Age Estimation Using Support Vector Machine–Sequential Minimal Optimization

Julianson Berueco, Kim Lopena, Arby Moay, Mehdi Salemiseresht, and Chuchi Montenegro
Department of Computer Science, College of Computer Studies, Silliman University, Dumaguete City, Philippines
Email: chuchi.montenegro@gmail.com

***Abstract*—This paper investigates the use of SVM-SMO algorithms in estimating the age of a person through the evaluation of its facial features on both front and side-view face orientation. Stephen-Harris algorithm, SURF, and Minimum Eigenvalue Feature Detection algorithms were also used for feature extraction. During experiments, training sets composed on 44 front view images and 44 side view images were used to train the network. Testing was performed to 140 front view images and 44 side view images. Result of the experiment shows age recognition of 53.85% for front view images and 14.3% for side view images.**

***Index Terms*—age determination, image processing, neural networks**

I. Introduction

The availability of robust face recognition algorithms brings vast studies in the area of image processing to perform expression recognition, smile detection, identity recognition, and much more. One of the more unexplored areas in face study is age determination. Not much study has been done in determining an age of a person through image processing and for those studies age determination is only tested on front-view images. Normally, to determine someone's age people usually turn to looking at a person's wrinkling face, similar to the study of [1], but it poses challenges for some people who have wrinkles due to frequent smoking, drinking, overexposure to sun, and sleep deprivation, among others. The other method to determine age is by the analysis of a selected set of facial feature points, for instance the structure of facial bones, similar to the study of [2] on how the mandible continues to enlarge in the course of life. Support Vector Machine-Sequential Minimal Optimization has been used in age classification and has proven to have good accuracy but has been applied to a wide range of age groups. In the study of [3], they used what they determined to be the optimized facial feature points for the facial measurements in classifying age.

The purpose of this study is to be able to determine a person's age by analyzing a set of facial feature points from an image. The method is divided into two parts: (1) face detection and facial feature identification, and (2) age determination. For the first part, the Viola-Jones Object Detection framework was used to detect if an image contains a face and to identify the facial feature points required. The second part measures the distances and angles between the selected set of facial feature points whose results was fed to SVM-SMO classifier for age determination.

II. Scope and Limitation

The study covers multiple human faces in an image regardless of size or distance and orientation (front or side-view). The image can also be either a colored or in black and white. It should also be able to estimate the age regardless of expression or if the subject of the image has wrinkles because of smoking, drinking, sleep deprivation, overexposure to the sun and other external factors that contribute to them "looking old." For ease, images of Filipino faces were used for the study.

The study did not cover faces that are obscured with sunglasses/eyeglasses, masks, tattoos or any foreign objects that covers the feature points used by the application. The images were not blurry or fuzzy and the luminance levels were normal or where the face is recognizable. It also did not cover images whose subject's face has been altered by cosmetic surgery, injury, illness or scars.

III. Methods

The overall process used in this study is reflected in Fig. 1. Training images were initially fed to the model before testing of other images was performed. In both scenarios, all images passed through pre-processing, feature extraction, and classification components.

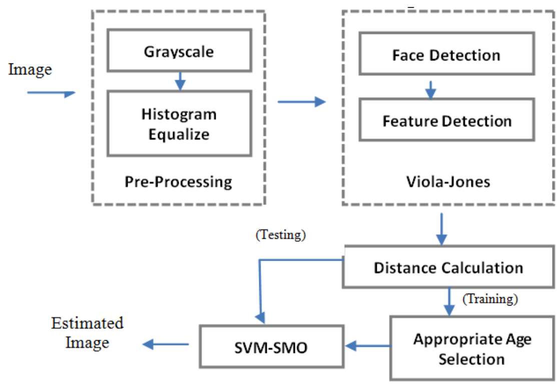

Figure 1. System block diagram

Manuscript received June 25, 2014; revised November 19, 2014.

doi: 10.12720/joig.2.2.145-150

A. Image Acquisition

All images used in the training and testing phase of this research were captured using a typical digital camera having a resolution of 1280×760px. Image acquisition was also done in a controlled environment, that is, proper illumination and lighting is observed, and with a light colored background. For the grayscaling of several images, it was done using the grayscale option in Photoshop and as well in Photoscape.

