Face Gender Classification using Combination of LPQ-Self PCA

Tio Dharmawan[1], Danu Adi Nugroho[2*], Muhammad Arief Hidayat[3]

[1,2,3]Departement of Informatics, Faculty of Computer Science, University of Jember, Indonesia

*corr_author: danuadinugroho65@gmail.com

Abstract - The age factor had a significant impact on human faces, potentially influencing the performance of existing gender classification systems. This research proposed a new method that combined local descriptors such as Local Binary Patterns (LBP) and Local Phase Quantization (LPQ) with Self-Principal Component Analysis (Self-PCA) as a feature extraction technique. The use of Self-PCA was chosen for its ability to address the age factor in human facial images, while also leveraging local descriptors to capture features from these images. The primary focus was to compare the performance of Self-PCA with LPQ+Self-PCA, along with the additional comparison of LBP+Self-PCA, in the task of gender classification using facial images. Euclidean distance served as the classifier, and the evaluation was conducted using the FG-Net and ORL datasets. The combination of LPQ+Self-PCA showed an improvement in accuracy by 57.85% compared to the combination of LBP+Self-PCA, which provided an accuracy of 56.47%. Meanwhile, using Self-PCA alone gave an accuracy of 55.37% on the FG-Net. In contrast, on the ORL dataset, both combinations gave the same accuracy result as Self-PCA, which was 90.14%, for images without blurring.

Keywords: Self-Principal Component Analisys, Local Binary Patterns, Local Phase Quantization, face gender classification

I. INTRODUCTION

Face recognition by computers is more complex and requires more time than recognition performed by humans. Humans can easily and accurately identify a person's gender through their face, voice, behavior, and other factors. In contrast, computers need to process attributes, such as facial images, before they can identify gender. Therefore, specific methods have emerged for feature extraction from facial images.

One of highly popular feature extraction method is Local Binary Pattern (LBP) [1]. Since its introduction, this method has gained significant popularity due to its effectiveness and relatively simple calculation [2]. Given its reputation, LBP has attracted researchers attention and has been applied in various Computer Vision tasks, such as face identification [3]-[5] and texture identification [6]- [8]. In prior study, LBP were applied for gender classification using facial images [9]. The research demonstrated that using LBP for feature extraction in faces can enhance the performance of Support Vector Machine (SVM) in gender classification. Despite its popularity, LBP has a limitation in extracting features from blurry images [7]. To overcome this limitation, Local Phase Quantization (LPQ) was developed [10]. This method changes the feature extraction approach by using Fourier transformation. Another development from LBP is Binarized Statistical Image Features (BSIF) [11], which creates a filter available in the image without manually creating filters like LBP and LPQ. Faced with a numerous of local descriptor options, determining the most suitable method for feature extraction poses a challenge. Consequently, a comprehensive comparison is conducted among various local descriptors, including LBP, LPQ, and BSIF, to perceive their respective performances and applications [7]. The results showed that feature extraction with LPQ outperformed LBP and BSIF.

Meanwhile, recognizing the impact of age on gender classification performance, there is a suggestion to enhance the classification process by employing the Self-Principal Component Analysis (Self-PCA) method [12]. Self-PCA is a development of PCA that involves creating specific eigenfaces for each class to be identified, resulting in more accurate eigenfaces for each class [13]. The research demonstrated that the application of Self-PCA to facial images for gender classification, especially when considering the age factor, gives better performance than traditional PCA, using Euclidean distance measurement as the classifier.

Results from previous studies indicate that the LPQ feature extraction method provides the best results in gender classification and remains consistent even when images experience blur disturbance. If there is an additional age factor in gender classification, the Self-PCA method also proves to be superior to other methods.

II. METHOD

The stages in this research can be illustrated as shown in Fig. 1, starting with the identification and downloading of two datasets used. After that, preprocessing is performed on each dataset, involving conversion to grayscale, face detection, facial region cropping, resizing images to 100x100 pixels, and applying Gaussian Blur. Feature extraction is then conducted using LBP and LPQ, followed by the creation of eigenfaces for each gender. Once eigenfaces are obtained for each gender, the test images are projected onto these eigenfaces, and their Euclidean distances are calculated. Subsequently, the gender classification results for the test images are determined.

