Age Estimation on Human Face Image Using Support Vector Regression and Texture-Based Features

Jesy S Amelia, Wahyono*
Department of Computer Science and Electronics
Universitas Gadjah Mada, Yogyakarta
Indonesia

Abstract—This paper proposed a framework for estimating human age using facial features. These features exploit facial region information, such as wrinkles on the eye and cheek, which are then represented as a texture-based feature. Our proposed framework has several steps: preprocessing, feature extraction, and age estimation. In this research, three feature extraction methods and their combination are performed, such as Local Binary Pattern (LBP), Local Phrase Quantization (LPQ), and Binarized Statistical Image Feature (BSIF). After extracting the feature, Principle Component Analysis (PCA) was performed to reduce the feature size. Finally, the Support Vector Regression (SVR) method was used to predict age. In evaluation, the estimation error will be based on mean average error (MAE). In the experiment, we utilized the well-known public dataset, face- age.zip, and UTK Face datasets, containing 15,202 facial image data. The data were divided into the training of 12,162 images and the testing of 3,040 images. Our experiments found that combining BSIF and LPQ with PCA achieved the lowest MAE of 9.766 and 9.754. The results show that the texture-based feature could be utilized for estimating the age on facial image.

Keywords—Age estimation; LBP; BSIF; LBQ; MAE; PCA; preprocessing; feature extraction; Support Vector Regression (SVR)

I. INTRODUCTION

The face is a part of the human body that has an important role. Characteristics of the face such as eyes, nose, mouth, eyebrows, and wrinkles or aging provide much information. One of them is age estimation. It is possible to limit that under age to accessing content (related to violence) in the web browser through age estimation. Shopping centers can use age estimation to determine advertising strategies based on the age of visitors and provide recommendations. It is necessary to have a method used to detect the observer’s age. We must have an extensive collection of facial images that will be used in age detection and pay attention to the face, such as wrinkles or aging that will be used in the age detection process to produce better estimation.

Generally, the age estimation process has several steps: preprocessing, feature extraction, and age estimation. The first is preprocessing (resizing and grayscale image) to be uniform. Then, the feature extraction step is to find the unique facial features in the age estimation process. The features were the Binarized Statistical Image Feature (BSIF), Local Binary Pattern (LBP), and Local Phrase Quantization (LPQ) methods and their combination: BSIF + LBP, BSIF + LPQ, and LBP+LPQ. In addition, Support Vector Regression (SVR) was used in the age estimation stage. SVR is a machine learning algorithm that provides reliable performance in predicting time series data and can classify data that cannot be separated linearly [1].

Several previous studies have been proposed for age estimation, including Ingle et al. [2] and research by Thukral et al. [3] and analysis by Khunteta, Ajay, et al. [4]. However, these three studies only predict a group of ages (e.g., categories 25, 35, 45, 55, 65, 65, and 75 years old), not the age. For this reason, in this study, age estimation was conducted not based on age group but on exactly one age estimation value. Furthermore, the texture of the face is one of the essential features in determining age. This study will detect age on the image containing a single face. The age estimation is carried out only at 1-70 years old due to limited data on ages over 70.

Our research aims to find the best feature extraction method based on texture in detecting age with the lowest Mean Absolute Error (MAE) level from the facial images. Several texture-based feature extraction methods obtain good accuracies, such as LBP, BSIF, and LPQ. The BSIF and LPQ methods provide the best results in the age estimation process with an MAE of 3.58 using the MORPH II dataset and an MAE of 9.07 with the same extraction feature method (BSIF+LPQ) using a Subset of LFW [5]. Meanwhile, LBP is one of the best texture description methods for texture classification, face detection, face recognition, gender classification, and age estimation [6]. In addition, LPQ is proven robust for blur images and is considered to show better performance in texture classification, and the combination method gives better performance results [7].

The paper will be divided into several sections. The second section will explain in more detail how the proposed framework has been applied. The third section will discuss the result of the experiment. Then, the fourth section will conclude our work.

