute texton dictionary
   d. Compute bag of textons representation
D. Run classification/regression

Locating facial features

To locate important facial features the researcher parsed the corresponding points file contained in the FG-NET database, because localizing facial features is especially difficult when the images have huge variations in lighting, pose, and facial expression.

Computing facial feature ratios

Facial feature ratios were used for classification because they are robust to changes in facial expression and image quality. Of the eight ratios used, six are from the Kwon-Lobo paper on age classification.

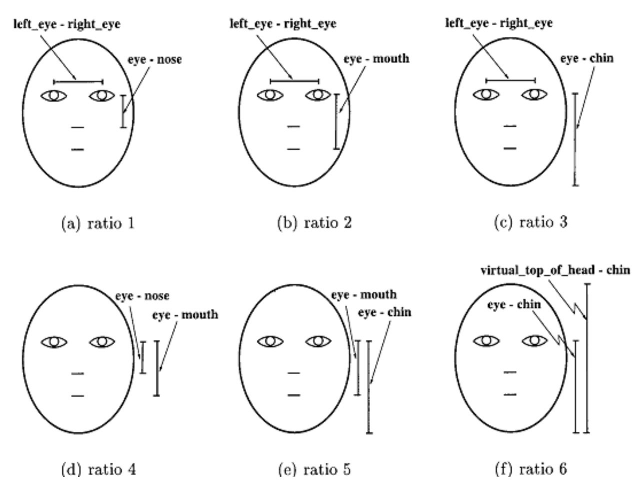

Figure 1: The six ratios from Kwon-Lobo

The six ratios from Kwon-Lobo were computed using the locations of salient features. Simple linear algebra was used to find the distances between features.

Two novel features were also used to improve classification accuracy. The first is based on face shape. Face shape in infancy is close to circular but as aging occurs the face elongates and becomes less circular. To calculate the circularity of the face, an ellipse is fit to the jawline. The ratio is calculated as follows:

$$\text{first ratio} = \left|\frac{major\ axis}{minor\ axis} - 1\right|$$

This ratio was chosen over eccentricity because it resulted in much better classification accuracy.

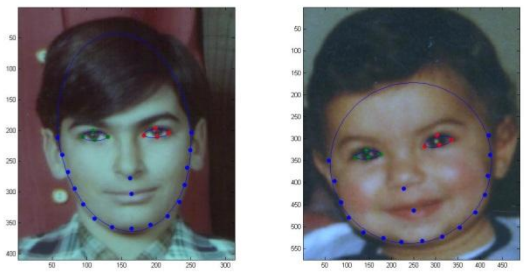

Figure 2: Sample images displaying the computation of two novel facial feature ratios

The second novel feature is based on the relative areas of eyes to the rest of the face. Both the eyes and the face are approximated as ellipses. The ratio is calculated as follows:

$$\text{second ratio} = \frac{total\ area\ of\ eyes}{total\ area\ of\ the\ face}$$

Wrinkle Analysis

The first step taken in wrinkle analysis was to mask wrinkle-prone areas of the face. In this case, the left and right cheeks and chin were masked. To speed up computation, the masked areas were cropped and resized.

A bank of 8 Gabor filters at 5 scales was used to filter the masked images. In turn, these filtered images were used to generate a texton dictionary using kmeans clustering.

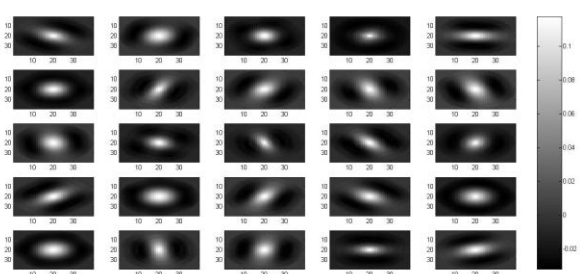

Figure 3: Texton dictionary created from cheek images

Using these texton dictionaries, a bag of textons representation is created. This stores the best matching texton for each pixel in the image by computing the Euclidean distance between the filter output and the dictionary of textons.

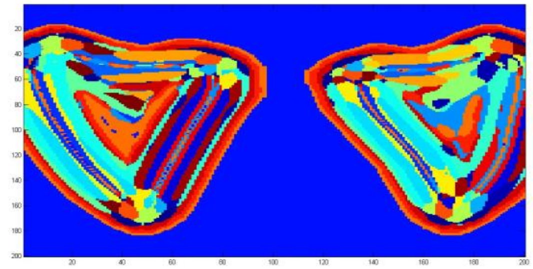

Figure 4: Bag of textons representation for cheek image

The bag of textons representation is then used to calculate a histogram of textons for each image. A texton histogram for the cheeks and the chin was computed. Each feature was calculated using their respective texton dictionaries. For example, the chin wrinkle analysis used the texton dictionary computed from chin images.

Classification/Regression

This research utilized the commonly used and freely available LIBSVM software package. Binary classification between infant/non-infant classes was used to test initial age classification performance. Binary classification was also used to evaluate the saliency of different features, via their weights assigned during classification. Regression was then used to test age estimation.