A. Dataset Identification

This research was conducted using facial images obtained from publicly available datasets. There are two datasets used in this research, which is FG-Net and ORL. FG-Net, also known as the FG-Net Aging Database, usually used for age estimation and cross-age face recognition [14]. This dataset contains 1002 facial images spanning ages from 0 to 69 years. The ORL dataset, now referred to as Our Database of Faces, features 40 subjects, each with 10 images, totaling 400 images [15].

B. Preprocessing

Analyzing high-resolution images with complex backgrounds can be time-consuming and may reduce the success rate [16]. Therefore, a preprocessing stage is necessary to address this challenge. The preprocessing stage refers to a set of techniques performed to enhance the quality of data [17]. In this study, the preprocessing steps are as follows:

1. Converting to Grayscale
2. Face Detection using Viola-Jones
3. Cropping Face Region
4. Resizing image to 100x100 pixel
5. Adding Gaussian Blur

C. Feature Extraction

In this research, LPQ, LBP, and Self-PCA are used for feature extraction. There are three combinations of feature extraction methods used in the feature extraction stage, as shown in the Table I.

Self-PCA feature extraction is used to create a benchmark for this research. This method directly generates eigenfaces without the extraction of features with other local descriptors. Of the three combinations mentioned above, the research involves training and testing data using three different datasets and variations in Gaussian Blur values, window size, and the number of local samples.

TABLE I
COMBINATION OF FEATURE EXTRACTION

Research	Feature Extraction
Nayak and Indiramma [12]	Self-PCA
Proposed Method	LBP + Self-PCA
Proposed Method	LPQ + Self-PCA

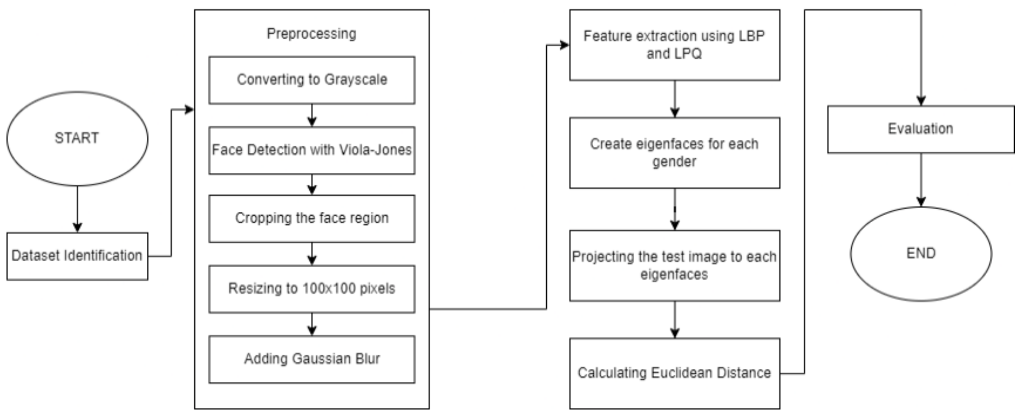

Fig. 1 Stages of research diagram

D. Eigenface Creation

The procedure for generating eigenfaces for male and female genders shown on Fig. 2, begins by transforming an image of size P x P into a column vector of dimensions P^2 x 1. Then, the column vector is subtracted with the average of the transformed vector images. Next, the eigenvalues along with the eigenvectors and covariance matrix C are computed. In the end, two sets of eigenfaces are obtained, namely male eigenfaces and female eigenfaces.

E. Test Image Projection

This step is similar to the creation of eigenfaces shown on Fig. 3. In this step, the test image with pixel dimensions P x P is transformed into a vector of size P^2 x 1. Then, the column vector is subtracted with the average of the transformed vector images. Subsequently, this vector is projected into the eigen space generated from training images of both males and females. This projection enables the system to comprehend the characteristics and patterns of the test image within the context of the eigen space adapted for both genders.

F. Calculating Euclidean Distance

This step is a continuation of the projection step, where classification is performed based on the Euclidean distance obtained from the projection. The smallest Euclidean distance from the test image in the eigen space projected for male images is referred as d_m, while in the eigen space projected for female images, is referred as d_f. The classification process is executed based on the smaller value between d_f and d_m. An example of the classification result is as follows: if d_m is less than d_f, then the image is classified as male.

