Naïve Bayesian Minor Detection Using Facial Anthropometric Ratios

Chad M. Steel,
Virginia Polytechnic Institute and State University
Computer Science Department
csteel13@vt.edu
P.O. Box 6136
McLean, VA 22101

Abstract

Facial anthropometry is the measurement of facial features and proportions for medical purposes. The use of anthropometric measures as dimensions in an age detection naïve Bayesian classifier is used to accurately determine if the individual depicted in a digital image is a minor (<18 years of age) or not. The resulting classifier is shown to have excellent performance (Precision - 0.89418, Recall - 0.871134) in discriminating images of children from those of adults. Additionally, the key anthropometric ratios for age recognition are identified.

Keywords

Age detection, facial recognition, anthropometric ratios

Highlights

Our research provides several major contributions to the field of age classification:

- First, a classifier is built that outperforms the current-best facial age classifiers.
- Second, the best ratios for facial anthropometry are identified.
- Third, the use of anthropometric ratios in a naïve Bayesian classifier is shown to have an experimental performance greater than that of cardioidal strain.

- Finally, it is shown that anthropometric ratios are correlated with age, but are more strongly correlated with developmental stage.

1 Introduction

A specialized application of age detection algorithms is the detection of minors based on facial characteristics. The ability to detect if an individual is below the age of 18 is useful forensically in the detection of child pornography. The use of facial features to appropriately classify age expands the applicability of classifiers based on the Tanner Scale [1] to applications where body-specific feature classifiers are not possible due to clothing obscuration. Specific applications include automated cigarette machines validation [2], scanning of casino surveillance for under-age patrons, and entry into nightclubs.

Using facial anthropometry we identify the key ratios associated with age, classify a set of labeled images based on the ratios, and compare the results to current techniques. The application of a Bayesian classifier to facial anthropometric ratios is used to train and classify images of individuals ranging from a few months old to elderly. Finally, a test of efficacy of pubescence detection vis a vis age detection is performed.

Overall, the use of anthropometric ratios are shown to be an effective classifier for age. They outperform current techniques, are efficient to calculate, and have a traceable accuracy that can be manually verified.

2 Theory

The use of faces to differentiate between pre and post-pubescent images relies on changes in craniofacial geometry. The basic ratios present in the face alter in proportion to cranial development. With infants, an increase is vertical relative to horizontal size is present, along with a less pronounced and more flat nose and an increased eye size relative to head size. These proportions gradually change until growth stops, at which point adult ratios are present. These same changes are perceptually noted by individuals when assessing the age of children as discriminators [3].

A mathematical basis was developed for the facial changes which was characterized as cardioidal strain [4]. The strain equations provide a three dimensional transform which changes the shape of the cranium (and thereby the face) in relation to perceived age [5]. These were shown to be responsible for the majority of developmental difference in cranial structure throughout the growth process [6].

Building on the anatomical work, Ramanathan et al summarized the techniques for computational age estimation based on facial characteristics [7]. Similarly, Geng et al use an eigenspace projection based on known age progressions to project an estimated age for an unknown face [8].

Guo et al found that age estimation based on a manifold model of the face was possible. Their work identified Conformal Embedding Analysis with linear regression as the best performing technique, with a mean absolute error of 5-6 years. Their error estimates were not broken up into pre-and-post pubescent groups, however, and their applicability to minor detection would be based on a lower error rate in the pre-pubescent category [9].