Results

Binary Classification

Using only the six ratios from the Kwon-Lobo paper (See Figure 1) to classify images into the infant/non-infant classes resulted in a total accuracy of 96.94% and 92.94% area under the ROC curve.

Using the two novel features in addition to the previous six features increased total accuracy to 97.17% and 97.85% area under the ROC curve.

Using only the bag of textons representations for chin and cheek areas resulted in 96.12% total accuracy but only 75.74% area under the ROC curve. This is because the classifier identified all the test images as non-infants. Since there are not many infant images, the total accuracy was not affected much by the incorrect classification. However, using wrinkle analysis with the eight facial feature ratios resulted in a total accuracy of 96.13% and 97.99% area under the ROC curve.

Although the classification using all eight ratios and wrinkle analysis resulted in lower total accuracy (96.13% vs. 97.17%), the standard deviation of the accuracy was much lower (0.5259% vs. 0.0768%) and more area under the ROC curve (97.99% vs. 97.85%), meaning there were less false negatives and false positives.

Table 1: Results of binary classification

	Total Accuracy	Standard Deviation	AUC
6 Ratios (Kwon-Lobo)	96.94174757	0.4817	0.929404
8 Ratios (2 novel features)	97.1723301	0.5259	0.978461
Just wrinkle analysis	96.11650485	0	0.757353
8 Ratios with wrinkles	96.12864078	0.0768	0.979909

Looking at the weights assigned to features during binary classification yields information about the saliency of various features in age estimation. As is apparent in the figure below the first four ratios identified by Kwon-Lobo are especially important during age classification, as well as the two novel ratios.

This reflects actual human age estimation behavior. A study of gaze behavior during age estimation showed that participants' eyes remained fixated around the eyes, nose, and mouth. Likewise, the features that were important during classification were concentrated around the eyes, nose, and mouth (as opposed to the chin and top of the head).

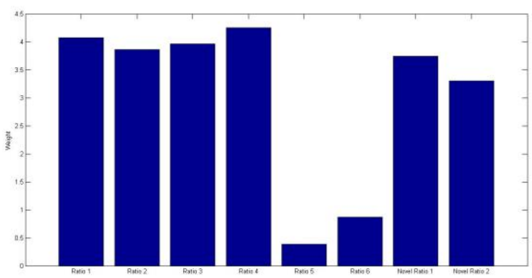

Figure 5: Weights of features used during binary classification

Regression

The results of SVM regression appear below. Unlike binary classification, adding wrinkle analysis made regression perform much worse, with a mean absolute error of 11.75 years versus the error of 9.36 years for just using the eight facial feature ratios.

Table 2: Results of Regression

	Mean MAE	Standard Deviation
6 Ratios (Kwon-Lobo)	9.4647	0.6694
8 Ratios (2 novel features)	9.3671	0.6604
Just wrinkle analysis	12.9042	1.2821
8 Ratios with wrinkles	11.7527	1.0418

Given that the FG-NET database over represents younger ages, it may be somewhat obvious that the regression performs better at younger ages.

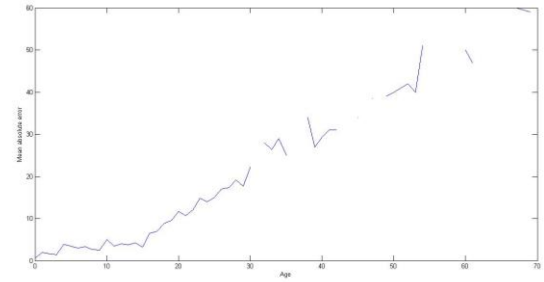

Figure 6: Age vs. mean absolute error

Conclusion

This research investigated computational age estimation from facial images. The goal was to approach human performance in age estimation. Using two novel features as well as the features from the Kwon-Lobo paper, binary classification accuracy improved (97.84% vs. 92.94% AUC).

Incorporating wrinkle analysis pushed classification accuracy even higher, resulting in around 96% total accuracy and 98% area under the ROC curve.

Binary classification also agreed with evidence that humans estimate age by fixating on the center parts of the face: the eyes, nose, and mouth.

On the other hand, regression almost approached human age estimation accuracy, with mean absolute errors under 10 years for ages between 0 and 21. However, adding wrinkle analysis drastically worsened regression accuracy (11.75 vs. 9.37 years MAE). This could be improved by having higher resolution images or experimenting with alternative texton dictionaries.

Acknowledgments

The author is grateful to T. Serre for his helpful advice and leading a great class; and to A. Arslan for his advice.

References

Guo, G., G. Mu, Y. Fu, and T. Huang. "Human Age Estimation Using Bio-inspired Features." *CVPR* (2009): n. pag. Print.

Kwon, Young H., and Niels V. Lobo. "Age Classification from Facial Images." *Computer Vision and Image Understanding* 74.1 (1999): 1-21. Web. 15 Dec. 2012.