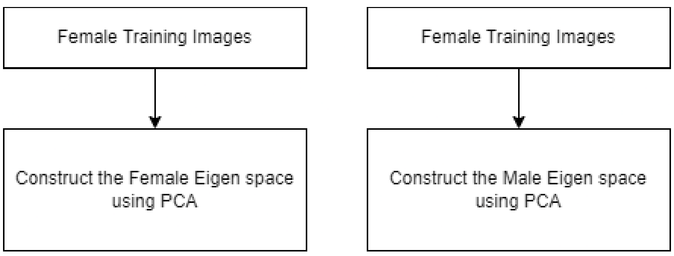

Fig. 2 Eigenface creation steps [12]

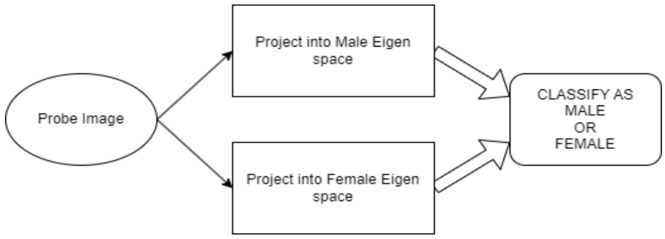

Fig. 3 Projection steps [12]

G. Evaluation

In the last step of research, the model testing results are evaluated using metrics such as Accuracy, Precision, Recall, and F1-Score to assess the performance on the FG-Net and ORL datasets and their respective classes during the validation process. This research employs the weighted average score for evaluation. The calculation algorithm for Precision, Recall, and F1-Score can be seen in (1-3).

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{1}$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{2}$$

$$F1\ Score = \frac{2{*}Precision + Recall}{Precision + Recall} \tag{3}$$

III. RESULT AND DISCUSSION

A. Dataset Identification

The data used in this research consists of the ORL and FG-Net datasets. Both datasets were obtained through access to different repositories. The FG-Net dataset can be accessed through the repository owned by Yanwei Fu, while the ORL dataset is available through the repository managed by AT&T Laboratories Cambridge. The initial amount of data in the ORL dataset is 400, with 40 images representing female gender and 360 images representing male gender. Meanwhile, the FG-Net dataset contains 1002 data, with 598 images representing female gender and 404 images representing male gender. Class distribution on ORL dataset can be seen in Fig. 4.

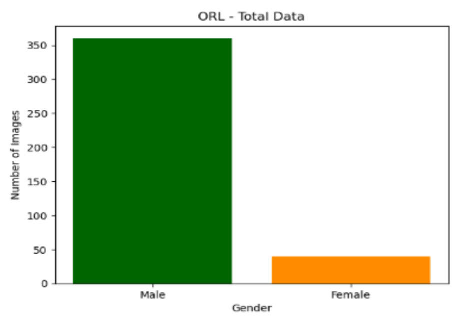

Fig. 4 Total images in ORL dataset

The naming of images in the FG-Net dataset contains several important attributes; for example, 001A02.jpg signifies the subject with ID 001 at the age of 02. Therefore, from the naming, the age of the subject in the image can be determined. In FG-Net, the age range of subjects starts from 0 years to 69 years. In this research, the images used are those with ages between 10 and 60 years. Thus, the total usable data consists of 604 images, comprising 334 images of male gender and 270 images of female gender. Class distribution on FG-Net dataset can be seen in Fig. 5.

B. Preprocessing

1) Converting to Grayscale: The image conversion process from RGB to grayscale is a crucial step in image processing. Besides reducing computational time, this process is necessary to enable the efficient execution of the Viola-Jones algorithm on the image. This process transforms each pixel in the RGB color image into a grayscale image based on the light intensity of each pixel. The result of this process is a grayscale image, where color and light intensity have been converted to shades of gray.

2) Face Detection using Viola-Jones: In this research, facial images of individuals serve as the basis for gender classification. Therefore, an additional step is required to detect the positions of individual faces in these images. This process is accomplished using the Viola-Jones algorithm, commonly known as Haar Cascade. After this detection step, due to some faces not being detected by the Viola-Jones algorithm, a total of 590 images from the FG-Net dataset and 355 images from the ORL dataset can be used for further analysis.