Yang and Ai used Local Binary Patterns with AdaBoost, and they showed that a texture-based classifier outperformed one which used Haar-like features to segregate individuals into three categories – child, youth, and old. While they do not define the age ranges, the trinary classification showed an average error rate of approximately 10%. Given the model, tuning could be used to provide a binary classifier that would be tunable to the minor detection problem [10]. A more detailed overview of general techiques is available from Fu et al. [11]

Kwon and Lobo used anthropometric ratios to classify age on a small scale coupled with wrinkles. Their work pre-selected seven ratios and used forty seven images. The technique showed promise, but their dataset was limited and the ratios were not necessarily the optimal ratios [12]. Similary, Gunay and Nabiyev propose the use of facial measurements with a neural network to do age estimation, though they do not provide extensive experimental results to evaluate against. [13]

The use of facial anthropometric measures in age detection has generally been limited to direct measurement and comparison to physical norms. For digital images, this implies the need to have the exact scale of the face in the image to renormalize the measurements to their physical counterparts. Estimating the scale introduces an additional error factor in the calculations for single image measurement and a compounded error for multi-image age comparisons.

Our approach uses ratios between defined facial points instead of direct measurements. By utilizing ratios instead of simple measurements, direct comparisons between images of different sizes is possible, and direct estimation of age can be performed without knowledge of the scale in the image.

Prior work in medicine has focused on direct measures and a few pre-determined ratios. [13] The determination of the optimal ratios for facial recognition was computationally unfeasible, however, when the initial measurement work was performed. For the corpus used in the primary experiments, there are 68 individual points. For an n-point measurement, there are $\frac{n^2}{2}$ two dimensional point measures. From these measures, there are $\left(\frac{n^2}{2}\right)^2$ separate ratios (inverse ratios not included), yielding 5.3 million ratios per image.

To identify minors, we calculate all possible ratios for a labeled data corpus, and perform multiple experiments on the corpus. The ratios with the greatest linear separability are identified, and these are then used in a naïve Bayesian classifier on a test set of images.

Theoretically, the use of anthropometric ratios should outperform cardioidal strain. The traditional equations for cardioidal strain (as depicted in Figure 1 below) are as follows [4]:

$$P \propto R_0(1 - \cos(\theta_0))$$

$$R_1 = R_0 + k(R_0 - R_0 (\cos\theta_0))$$

$$\theta_1 = \theta_0$$

To show anthropometric ratios can be used to calculate cardioidal strain, we use to the most basic orthogonal measurements – the ratio from the nose to the ear and from the nose to the center of the forehead. This ratio uses two radii, one at 90 degrees (R_1) and one at 0 degrees (R_2). Given an arbitrary R_0, the radii are defined as:

$$R_1 = R_0 + k(R_0 - R_0 (\cos\theta_1))$$

$$R_2 = R_0 + k(R_0 - R_0 (\cos\theta_2))$$

Because $\theta_1 = 0$, $R_1 = R_0$. Substituting to solve for R_2, we find:

$$R_2 = R_1 + k(R_1 - R_1 (\cos\theta_2))$$

which reduces to:

$$R_2 = R_1(1 + k)$$

therefore,

$$\frac{R_2}{R_1} = 1 + k$$

Since k is a growth factor and is directly proportional to the age of the child, the single ratio $\frac{R_2}{R_1}$ is therefore directly proportional to the age. Thus, anthropometric ratios estimate age equivalent to cardioidal strain with a single ratio, and a Bayesian classifier incorporating multiple ratios should perform equal to or better than a cardioidal strain measurement.

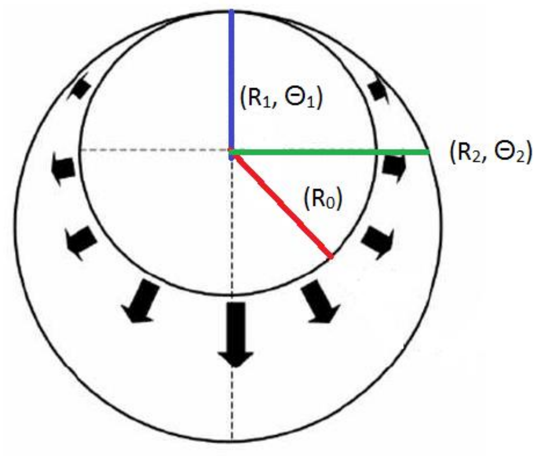

Figure 1 - Cardioidal strain

3 Methods and Results

3.1 Data Set

For this research we make use of the FG-Net aging corpus, which consists of 1002 labeled images of individuals ranging from infants to 68 years of age. The corpus contains color images of different dimensions representing human faces. The faces are pose-normalized to be largely forward-facing (though they may be angled vertically or horizontally) and have point files defining the limits of the faces within the images [14]. A sample image with the points labeled is shown in Figure 2.