II. PROPOSED METHOD

Age estimation is one of the most challenging and crucial issues in utilizing the facial area to produce helpful information. Age estimation can be used in solving scientific problems as well as in subproblems of facial recognition. The prediction of age is formed because the human brain cannot predict age correctly, so it is necessary to have a system that...
can predict age correctly [8]. Many factors can be used in predicting the age of the facial area, including aging, wrinkles on the face, wrinkles in the eye, mouth area, and other features in the facial area. The estimated age itself can be detected using two methods. These methods are classification and regression methods [9]. Age estimation using classification can only estimate based on age range, while age extraction using the regression method can estimate age with exactly one output value of age. It is the difference in age detection using classification and regression methods, as shown in Fig. 1.

Our proposed framework was divided into several significant steps: image preprocessing, feature extraction, and age estimation.

A. Preprocessing

Preprocessing is an initial process that aims to improve an image. It is for eliminating noise in the image. In preprocessing stage, we also uniform the image in color and size [10]. The purpose of this process is to make it easier to perform feature extraction at the next stage and to create an age estimation model. This research has two preprocessing steps: resizing and gray scaling.

1) Resizing: Since the dataset contains images with different sizes, the image should be resized to a specific size. The resizing process is used to change the size of the image to either enlarge or reduce size. The resizing function is carried out using the interpolation method. The following is a calculation of the resizing process using the linear interpolation formula:

\[
\frac{y - y_0}{x - x_0} = \frac{y_1 - y_0}{x_1 - x_0}
\]  

where \((x_0, y_0)\) is two points of initial coordinates, \((x_1, y_1)\) is destination coordinates, and \((x, y)\) is coordinate prediction.

2) Grayscaling: Grayscale is one of the preprocessing steps to change the image from an RGB image to a gray color (grayscale). RGB is an image with three combination colors: Red, Green, and Blue (RGB). The value of RGB is represented in three dimensions XYZ, consisting of lightness, chroma, and hue. The process aims to convert the value of RGB (24-bit) into grayscale value (8-bit) [11]. A grayscale image is an image that has only one value for each pixel. This value is used to indicate the level of intensity. The three colors in the grayscale process are white, black, and gray. The following is a formula for the grayscale process:

\[
W = 0.299R + 0.5870G + 0.1140B
\]  

where W, R, G, and B are grayscale, red, green, and blue values of the image, respectively. Fig. 2 illustrates how the RGB image was converted into a grayscale image.

B. Feature Extraction

Feature extraction is a process used to generate unique feature vectors from facial images that will be used to create age estimation models. This feature extraction stage uses three different texture-based methods, including Local Binary Pattern (LBP), Local Phrase Quantization (LPQ), and Binarized Statistical Image Features (BSIF), and their combination, BSIF+LBP, BSIF+LPQ, LPQ+LBP.

1) Local Binary Pattern (LBP): LBP gives the best results in various applications, especially for texture classification, segmentation, face detection, face recognition, gender classification, and age estimation [12]. The LBP method is grayscale invariant and can be combined easily with simple contrast by calculating each gray level of each pixel [13]. LBP considered each pixel has a code. It is called Local Binary Pattern code or LBP code. It works by comparing the value of the center image (pixels) with each neighborhood which amounts to 8 pixels, and thresholding of 3x3 neighborhood [14] divided into eight blocks. The following is a simple example of calculation using LBP at 3x3 pixels:

Given a pixel at \((x_c, y_c)\), the resulting LBP can be expressed in decimal form as follow:

\[
LBP = \sum_{p=0}^{7} S(H_p - H_c)2^p
\]

where \(H_c\) is a central pixel, \(H_p\) start from \((p = 0, 1, \ldots, 7)\), it is neighborhood from \(H_c\), and \(S\) is a function of thresholding, which is defined as:

\[
S(H_p - H_c) = \begin{cases} 1 & H_p \geq H_c \\ 0 & H_p < H_c \end{cases}
\]

After comparing the center image with its neighborhood, the result will be converted into a binary number. As shown in Fig. 3, a binary number generated from pixel 3x3 is 01010100 = 84. The value will be inserted at the center image. Each value of LBP is formed into 256 histograms which will be used as a texture descriptor.