3) Cropping Face Region: After successfully detecting the facial positions using the Viola-Jones algorithm, the next step is to crop the identified facial regions. The cropping process is adjusted according to the positions identified by the Viola-Jones algorithm. The purpose of this cropping is to focus the image more on the individual's face area. By isolating and retaining only the facial region, it aims to minimize disturbances caused by the background in the image. This can enhance the gender classification system's ability to identify gender characteristics present in the facial area by disregarding irrelevant background information. There is no reduction in the number of images during this process.

4) Resizing Image to 100x100 Pixel: In this research, the image size used is 100x100 pixels. It is necessary to resize the original images to 100x100 pixels. This process involves scaling each pixel to achieve the desired size, making the original image smaller or larger while still retaining relevant visual information. Resizing the images to a uniform size facilitates subsequent processes in this research.

5) Adding Gaussian Blur: The addition of Gaussian Blur aimed to reduce noise and smooth details in the image. This Gaussian Blur addition is performed to enhance the system's ability to handle small variations and optimize feature representation. This process can mitigate potential disruptions from small variations that might impact the accuracy of gender classification based on facial images. In this research, images are subjected to various configurations of the Gaussian Blur radius, namely with values of 1 (no blur), 3, and 5.

C. Feature Extraction

In this research, two local descriptor methods are employed for additional feature extraction, alongside Self-PCA, namely Local Binary Patterns (LBP) and Local Phase Quantization (LPQ).

LBP is utilized to identify local texture patterns in images without being affected by scale changes. It is applied to extract facial texture features at the pixel level, where each pixel in the image is compared with the intensity value of its central pixel and labeled binary based on this comparison. The result is a series of binary patterns representing the texture distribution in the local pixel environment that can be seen in Fig. 6.

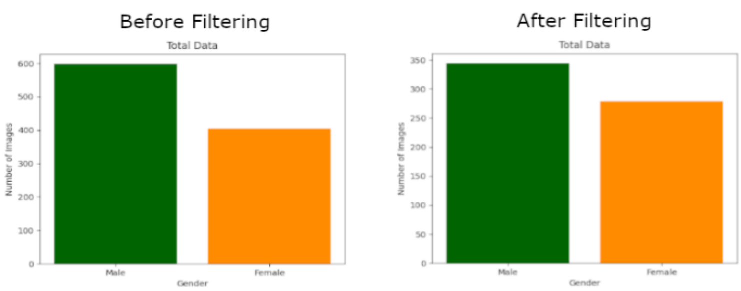

Fig. 5 Total images in FG-Net dataset

On the other hand, LPQ is used to identify local frequency patterns in facial images without being influenced by blur effects. This process involves converting pixel intensity values into frequency information through Fourier transformation, followed by quantization using the signum function. Each pixel in the image is then assigned a binary label based on the quantization results, forming a series of binary patterns representing the frequency distribution in the local pixel environment that can be seen in Fig. 7.

D. Eigenface Creation

This step commences by taking a test image, transforming it into a single-column vector for PCA analysis. Subsequently, image normalization is carried out for each class to be created—male and female, in this case. This involves converting them into a single-column vector and subtracting the average of the test image for each image before storing them in a matrix. Following this, two PCA objects will be created for each class, which are then trained with the pre-existing training data. PCA is trained using the transpose of the previously normalized matrix. This training process generates eigencomponents (eigenfaces) that represent the variations in the data for each class. Result of the eigenfaces used in this research can be seen in Fig. 8 for each genders.

E. Test Image Projection

The steps in performing the test image projection begin with converting the image into a single-column vector. Next, the vector is normalized by subtracting its mean. This normalization process aims to eliminate common (global) features, allowing a more focused emphasis on specific features. After the normalization process, the vector of the test image is projected into the eigen space generated from the previous PCA analysis. This projection is carried out in two eigen spaces: male and female classes. The result of this projection can be seen in Fig. 9, provides a representation of the test image in the relevant eigen space.