Porcheron, A., W. Jenner, J. Pripfl, M. Pivec, M. Binder, and F. Morizot. "Perception and Gaze Behaviour: How Do We Look at Faces?" *Perception* 38 (2009): 76. Web.

Additional Figures

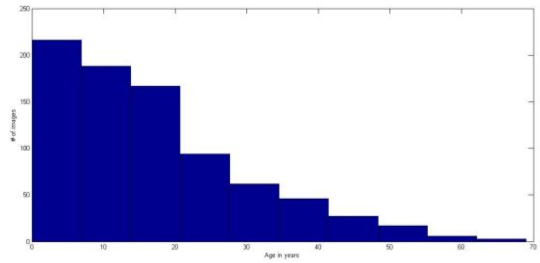

Figure 7: FG-NET databse age distribution

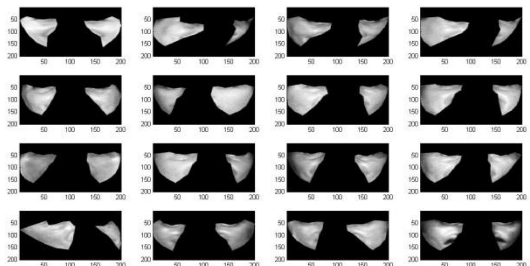

Figure 8: Cheek masks cropped and resized to 100x100

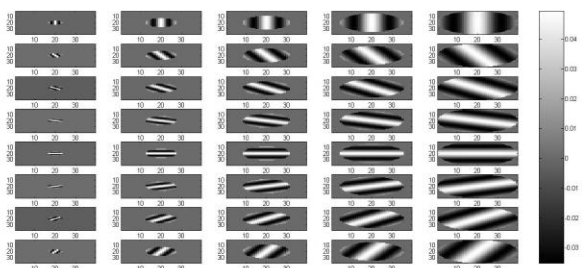

Figure 9: Bank of gabor filters used for wrinkle analysis

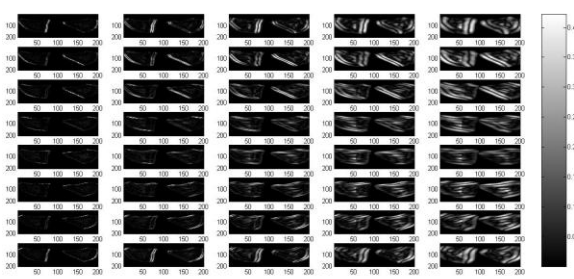

Figure 10: Sample filtered cheek images

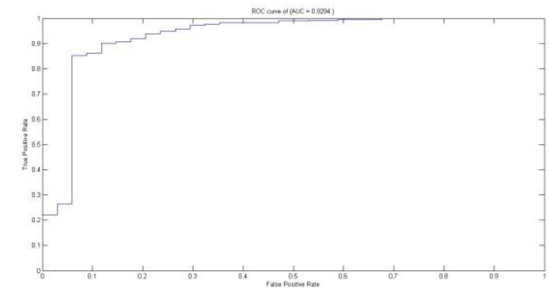

Figure 11: ROC curve for binary classification using the 6 Kwon-Lobo Ratios

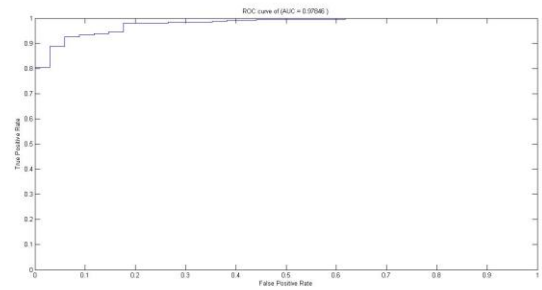

Figure 12: ROC curve for binary classification using the 6 Kwon-Lobo Ratios and two novel ratios

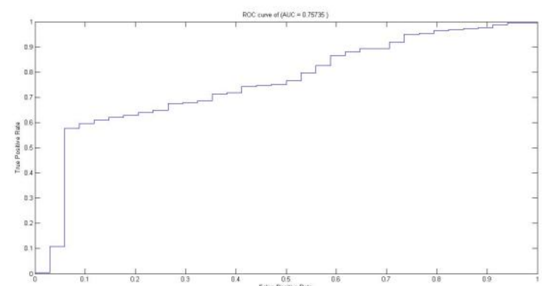

Figure 13: ROC curve for binary classification using just wrinkle analysis

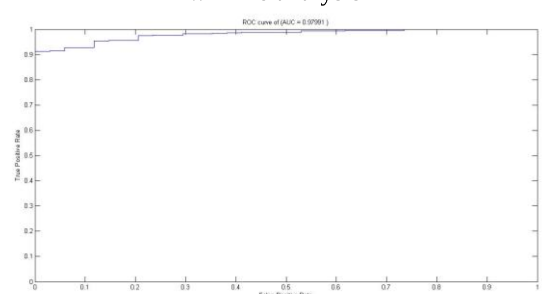

Figure 14: ROC curve for binary classification using all eight ratios and wrinkle analysis