For the purposes of the experiments below, the dataset was broken into a training dataset and a test dataset. The training dataset consists of 501 images and the test set the remaining 501 images. Images were randomly assigned to a set, however images of the same individual (where multiple ages were present) were grouped into the same set.

3.2 Experiment 1: Identification of Optimal Ratios

Prior to running a Bayesian Classifier, the best ratios for minor discrimination (based on linear separability) were identified. The ratios for minors and the ratio for adults in the training set was calculated. Each image in the test set was then compared to the training set to determine which classification (adult or minor) the ratio was closest to and classified as such. The top ratios in terms of individual performance showed a classification accuracy of 81.7%, with over 100 ratios providing better than 80% discrimination. The top ratios are show color-coded in Figure 2 below.

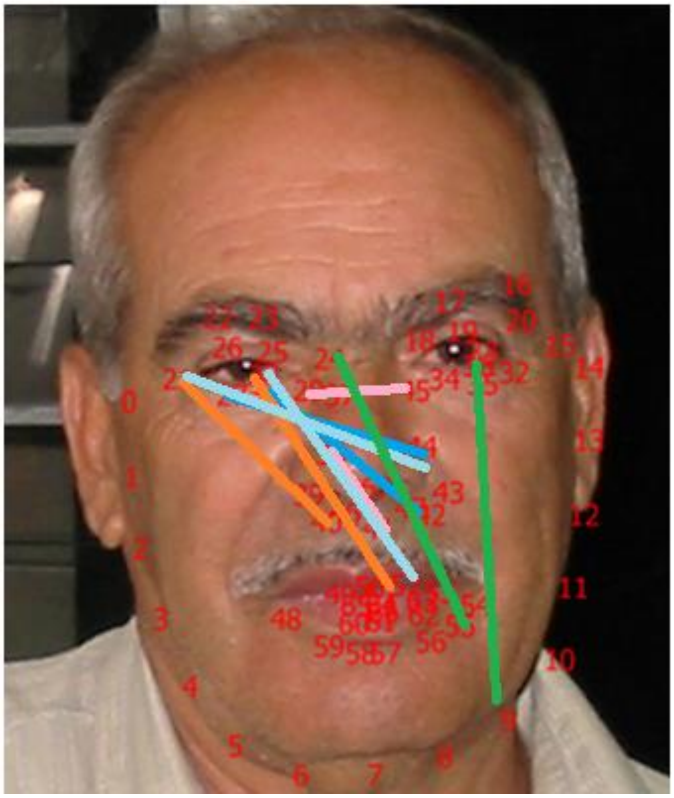

Figure 2 – Top ratios for minor detection

Overall, the top ratios tended to be non orthogonal, internal diagonal ratios. Specifically, the facial feature edge detection (points 0-14 in the diagram) appeared to be less effective overall in determining age when compared with the internal feature measurements (ratios between the nose, eyes, and mouth). This is consistent with fixed feature distances – the outer perimeter features are much more susceptible to weight fluctuation, and therefore expected to be less reliable as an overall predictor of age. The top 10 ratios, all of which have an 81% individual accuracy, are shown in Table 1 below.

Point 1-1	**Point 1-2**	**Point 2-1**	**Point 2-2**
27	44	30	42
38	42	27	44
38	41	27	44
21	40	31	51
38	41	29	45
30	52	21	44

38	41	26	46
9	36	24	55
30	51	21	44
40	51	21	30
41	46	24	44

Table 2 – Top ratios for minor discrimination

3.3 Experiment 2: Age Accuracy of Ratio Identification

Because other techniques for age detection have focused on identification of age ranges, we analyzed the ability of ratios to predict age-ranges. The age ranges were set to the year identified as labeled for 0-17 and all ages above 17 were set to 18. For any adults (those 18 or older) the ratios were not expected to change following the end of puberty. As such, they would be expected to remain fixed – therefore a single category of above-18 was selected.