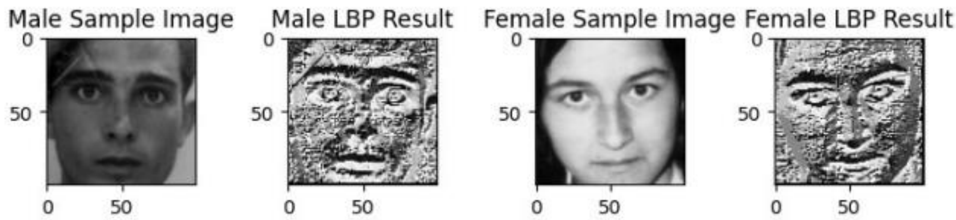

Fig. 6 Feature extraction result using LBP

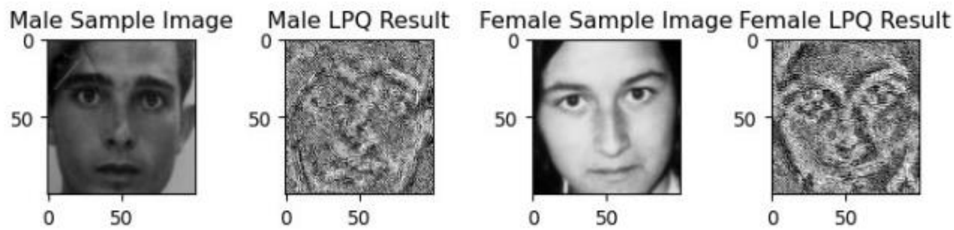

Fig. 7 Feature extraction result using LPQ

Fig. 8 Eigenface result for each genders

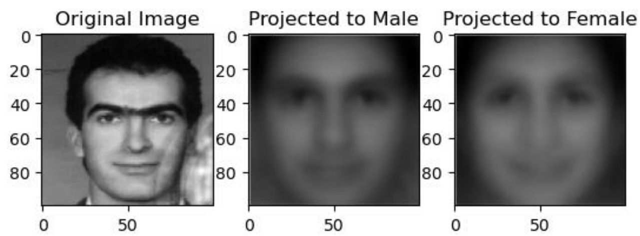

Fig. 9 Projection result

F. Calculating Euclidean Distance

To obtain the classification result, the initial step involves computing the Euclidean distance between the test image projected into the male eigen space. This process returns a matrix that includes the distances between the test image and each relevant eigen vector. A similar step is also applied to the female eigen space. Subsequently, the search for the smallest value among these distance calculations is conducted. This value indicates how close the test image is to a particular image representation in the eigen space. Following this, a comparison is performed between the smallest distances for each class to ascertain the gender prediction for the image.

G. Evaluation

In this research, testing was conducted on three different set data, with each set being. The three datasets were shuffled using seed values of 42, 65, and 131. The accuracy results reported in this research represent the average accuracy from these three set data shuffling experiments. The utilization of three set data serves as a measure to ensure the consistency of the experimental outcomes.

Before conducting tests on the proposed method, testing will be performed on the foundational method of this research, the Self-PCA method. After obtaining the results from Self-PCA, the next step is to conduct tests using the combined method of Self-PCA, LBP, and LPQ. In the LBP and LPQ methods, there are adjustable parameters according to the needs. For LBP, the window_size parameter is used to set the window size calculated from the center point, and the n_points parameter is the number of neighboring points/pixels used. Meanwhile, for LPQ, the window_size parameter is used to set the overall calculated window size (from most left pixel to most right pixel of the window), and the parameter 'a' is the extracted frequency value. In this experiment, the window_size values used are 1, 2, and 3 for LBP, 3, 5, and 7 for LPQ. The n_points values are 3, 4, 5, 6, 7, 8, and the 'a' value is 2, 3, 4.

Utilizing local descriptors in conjunction with Self-PCA on the FG-Net dataset has been proven to enhance gender classification performance as seen on Table II, especially on images with significant blur (minimal noise). This indicates that the combined method proves to be more effective in capturing relevant facial information in images with minimal noise. The integration of these techniques showcases the potential for improved classification accuracy, particularly in scenarios with blurred facial features and minimal noise interference. However, on the ORL dataset, the accuracy results remain stagnant and consistent across all methods on data without blurring. This could be attributed to the class imbalance in the ORL dataset, which will be discussed later.