![Fig. 1. Age estimation using classification in the left side and regression in the right side.](image1)

![Fig. 2. Image conversion sample from RGB to Grayscale.](image2)
2) **Local Phrase Quantization (LPQ):** The LPQ descriptor is used in the texture descriptor and classification for blur images [15]. LPQ method is efficiently used to solve the problem of variation expression in the face verification system. LPQ aims to maintain images in local invariant information with various blurring images. Fig. 5 shows the stages of the LPQ. As can be seen in Fig. 5, the $q(x)$ is defined as the following formula:

$$
q_j(x) = \begin{cases} 
1 & \text{if } q_j \geq 0 \\
0 & \text{if } q_j < 0 
\end{cases}
$$

(5)

Lastly, the LPQ generates a binary number by combining the pixel values obtained from $\text{Re}\{F \times\}$ and $\text{Im}\{F \times\}$. Fig. 5 received a binary number of 89 for the center image.

3) **Binarized Statistical Image Features (BSIF):** Inspired by LBP and LPQ, Kannala et al. [16] proposed a new local descriptor called BSIF (binarized Statistical Image features). The basis vectors of a subspace into which local image patches are linearly projected are obtained from images by using Independent Component Analysis (ICA). The coordinates of each pixel are threshold, and thus a binary code is computed. A value represents the local descriptor of the image intensity patterns in the neighborhood of the considered pixel [17]. The following is the formulation of BSIF:

$$
S_i = \sum_{u,v} W_i(u,v) \ast X(u,v) = W_i^T x
$$

(6)

where $X$ is image with size of $M \times M$, $\text{Filter } W_i$ learning from **Independent Component Analysis (ICA)** by maximizing the statistical independence of $S_i$.

$$
b_i(x) = \begin{cases} 
1 & S_i \geq 0 \\
0 & S_i < 0 
\end{cases}
$$

(7)

where $b_i$ is a binary number obtained from $S_i$. In all our experiments, the BSIF descriptor has been used with filters of size 9x9 and 8-bit.

**C. Principal Component Analysis (PCA)**

Principal Component Analysis (PCA) is a statistical method to distinguish patterns and signal processing. PCA reduces dimensional data, feature extraction, and facial recognition [18]. PCA is a powerful method, especially for high-dimensional data. That technique reduces the dimensions of the dataset by extracting the essential components.

**D. Age Estimation Using Support Vector Regression (SVR)**

Support Vector Regression (SVR) is a supervised learning algorithm to predict discrete values. SVR can solve linear and non-linear problems using kernel functions [19]. The advantage of SVR is that the computational complexity of SVR does not depend on the dimensions of the input space. In addition, SVR has excellent generalization capabilities and high prediction accuracy.

The SVR method is effective for solving problems related to the estimation function. So SVR is the best method for age estimation. SVR is the same as SVM. The idea of SVR is to determine the best fit line. In SVR, the most appropriate fit line is the hyperplane which has the maximum number of points. So, the SVR tries to match the best line within its threshold value. The threshold value is the distance between the hyperplane and the boundary line. So, the regression problem is to determine a function that approximates the mapping from the input domain to real numbers based on the training sample.

**E. K-Fold Cross Validation**

Cross Validation is a testing method in data mining and machine learning where the dataset will be divided into $k$ parts or folds randomly. Cross Validation is used to evaluate the performance of the algorithm. The data will be separated into two subsets: the learning subset (training) and the validation/evaluation subset [20]. First, the algorithm will be trained using a learning subset, and then the model will be validated using a validation subset. The illustration of the cross-validation scheme is shown in Fig. 6.
III. EXPERIMENTAL RESULT

In this section, we first introduce the databases and the protocol evaluation of our experiments. After that, we will describe experimental setups and their results.