The training set was used to calculate the average ratios as above but for each age bracket. For each image present in the training set (for each ratio), the nearest-match ratio was identified and the age selected. The difference between the predicted age was then determined and the mean absolute error (MAE) for each ratio pair over all images:

MAE = $\sum_{k=1}^{N} |l'_k - l_k|/N$

Where l'_k is the actual age and l_k is the predicted age. The top-performing single ratio measures indicate a mean deviation of 3.3 years, as shown in Table 2 below. Unlike the internal measures identified as top ratios, when only child age-brackets are considered (the adults are collapsed into a single category), the external measures become more relevant. Specifically, longitudinal/latitudinal ratios from perimeter locations are effective measures. Because of the general stability in relative fat distribution changes over childhood development (as opposed to adulthood), external measures can be used as age predictors once the minor/adult determination has been made.

Point 1-1	Point 1-2	Point 2-1	Point 2-2	MAE
2	4	0	14	3.30
0	14	9	13	3.31
1	5	0	14	3.36
0	13	1	4	3.36
0	14	9	12	3.36
1	4	0	14	3.39

0	14	2	5	3.42
0	14	10	11	3.42
2	4	0	13	3.42
8	35	0	15	3.43
2	4	0	14	3.30

Table 2 – Best ratios for the prediction of exact age

Using single ratios, the MAE for facial anthropometric measures are significantly better than those previously identified (3.30 v. 5.07). A comparison of measures as identified in [9] is shown in Table 3 below.

Algorithm	**FG-Net Results**
WAS [8]	8.06
AGES [8]	6.77
QM [15]	6.55
MLPs [15]	6.98
LARR [9]	5.07
BEST_RATIO	**3.30**

Table 3 – Cross-algorithm MAE comparison

3.4 Experiment 3: Naïve Bayesian Classifier

For experiment 3 we built a Naïve Bayesian classifier to utilize the anthropometric ratios identified above to perform a binary classification of the facial images in the corpus as minors or adults. As with the above experiments, all possible ratios for the set of training images were calculated as part of the classifier.

Because many of the ratios were not good predictors of age, basic dimensionality reduction was performed prior to running the classifier. Only ratios where a separation of at least .25 (determined heuristically) between the median ratio for adults and the median ratio for minors were used.

To make the classifier adjustable for different detection thresholds, a simple bias, β, was incorporated into the ratio classification such that:

$$\forall r\!: \; d_n = (r_{adult} - r_{minor}) \times \beta$$

$$(r_{minor} + d_n) \leq r_n\{\, Class = \; minor$$

Using the tunable classification above, the receiver operating characteristic curve was plotted for the classifier as noted in Figure 3.

The highest f-score performance (.89) occurred with a precision of .82 and a recall of .97. Specifically, only 3% of minors were misclassified as adults, with 18% of adults misclassified as minors. For the detection of child pornography, identifying the vast majority of the images that are children and misclassifying adults is preferable for a human-reviewed system, whereas an automated scan which reports the likelihood of child pornography being present in several thousand images would use a much higher precision and lower recall point.

In terms of simple accuracy, the non-adjusted performance resulted in 82% accuracy. Is should be noted that this is not substantially higher than the accuracy identified in Experiment 1 for single ratio measurements. Because the Naïve Bayesian classifier assumes dimensional independence, and because all facial anthropometric measures are highly correlated, the Naïve Bayesian performance is only slightly better than individual ratio performance.