B. Face Detection for Front and Side View

Detection of the face region for both front and side view faces is described below.

1) Front view viola jones

In this study, the Vision toolbox in MATLAB was used to detect if an image contains faces in front-view. The toolbox uses the Viola-Jones Object Detection Framework as a method for detection.

2) Side view training using cascade training GUI

The Cascade Training GUI is an interactive GUI for managing the selection and positioning of rectangular ROIs in a list of images, and for specifying ground truth for training algorithms. In this study, it was used to train the detectors needed for the side view face detection. Each side and features, ears, chin and eyes were trained separately for a higher accuracy detection rate. Each training sets were trained differently in order to meet the programs requirements. The GUI consisted of stages that have to be manually adjusted to minimize false detections and each set was specified to have the feature type set to LBP (local binary patterns).

a) Side view

Training of the side views had to cover all the features that were used for feature extraction. The ROIs start just above the eyebrows, below the chin and behind the ears. The false alarm rate was set to 0.1, the numbers of stages were set to eight, and the negative samples were set to 5. Manually inputting the object training to 32×34 increased the detection rate. Fig. 2 shows the selection for side view face detection training.

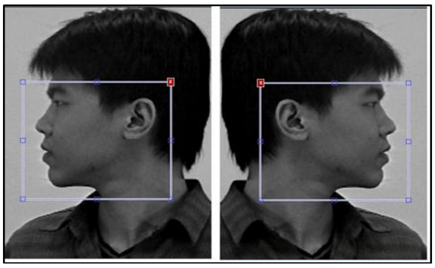

Figure 2. Region of interest selection

b) Ear

The training of the ears had the same concept of the other detectors, but balancing the stages and negative samples was difficult due to the images used for training. The images had to be manually selected from their directories; the pattern of the ear can be easily obscured by hair and would affect the training data. Earrings, specifically studs, were not much of a factor Fig. 3, shows the selection of side view ear detection training.

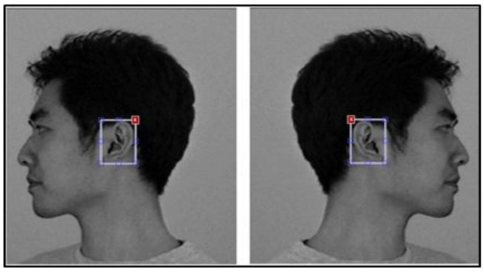

Figure 3. Ears region of interest

c) Chin

The training of the chin was quite an ordeal, same with the eye and the ear. A similar procedure with the ears was used for the chin, in which we selected the facial feature using the cascade training GUI by manually inputting the ROI and forming a rectangular over the chin, starting from its base and going near the lips, but forming the smallest possible rectangle on both sides as shown Fig. 4.

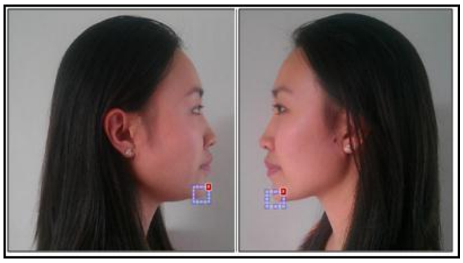

Figure 4. Chin region of interest

During training, several images up to 200+ for each side were conducted for the purpose to increase detection. Numerous changes on the settings were done to get a close 100% accuracy. Most of the changes were in the per – stage false alarm rate and per – stage true positive rate. Cascade stages were lowered to 7 stages to increase accuracy and having the negative samples factor to 4 in order for it to compare several hundreds of negative samples and making the object training size equal to 23×26 to specify a small size will escalate precision.

d) Eye

Figure 5. Eye region of interest

Cascade training GUI for training side-view eye detector to detect eyes for both sides was used. In cascade GUI, 227 positive picture for training left-side eyes and 276 for right side eyes was used, selecting the best ROI for each side-eye as shown in Fig. 5.