A. Dataset

The facial dataset used in this research was downloaded from face-age.zip [12]. The facial image size is 200×200 pixels with a total of 9,778 RGB face images. In this face-age.zip face image, there are facial images with an age range of 1-110 years, but this research uses face images from 1-70 years of age with the female and male gender. First, a selection process is carried out on the facial image, only selecting a face without glasses and a watermark. In addition, we only utilized images with ages between 1-70 years due to data limitation in the face image dataset from face-age.zip for ages above 70 years. From the face-age.zip dataset with the selection process, we choose a total of 3,040 images for our experiment.

Furthermore, the face image dataset was added using the UTK Face dataset. We then selected 12,162 face images from this dataset. The UTK Face dataset is also obtained from downloading through the same source on the face-age.zip dataset, and the total facial image dataset owned is 15,202 images. Finally, the dataset is divided into training and testing data, where the training data is 12,162 face images, and the testing data is 3,040 facial images (20% of the total facial images) which contain facial images 1-70 years old for training and data testing. Table I shows the selected images used in our experiment.

B. Experiment Scenario

The scenario of the system is divided into two main stages: training and testing. The training stage is applied to find the relationship between one face and another and create a model for age estimation. In contrast, the testing stage is performed to detect the estimated age of the face image. Fig. 7 shows the scenario of the experiment for our proposed framework.

In this research, to produce the best age estimation model, experiments were carried out on feature extraction methods based on texture, such as LBP, LPQ, BSIF, and the combination of these methods, BSIF + LBP, BSIF+LPQ, and LPQ + LBP. From the three extraction methods and the combination, it will be seen which method will produce the lowest MAE level in the age estimation process. In addition, there are several experiments: feature reduction extracted using PCA with values of 40, 50, 60, 70, 80, and 90 features. For age estimation using the SVR method.

Since the estimation output will not be in discrete values, for evaluation purposes, we rounded the estimation output using two strategies. The first strategy is rounding the value up from the detected age. For example, the detected ages 32.5, 17.8, and 23.2 will be 33, 18, and 24. The second strategy is rounding the age value down from the detected age. For example, the detected ages 32.5, 17.8, and 23.2 will be 32, 17, and 23.

C. Evaluation Matricts

To measure the performance of our approach, we adopted Mean Absolute Error (MAE). MAE is a method used to evaluate the performance of age estimation. MAE is the average absolute error between the ground truth age and the predicted one [5]. The age estimation model can be measured by using MAE. The following is the formulation of MAE:

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |P_i - G_i|
\]

where \(P_i\) is the estimated age, \(G_i\) is the corresponding ground truth, and \(N\) is the total number of samples.

<table>
<thead>
<tr>
<th>TABLE I. THE SAMPLE IMAGES IN THE DATASET</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Image</td>
</tr>
<tr>
<td>Image</td>
</tr>
</tbody>
</table>
D. Results and Discussion

1) Experiment scenario: In this study, age estimation was carried out with various scenarios of experiments as follows:
   a) Resizing the image: 180×180 pixels, 160×160 pixels, 140×140 pixels, and 120×120 pixels.
   b) Feature Extraction using different methods and the combination (LBP, LPQ, BSIF, BSIF+LBP, BSIF+LPQ, and LPQ+LBP).
   c) LBP Feature extraction using different radius (radius = 1, radius = 2, radius = 3 with same P (total number of neighborhood)).
   d) Using the first strategy and second strategies for the age estimation, as explained in section III.B.
   e) Comparing MAE from extracted features without PCA and with PCA.
   f) Tried several variations of PCA, such as 40, 50, 60, 70, 80, and 90 features and compare MAE.
   g) Find the optimum parameter of BSIF Filter by ICA.
2) Evaluation with different images scale and without PCA: In this paper, experiments were carried out on various image sizes such as 180×180, 160×160, 140×140, and 120×120 with the first strategy and second strategy of age estimation using LBP, LPQ, BSIF, BSIF+LBP, BSIF+LPQ, and LPQ+LBP. Fig. 8 and 9 show the MAE results when applying the proposed framework without using PCA on several image sizes. From these results, it can be concluded that by resizing images to 180×180, 160×160, 140×140 and 120×120 with feature extraction without PCA, the lowest MAE when using the BSIF method with an image size of 180×180 of 10.612 for the first strategy and 10.600 for the second strategy for age estimation. Furthermore, from the results, we could also find that in most scenarios, if the image is resized to a smaller size, the MAE decreases for LBP, LPQ, and BSIF. However, the results are not better compared to the combination of features.