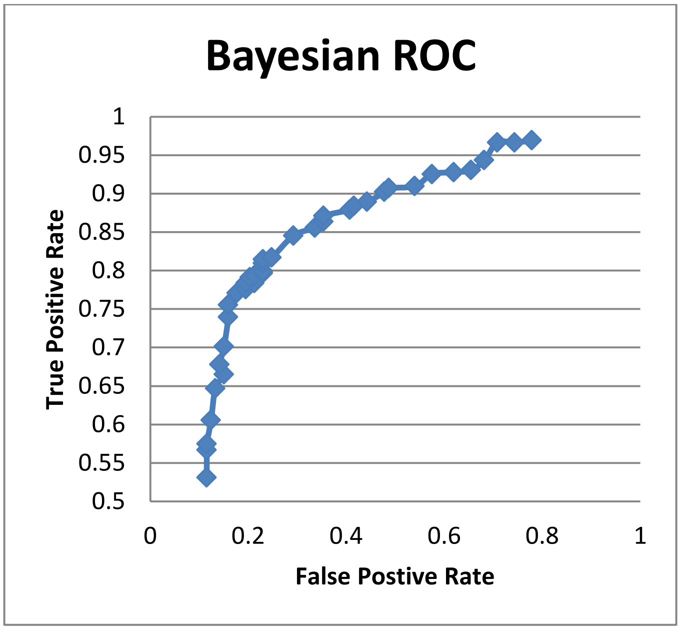

Figure 3 – Receiver Operating Characteristics for Naïve Bayesian Classifier

3.5 Experiment 4: Age demarcation

As previously noted, one limitation of age detection algorithms using body characteristics is that they are not detecting age but detecting developmental maturity, which is a Gaussian distribution over a range of ages and not fixed at a given age. While different characteristics peak at different points for each gender, facial structure is generally developmentally complete before 18 years of age. As such, it would be expected that moving the detector from a minor detector to a pubescence detector would result in a higher recall and precision.

For this experiment, we used the Naïve Bayesian classifier noted in Experiment 3 above. The classifier was trained multiple times using the ages 12 and 18 (representing the expected age range of the completion of puberty). The results are shown in table 4 below.

Age	Precision	Recall
12	0.79	0.88

13	0.85	0.88
14	*0.88*	*0.88*
15	0.89	0.86
16	0.90	0.84
17	0.90	0.83
18	0.92	0.82

Table 4 – Best delineation age

The best age for delineation was found to be 14 years old, representing the probabilistic end of puberty across genders. For the purposes of child pornography detection, the bulk of prosecutions occur for distribution of images under 14, which would indicate a detector set younger than 18 might be warranted. When the age is set to 14, 88% of the individuals in the images identified are actually under 14.

Using an age of 14, further tuning as performed in Experiment 3 above could be conducted depending on the application. Additionally, a separation of genders would likely result in a higher precision and recall, however limits in current gender detection algorithms would result in an overall lower performance due to gender misclassification error.

4 Discussion

The FGNet corpus provides a baseline for testing age detection methods, however a larger and more ethnically diverse database would provide greater assurances that detection techniques are cross-cultural. Additionally, there is no consistent measure in use for testing age detection algorithms. The use of MAE appears to be an effective measure, but many papers attempt to create age-range buckets and thus measurements are based on bucket size in addition to algorithmic effectiveness.

In addition to a larger dataset, this experiment grouped both male and female faces together. While separating male and female into separate ratio groups will likely increase performance, the multiplicative error expected from a pre-classifier as to gender would likely reduce overall effectiveness, however this has not been proven experimentally.

Finally, the selection of key facial points on which to base ratios is not a single-factor (accuracy) choice. The ability to accurately identify ratio endpoints in an automated fashion is a key factor – the eyes, the nose, the center of the mouth and the extremes of the lower jaw are likely candidates for ease of identification.

Because of the similarity in accuracy of the highest f-score ratios and the results of the Bayesian classifier, the use of other classifier techniques may yield higher results.

Overall, the use of anthropometric ratios for the detection of child pornography is likely to be effective as part of a cascade classifier or for discrimination in ranking images for a human examiner. However, the technique identified (or any of the techniques surveyed) are not applicable to the detection of child pornography amongst a large corpus of legitimate images due to the base rate fallacy.