For train cascade detector, 2049 negative picture was added. Final training settings were set at false alarm rate equal to 0.00150, true positive rate at 0.995, number of cascade stages at 7 and negative samples factor of 8.

Linear binary pattern facial feature extractor was used with an image size of 1280 by 720 pixels. The accuracy is 95% for both sides and false detector at less than 3%.

C. Feature Detection

1) Front-view Viola Jones

In a similar fashion to the studies of [4] and [5], the researchers needed to be able to detect the left and right eyeballs, and the mouth of a front view face. The researchers also incorporated elements from the study of [4] into the research.

For the progression of the research, the Vision Toolbox was used for detecting the different facial feature points on the front-viewing face. Similar to face detection, it also uses the Viola-Jones Object Detection Framework as its method of detection.

First, the 'EyePairBig' model was used to search if the face region contains both the left and right eyes. If ever this detection fails, image processing for this region is aborted.

Second, the 'LeftEye' model was used to search the face region for the left eye. To eliminate the various false detections that this model may cause, the paper compares each of the detected LeftEye regions' point coordinates with the coordinate of the 'EyePairBig' region. The nearest 'LeftEye' region is selected as the 'LeftEye' region, the rest are deleted.

Third, the 'RightEye' model was used to search the face region for the right eye. It followed a similar process with the 'LeftEye', except it compared each of the detected 'RightEye' regions' point coordinates plus the length of the region with the 'EyePairBig' region's point coordinate plus the length of that region.

Then, the 'Mouth' model was used to search the face region for the mouth.

2) Side-view SURF feature and Harris-Stephens algorithm

a) Ear

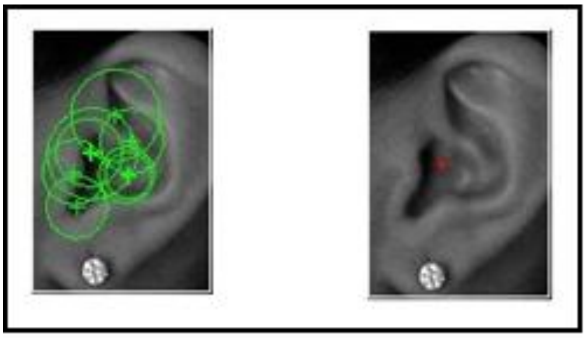

Figure 6. Strongest points using SURF

The corner detector used to extract the feature from the ear is SURF (Speed Up Robust Feature). After the ear was detected, the region where the ear is located was cropped and converted into grayscale, as needed for all corner detectors. The corners were then detected by using detectSURFFeatures() method and the features were extracted with using extractFeatures() method of the toolbox. The strongest points were located by computing the mean that is used to plot the location of the ear as shown in Fig. 6.

b) Eye

The corner detector used to extract the feature from the eye was the Harris-Stephens algorithm. After the eye was detected, the region where the eye is located was cropped and converted into grayscale. The corners were then detected by using detectHarrisFeatures() method and the strongest features are extracted using corners.selectStrongest() method of the toolbox.

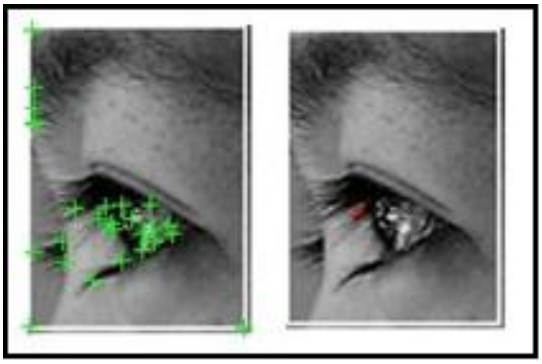

Figure 7. Strongest points of eye

The strongest points were located by computing the mean that is used to plot the location of the eye as shown in Fig. 7.

c) Chin

The corner detector used to extract the feature from the chin was the Harris-Stephens algorithm. The procedure is similar to the eye detection.