Not only in terms of classification accuracy but also, the confusion matrix results on Fig. 10 give interesting findings. By utilizing local descriptors, the system successfully recognizes the characteristics of each gender more effectively, as reflected in the more balanced classification distribution between male and female classes. This differs from the Self-PCA method, which tends to classify into the dominant class, suggesting that the combined use of local descriptors with Self-PCA provides a more balanced approach in identifying gender features in facial images.

TABLE II
COMPARISON OF ACCURACY RESULT OF EACH METHOD

Dataset	Blur	Self-PCA	LBP + Self-PCA			LPQ + Self-PCA		
		Accuracy	window_size	n_points	Accuracy	window_size	a_value	Accuracy
FG-Net	1	55.37%	**1**	**4**	**55.64%**	5	2	54.55%
	3	55.37%	1	4	56.47%	**5**	**5**	**57.85%**
	5	55.37%	1	4	54.55%	**5**	**3**	**57.85%**
ORL	1	**90.14%**	**1**	**3**	**90.14%**	**3**	**3**	**90.14%**
	3	**89.04%**	2	3	86.30%	**3**	**3**	**89.04%**
	5	**89.04%**	1	3	84.02%	**3**	**3**	**89.04%**

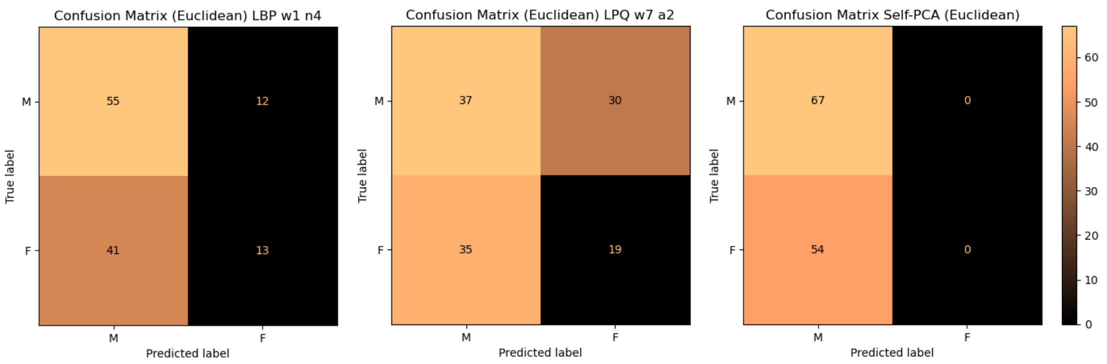

Fig. 10 Confusion matrix for each method in FG-Net dataset

On the other hand, in the ORL dataset, it is observed that the use of local descriptors can achieve similar performance to the Self-PCA method. This situation indicates that there are specific conditions in the dataset that result in comparable performance for each method. One possible cause is the imbalance in the number of classes in the dataset. With the potential class imbalance in the ORL dataset, relying solely on accuracy metrics is insufficient to evaluate the model's performance. To address this, metrics such as Precision, Recall, and F1 Score are employed (Table III and IV).

It can be seen from Table III and IV that in the ORL dataset, the precision, recall, and F1 Score values are quite low close to zero and the majority is zero. From that table can be seen that the system tends to prioritize classification of one gender, especially the male gender, due to the dominance of male data. This limitation is attributed to the limited amount of female gender data in the dataset, causing the system to face challenges in recognizing and classifying features of the female gender.

Apart from that, the computation time during the training process can be observed in Fig. 11 and 12. It can be seen that the LPQ+Self-PCA method takes more time to extract features, with LPQ requiring 1046.39 seconds on the FG-Net dataset and 699.54 seconds on the ORL dataset for training. Meanwhile, the LBP+Self-PCA method takes 44.89 seconds on the FG-Net dataset and 26.14 seconds on the ORL dataset. The Self-PCA method alone requires only 2.30 seconds on the FG-Net dataset and 1.30 seconds on the ORL dataset. For testing on a single image, Self-PCA takes 0.002 seconds, LBP takes 0.35 seconds, and LPQ takes 1.87 seconds.