3) Evaluation of LBP using different radius without PCA: In this study, experiments were also carried out on the radius of the LBP. Radius tested are 1, 2, and 3 to obtain the best age detection results with the lowest MAE. As shown in Fig. 10 and 11, it is found that the lowest MAE in age detection using LBP radius 3. Therefore, the LBP with a radius of three will be utilized in all experiments.

4) Optimal parameter evaluation for LPQ: Feature extraction on Face recognition uses LPQ with various window sizes: 5, 7, and 9, and the best results were obtained using the LPQ method with window size 7 [10]. For this reason, this study used a window size of 7.

5) Optimal parameter evaluation for BSIF: Parameters that can affect the BSIF method are texture filters. Research for Face Recognition using texture filters BSIF with a size of 7×7, 9×9, and 11×11 in 8 bits and the best result using 7×7 [10]. Meanwhile, in the other research, the BSIF descriptor has been used with filters of the size of 13×13 and 8 bits for age estimation [5]. So, we will test texture filters of size 7×7, 9×9, and 13×13 with 8 bits. From Tables II and III, it can be found that BSIF Texture Filter 9×9 8 bits has the lowest average MAE. This texture filter will be used in the entire experiment.

<table>
<thead>
<tr>
<th>Image Size</th>
<th>BSIF 7x7</th>
<th>BSIF 9x9</th>
<th>BSIF 13x13</th>
</tr>
</thead>
<tbody>
<tr>
<td>180x180</td>
<td>10.771</td>
<td>10.612</td>
<td>11.015</td>
</tr>
<tr>
<td>160x160</td>
<td>10.997</td>
<td>10.855</td>
<td>10.627</td>
</tr>
<tr>
<td>140x140</td>
<td>10.442</td>
<td>10.612</td>
<td>11.269</td>
</tr>
<tr>
<td>120x120</td>
<td>10.715</td>
<td>10.749</td>
<td>11.063</td>
</tr>
<tr>
<td>Average</td>
<td>10.731</td>
<td>10.707</td>
<td>10.994</td>
</tr>
</tbody>
</table>

Fig. 8. Comparison results of different image size without using PCA in the 1st strategy.

Fig. 9. Comparison results of different image size without using PCA in the 2nd strategy.

Fig. 10. Comparison results of LBP radius without using PCA in the 1st strategy.

Fig. 11. Comparison results of LBP radius without using PCA in the 2nd strategy.
TABLE III. BSIF FILTER COMPARISON WITHOUT PCA IN 2ND STRATEGY

<table>
<thead>
<tr>
<th>Image Size</th>
<th>BSIF 7x7 MAE</th>
<th>BSIF 9x9 MAE</th>
<th>BSIF 13x13 MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>180x180</td>
<td>10.762</td>
<td>10.634</td>
<td>11.004</td>
</tr>
<tr>
<td>160x160</td>
<td>10.982</td>
<td>10.607</td>
<td>11.272</td>
</tr>
<tr>
<td>140x140</td>
<td>10.402</td>
<td>10.714</td>
<td>11.07</td>
</tr>
<tr>
<td>Average</td>
<td>10.716</td>
<td>10.700</td>
<td>10.995</td>
</tr>
</tbody>
</table>

6) Evaluation age estimation based on extraction feature:
We use several variations of PCA feature vector sizes, such as 40, 50, 60, 70, 80, and 90, with various image sizes of 180x180, 160x160, 140x140, and 120x120 pixels with optimal parameters from LBP, LPQ, BSIF, and the combination BSIF+LBP, BSIF+LPQ, and LPQ+LBP. However, based on the result of our experiment, we will only display the results at 160x160 pixels because this image size produces the lowest MAE. Tables IV and V show the comparison results of several feature extraction methods for estimating. Based on these tables, it was found that the lowest MAE is generated using BSIF+LPQ with PCA 70. It achieved an MAE value of 9.766 in the first and 9.754 in the second strategies.