D. Training of SVM-SMO

To implement the SVM-SMO, the WEKA (Waikato Environment for Knowledge Analysis) data mining software was used. WEKA requires the creation of an ARFF file in order for it to perform its calculations. The ARFF file contains the data acquired from the previous steps, specifically, the measurements and angles of the facial feature points. Before the output model was created, a kernel function was first selected. The researchers used the default kernel function in WEKA, which is the Polynomial kernel with an exponent of 1.

IV. Experiments

A. Front-View Age Classification

Table I shows the number of subjects gathered for testing. During testing, 98.57% (138) were successful during the face detection stage. In the feature extraction stage, only 78 out of the 138 (56.52%) passed the detection, though most of them incurred some slight errors. For age classification stage, only 42 out of the 78 (53.85%) managed to correctly estimate the age category. Overall, the system only managed to correctly classify 42 out of the 140 total subjects, bringing its accuracy to only 30%.

TABLE I. AGE TEST DATA FRONT-VIEW

Age Category	Subjects	Face Detection	Feature Extraction	Age Classification
-10	5	5/5	5/5	1/5
11-15	13	13/13	7/13	1/7
16-20	59	59/59	44/59	39/44
21-25	24	24/24	11/25	0/11
26-30	6	6/6	0/6	0/0
31-35	2	1/2	0/1	0/0
36-40	7	7/7	2/7	0/2
41-45	10	10/10	3/10	0/3
46-50	6	5/6	2/5	0/2
51+	8	8/8	2/8	0/2
Total	140	138/140	78/138	42/78
Percentage		98.57%	56.52%	53.85%

Fig. 8 shows the different types of detections that the system does for the colored images. Fig. 8(1) shows a perfect detection and feature extraction, with the points exactly where they should be.

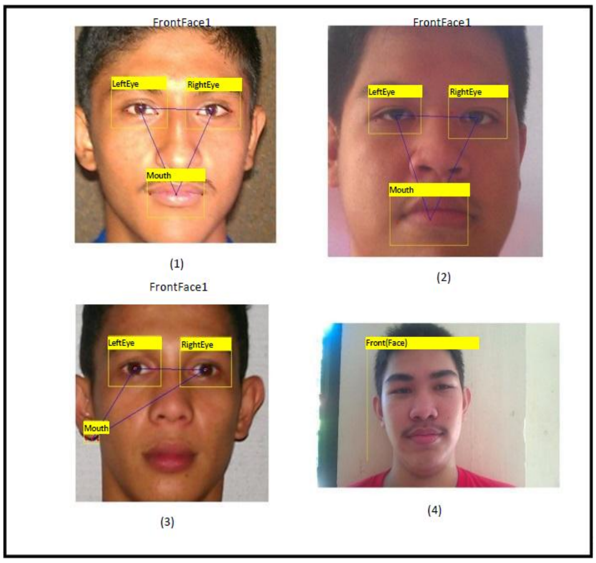

Figure 8. Front-view selection

Fig. 8(2) shows detection and feature extraction with slight errors. We still considered this during testing as it still gives out a prediction. Fig. 8(3) shows detection and feature extraction with a major error, and while the system still also gives a prediction for this, we regarded the image as being not detected since the mouth was not boxed. Fig. 8(4) shows detection and a failure during the primary eye detection. It will show a prediction since if a detector fails, it will stop processing that particular face image.

B. Side-View Age Classification

Eighty-eight (88) test subjects were used for side-view age classification (Table II), out of which, only 85 (96.6%) subjects had their face detected. Out of the 85, 14 (16.5%) had their features extracted, and out of the 14 subjects, 2 (14.3%) were correctly classified. Although the age classifier had an accuracy of 84.1% on its training set, the testing set only managed to have 2.27% overall.