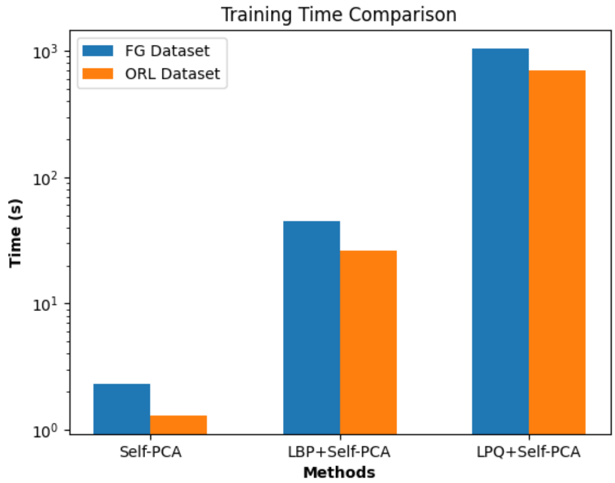

Fig. 11 Training time comparison for each method

Fig. 12 Execution time comparison for each method

TABLE III
COMPARISON OF PRECISION, RECALL AND F1 OF EACH CLASS USING LBP+SELF-PCA

Dataset	Blur	Accuracy	Male Precision	Male Recall	Male F1 Score	Female Precision	Female Recall	Female F1 Score
FG-Net	1	55.64%	0,5650	0,8657	0,6833	0,5127	0,1728	0,2544
	3	56.47%	0,5697	0,8756	0,6900	0,5407	0,1790	0,2660
	5	54.55%	0,5641	0,7860	0,6566	0,4848	0,2469	0,3258
ORL	1	90,14%	0,9014	1	0,948148	0	0	0
	3	86,30%	0,8911	0,9641	0,9260	0,0667	0,0417	0,0513
	5	84,02%	0,8845	0,9436	0,9127	0	0	0

TABLE IV
COMPARISON OF PRECISION, RECALL AND F1 OF EACH CLASS USING LPQ+SELF-PCA

Dataset	Blur	Accuracy	Male Precision	Male Recall	Male F1 Score	Female Precision	Female Recall	Female F1 Score
FG-Net	1	54,55%	0,5461	0,7463	0,6304	0,4139	0,2284	0,2930
	3	57,85%	0,5645	0,9353	0,7039	0,6128	0,1049	0,1755
	5	57,85%	0,5645	0,9353	0,7039	0,6128	0,1049	0,1755
ORL	1	90,14%	0,9014	1	0,9481	0	0	0
	3	89,04%	0,8904	1	0,9420	0	0	0
	5	89,04%	0,8904	1	0,9420	0	0	0

IV. CONCLUSION

This research uses a combination of feature extraction methods, utilizing local descriptors like Local Binary Patterns (LBP) and Local Phase Quantization (LPQ) alongside Self-Principal Component Analysis (Self-PCA) for human gender classification from facial images in the FG-Net and ORL datasets. The gender classification model proposed in this research achieves recognition accuracy of 56.47% using LBP+Self-PCA and 57.85% using LPQ+Self-PCA on the FG-Net image data, and 90.14% using either LBP+Self-PCA or LPQ+Self-PCA on ORL image data. This represents an improvement compared to just using Self-PCA [12] as the feature extraction process which only gives accuracy of 55.64% on FG-Net dataset. Notably, on the ORL dataset, with class imbalance, results in similar accuracies for both combinations with the same accuracy result as Self-PCA. Beyond accuracy, the confusion matrix of LBP + Self-PCA and LPQ + Self-PCA methods reveals more consistent and balanced gender recognition compared to the Self-PCA [12] method. Recommendations for future work include exploring additional parameters in LBP and LPQ, incorporating other local descriptor methods like BSIF, and optimizing computational time, particularly for the time-intensive LPQ + Self-PCA method. This could involve investigating more efficient methods or techniques for feature extraction.

ACKNOWLEDGEMENT

The authors extend sincere gratitude to Faculty of Computer Science, University of Jember for supporting and funding this research. The authors sincerely hope that this research contributes to the field of computer vision, advancing our understanding and paving the way for future developments.