7) Age estimation of public figures: After obtaining the best model from the previous experiment, we then utilized the model for estimating the age of two public figures: Ralph Fiennes and Indro Warkop. We utilized the BSIF+LPQ model with an image size of 160x160. Due to limited data on these two public figures, the PCA was not performed. Fig. 12 and 13 show the age estimation of these public figures at different ages. Our model achieved MAE of 10.428 and 11.142 at the first and second strategies for Ralph Fiennes. On the other hand, the model achieved an MAE of 12.142 and 13.142 in the first and second strategies for Indra Warkop.

TABLE IV. MAE COMPARISON WITH IMAGE SIZE OF 160X160 PIXEL IN 1ST STRATEGY

<table>
<thead>
<tr>
<th>Method</th>
<th>Without PCA MAE</th>
<th>PCA 40 MAE</th>
<th>PCA 50 MAE</th>
<th>PCA 60 MAE</th>
<th>PCA 70 MAE</th>
<th>PCA 80 MAE</th>
<th>PCA 90 MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPQ</td>
<td>11.297</td>
<td>10.986</td>
<td>10.635</td>
<td>10.709</td>
<td>10.95</td>
<td>10.696</td>
<td>10.672</td>
</tr>
<tr>
<td>BSIF+LBP</td>
<td>11.140</td>
<td>10.678</td>
<td>10.692</td>
<td>10.695</td>
<td>10.698</td>
<td>10.482</td>
<td>10.468</td>
</tr>
</tbody>
</table>

TABLE V. MAE COMPARISON WITH IMAGE SIZE OF 160X160 PIXEL IN 2ND STRATEGY

<table>
<thead>
<tr>
<th>Method</th>
<th>Without PCA MAE</th>
<th>PCA 40 MAE</th>
<th>PCA 50 MAE</th>
<th>PCA 60 MAE</th>
<th>PCA 70 MAE</th>
<th>PCA 80 MAE</th>
<th>PCA 90 MAE</th>
</tr>
</thead>
</table>
IV. CONCLUSION

In this research, the age estimation framework using Support Vector Regression (SVR) with texture-based feature extraction has been successfully implemented. The proposed framework can detect the exact age, not based on the age range. Furthermore, a comprehensive experiment has been carried out to determine the optimal model of age estimation, such as the different sizes of the image, utilization of PCA, and applying several feature extractions and their combination. Based on the experiment, we have found several essential findings in constructing of age estimation model. First, image size 160×160 produced the lowest MAE level compared to image size 120×120, 140×140, and 180×180 pixels. Second, utilizing the PCA after feature extraction gains computation time fast and gives the lowest MAE level in age estimation. Third, the optimal parameters for LBP, BSIF, and LPQ are a radius equal to three, a texture filter size of 9×9 8-bits, and a window size set equal to seven. Finally, the optimal model which produces the lowest MAE is the combination of BSIF and LPQ methods with a PCA dimension of 70. The model obtained MAE values of 9.766 and 9.754 for the first and second strategy is 9.754, respectively. However, the model has limitations for estimating the age of facial images from Asian countries since the dataset used for the experiment is mostly from Western countries. Therefore, future research might consider the additional dataset covering various ethnic groups. In addition, we may use another method like ANN or CNN and compare that with this method for age estimation.

ACKNOWLEDGMENT

The research and publication were funded through the Final Assignment Recognition Program (Program Rekognisi Tugas Akhir) Batch II Year 2022, Universitas Gadjah Mada with Assignment Number 633/UN1.P.III/KPT/HUKOR/2022 and 5722/UN1.P.III/Dit-Lit/PT.01.05/2022.

REFERENCES