5 Conclusions

The use of facial anthropometric ratios has been shown to provide a more accurate age classification than prior techniques, including cardioidal strain and manifold models. More specifically, the anthropometric ratio technique resulted in a leading classification performance for minor detection based on facial characteristics.

Additionally, brute-force tabulation previously unavailable has allowed for the identification of the key ratios associated with facial growth in minors. Previously, the best ratios for estimating age were based on measurements from a small number of subjects and were not comprehensive. For real-world detection tasks as well as medical evaluations a broader understanding of key ratios is critical knowledge.

For the purposes of age detection, it was shown that facial techniques could be more accurately characterized as developmental. This has broad implications for legal use of automated age detection (and manual detection methods).

Finally, a classifier based on anthropometric ratios was shown to be cost efficient and effective in identifying minors in a manner that can be reproduced and manually calculated if necessary. Because of the speed of classification when using a small number of ratios, this technique is applicable to real-time and high-volume recognition tasks.

References

[1] JM Tanner and Peter S Davies, Clinical longitudinal standards for height and height velocity for North American children, Sep. 01, 1985.

[2] Yoko Kubota. (2008, May) Wrinkles could be key to buying cigarettes in Japan. [Online]. http://uk.reuters.com/article/idUKT22491320080512?feedType=nl&feedName=uktechnology

[3] D H Enlow and J BOSTWICK III, "Handbook of facial growth," *Plastic and Reconstructive Surgery*, vol. 59, p. 116, 1977.

[4] JB Pittenger, "Shaw. RE (1975a). Aging faces as viscal elastic events: Implications for a theory of non-rigid shape perception," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 1, pp. 374--382.

[5] L S Mark and J T Todd, "The perception of growth in three dimensions.," *Perception & Psychophysics*, vol. 33, pp. 193--196, 1983.

[6] J T Todd and L S Mark, "Issues related to the prediction of craniofacial growth* 1," *American Journal of Orthodontics*, vol. 79, pp. 63--80, 1981.

[7] N Ramanathan and R Chellappa, "Face verification across age progression," , vol. 2, 2005, pp. 462-469 vol. 2.

[8] Xin Geng, Zhi-Hua Zhou, Yu Zhang, Gang Li, and Honghua Dai, "Learning from facial aging patterns for automatic age estimation," , 2006, pp. 307--316.

[9] G Guo, Y Fu, C Dyer, and T S Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Transactions on Image Processing*, vol. 17, pp. 1178--1188, 2008.

[10] Z Yang and H Ai, "Demographic classification with local binary patterns".

[11] Yun Fu, Guodong Guo, and Thomas S Huang, "Age Synthesis and Estimation via Faces: A Survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 32, pp. 1955 -1976, nov. 2010.

[12] Young H Kwon, "Age Classification from Facial Images," , 1999, pp. 762--767.

[13] Leslie G Farkas, *Anthropometry of the Head and Face*.: Raven Press, 1994.

[14] Face and Gesture Recognition Consortium. FG-NET Aging Database. [Online]. http://www.fgnet.rsunit.com/

[15] A Lanitis, C Draganova, and C Christodoulou, "Comparing different classifiers for automatic age estimation," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 34, pp. 621 -628, February 2004.

[16] C P Edwards, "The age group labels and categories of preschool children," *Child Development*, vol. 55, pp. 440--452, 1984.

[17] J O Britton and J H Britton, "Discrimination of age by preschool children," *Journal of Gerontology*, vol. 24, pp. 457--460, 1969.

[18] Narayanan Ramanathan, Rama Chellappa, and Soma Biswas, "Computational methods for modeling facial aging: A survey," *Journal of Visual Languages & Computing*, vol. 20, pp. 131 - 144, 2009.

[19] Haibin Ling, S Soatto, N Ramanathan, and D W Jacobs, , 2007, pp. 1 -8.