REFERENCES

[1] T. Ojala, M. Pietikäinen, and D. Harwood, “A comparative study of texture measures with classification based on featured distributions,” *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, Jan. 1996, doi: 10.1016/0031-3203(95)00067-4.

[2] A. Hadid, “The Local Binary Pattern Approach and its Applications to Face Analysis,” in *2008 First Workshops on Image Processing Theory, Tools and Applications*, Nov. 2008, pp. 1–9. doi: 10.1109/IPTA.2008.4743795.

[3] S. Najafi Khanbebin and V. Mehrdad, “Local improvement approach and linear discriminant analysis-based local binary pattern for face recognition,” *Neural Comput & Applic*, vol. 33, no. 13, pp. 7691–7707, Jul. 2021, doi: 10.1007/s00521-020-05512-3.

[4] D. Shi and H. Tang, "Face recognition algorithm based on self-adaptive blocking local binary pattern," *Multimed Tools Appl*, vol. 80, no. 16, pp. 23899–23921, Jul. 2021, doi: 10.1007/s11042-021-10825-z.

[5] H. N. Vu, M. H. Nguyen, and C. Pham, "Masked face recognition with convolutional neural networks and local binary patterns," *Appl Intell*, vol. 52, no. 5, pp. 5497–5512, Mar. 2022, doi: 10.1007/s10489-021-02728-1.

[6] Y. El merabet, Y. Ruichek, and A. El idrissi, "Attractive-and-repulsive center-symmetric local binary patterns for texture classification," *Engineering Applications of Artificial Intelligence*, vol. 78, pp. 158–172, Feb. 2019, doi: 10.1016/j.engappai.2018.11.011.

[7] A. Hadid, J. Ylioinas, M. Bengherabi, M. Ghahramani, and A. Taleb-Ahmed, "Gender and texture classification: A comparative analysis using 13 variants of local binary patterns," *Pattern Recognition Letters*, vol. 68, pp. 231–238, Dec. 2015, doi: 10.1016/j.patrec.2015.04.017.

[8] T. Song, J. Feng, S. Wang, and Y. Xie, "Spatially weighted order binary pattern for color texture classification," *Expert Systems with Applications*, vol. 147, p. 113167, Jun. 2020, doi: 10.1016/j.eswa.2019.113167.

[9] C. Shan, "Learning local binary patterns for gender classification on real-world face images," *Pattern Recognition Letters*, vol. 33, no. 4, pp. 431–437, Mar. 2012, doi: 10.1016/j.patrec.2011.05.016.

[10] V. Ojansivu and J. Heikkilä, "Blur Insensitive Texture Classification Using Local Phase Quantization," in *Image and Signal Processing*, A. Elmoataz, O. Lezoray, F. Nouboud, and D. Mammass, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2008, pp. 236–243. doi: 10.1007/978-3-540-69905-7_27.

[11] J. Kannala and E. Rahtu, "BSIF: Binarized statistical image features," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, Nov. 2012, pp. 1363–1366. Accessed: Oct. 21, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/6460393

[12] J. S. Nayak and M. Indiramma, "An approach to enhance age invariant face recognition performance based on gender classification," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 8, Part A, pp. 5183–5191, Sep. 2022, doi: 10.1016/j.jksuci.2021.01.005.

[13] J. S. Nayak, M. Indiramma, and N. Nagarathna, "Modeling self-Principal Component Analysis for age invariant face recognition," in *2012 IEEE International Conference on Computational Intelligence and Computing Research*, Dec. 2012, pp. 1–5. doi: 10.1109/ICCIC.2012.6510277.

[14] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 442–455, Apr. 2002, doi: 10.1109/34.993553.

[15] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proceedings of 1994 IEEE Workshop on Applications of Computer Vision*, Dec. 1994, pp. 138–142. doi: 10.1109/ACV.1994.341300.

[16] Ş. Öztürk and B. Akdemir, "Effects of Histopathological Image Pre-processing on Convolutional Neural Networks," *Procedia Computer Science*, vol. 132, pp. 396–403, Jan. 2018, doi: 10.1016/j.procs.2018.05.166.

[17] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, "A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data," *Frontiers in Energy Research*, vol. 9, 2021, Accessed: Oct. 05, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fenrg.2021.652801