TABLE II. AGE DATA SET SIDE-VIEW

Testing Control Summary				
Age Category	No. of Subjects	Face Detection	Feature Extraction	Correctly Classified
-10	10	9/10	1/9	0/1
11-15	10	10/10	0/10	0/0
16-20	10	10/10	8/10	2/8
21-25	10	10/10	5/10	0/5
26-30	6	6/6	0/6	0/0
31-35	5	3/5	0/3	0/0
36-40	9	9/9	0/9	0/0
41-45	10	10/10	0/10	0/0
46-50	8	8/8	0/8	0/0
51+	10	10/10	0/10	0/8
Total	88	85/88	14/85	2/14
Percentage		96.6%	16.5%	14.3%

In Fig. 9, it shows the image with the face detected, features detected and feature points extracted. The images used in the training set are manually chosen by testing all images in the database; the images with the most accurate feature points are then recorded and classified by SVM-SMO

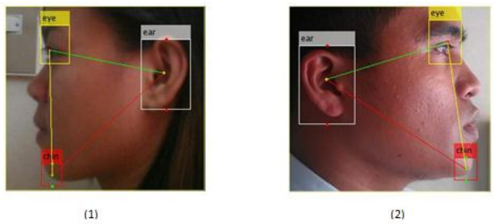

Figure 9. Side view training set

In Fig. 10, it shows the images with the face detected, but the required features were not. From our observations, the factors that hinder the feature detection: (1) the position of the head during side view, chinned up or down, (2) the position of the face in the image itself, the bottom of the chin and top of the head needs to be spaced, the minimum distance of the subject to the camera should be 2 ft. away, (3) the blurriness of the image and (4) after detecting the features, the values given by the edge detectors are not constantly in the same location

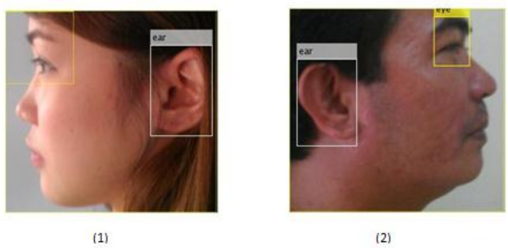

Figure 10. Side-view false detection

1) Group image

Samples of grouped images are also tested (Fig. 11.1) taken from the original image with a dimensions of 1280×720px. Observations from the testing shows: (1) the distance of faces can't detect the features well; (2) the system in detecting features wasn't trained on images of the faces (3) lastly the distance of the faces are quite far where the current resolution setting can't make out the features of the faces.

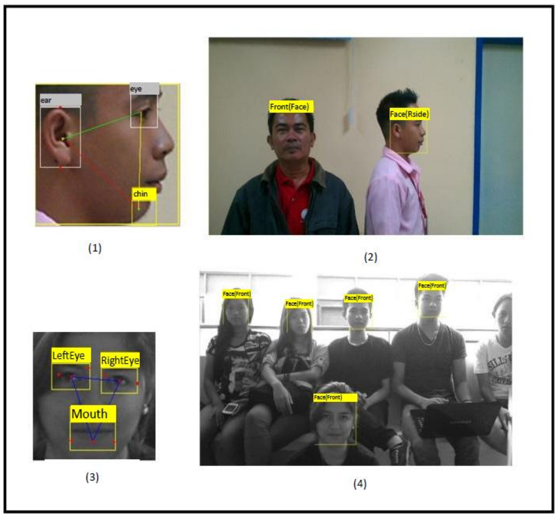

Figure 11. Group image pic

The face detection wasn't a problem at all for the system as it generated 53/57 (92.98%) face detection (Table III). The drawback for the distant images is the detectors for the facial features weren't trained for it, hence, the low detection of 17/53 (32.07%) which coincides with the classification of ages.

TABLE III. GROUP IMAGE SUMMARY

Front View Subjects	Side View Subjects	Face Detection	Feature Extraction	Correctly Classified
2	0	2/2	0/2	0/0
1	1	2/2	½	0/1
3	0	3/3	2/3	2/2
2	1	3/3	1/3	0/2
3	0	3/3	2/3	0/2
2	1	3/3	2/3	0/2
2	0	2/2	2/2	2/2
0	2	2/2	0/2	0/0
14	1	13/15	1/13	1/1
4	0	4/4	1/4	1/1
7	0	7/8	2/7	1/2
0	2	2/2	0/2	0/0
0	2	2/2	1/2	0/1
6	0	5/6	2/5	2/2
Total	54	53/57	17/53	9/17
Percentage		92.98%	32.07%	52.9%

V. CONCLUSION AND RECOMMENDATION

For the duration on this study, the researchers acquired images by taking pictures of random Filipino people of different ages using an 8.0 megapixel camera cellphone camera having resolution setting of 1280 × 728px, with a minimum distance of 2 ft. The images then went through the Viola-Jones Object Detection Framework, an algorithm used by the researchers for both face and feature detection. After that, the detected images went through SURF, Harris-Stephens, and Minimum Eigenvalue feature detection algorithms for feature extraction. Following the extraction the feature points were measured and calculated, and the data went through the SVM-SMO algorithm for age classification using WEKA.

The first scenario yielded only 53.85% for the front view and 14.3% for the side view but had zero percent accuracy on ages beyond 16-20 age categories. This is likely because the images used for training was lacking and/or zooming in for a standard sized image lost some pixels. The accomplishment of a low-percentage accuracy stems from the system's inability to accurately extract the feature points needed, and properly measure the distances and angles between these feature points. The measurement of these feature points were heavily dependent on the pixel size that any manipulation of the images, such as cropping and resizing, may cause it to lose some pixels and, therefore, lose the reliability of the measurements, causing further the SVM-SMO to be confused with the data, thereby causing it to predict inaccurate age estimates.

REFERENCES

[1] Y. H. Kwon and N. D. V. Lobo, "Age classification from facial images," *Computer Vision and Image Understanding*, vol. 74, no. 1, pp. 1-21, 1999.

[2] R. Shaw, E. Katzel, P. Koltz, D. Kahn, J. Girotto, and H. Langstein, "Aging of the mandible and its aesthetic implications: A three dimensional CT study," *AAPS Annual Meeting*, 2009.

[3] Z. Alom, S. Islam, N. Kim, J. H. Park, and M. L. Piao, "Optimized facial features-based age classification," *World Academy of Science, Engineering and Technology*, no. 63, pp. 448-452, 2012.

[4] K. C. Fan and C. Lin, "Triangle-based approach to the detection of human face," *Pattern Recognition Journal Society*, vol. 34, pp. 1271-1284, 2001.

[5] A. R. Chowdhury, R. Jana, and H. Pal, "Age group estimation using face angle," *Journal of Computer Engineering*, vol. 7, no. 5, pp. 35-39, 2012.

Julianson Beureco was born in Cavite, Philippines in 1991. He attained his BS Computer Science Degree in Silliman University, Dumaguete. He is currently working in the R&D Department as an Application Developer in E-Hors, Dumaguete City, Philippines. His current ambition is to attain a Master's Degree in Marine Biology.

Kim Lopena is a Filipino born in Dumaguete City in 1989. He recently attained his B.S in Computer Science degree from Silliman University, Dumaguete City in 2014. He is currently interning for Rentah Inc., Brooklyn, New York. His current endeavor is being an IT/Computer Systems (Back–End Developer).

Arby Moay is a Filipino born in Dapitan City in 1993. He was a scholar in Philippine Science High School - CMC, and attained a BS Computer Science degree in Silliman University, Dumaguete City in 2014. He is currently working as a R&D Engineer I for NetworkLabs Nokia, TechnoHub, Quezon City. He currently dreams of becoming a Software Architect for NWL.

Mehdi Salemiseresht was born in Tehran, Iran, in 1982. He recently attained his B.S in Computer Science degree from Silliman University, Dumaguete City in 2014. Currently working for NetworkLabs Nokia, Manila, Philippines.

Chuchi Montenegro was born in Dapitan City, Philippines, in 1971. She received her B.S. in Computer Engineering degree from Cebu Institute of Technology – University, Cebu City in 1992 and Master in Computer Science from the same university in 2009. She is an assistant professor of the College of Computer Studies, Silliman University, Dumaguete City. Her research interest is in the field of neural networks, signal processing, and speech recognition.

