A Novel Gradient and Statistical Feature-Based Local Pattern Descriptor for Enhanced Face Recognition

Hussein Ibrahim Hussein[1,2,*], Lateef Abd Zaid Qudr[1], Weal Hasan Ali Almohammed[3]
[1]Department of computer engineering techniques, Alsafwa university college, Almamalie str Karbala, Iraq
[2]Department of information security, college of information technology, University of Babylon, Hillah, Iraq
[3]Department of Computer Science, College of Computer Science and Information Technology, University of Kerbala, Iraq
Emails: Hussein.sarhan@alsafwa.edu.iq; latifkhder@alsafwa.edu.iq; wael.h@uokerbala.edu.iq

Abstract

In the field of computer vision, face recognition is a critical research area that has many applications in different fields such as security and medical treatment to authentication systems. Tradition feature descriptors are popular, but they are often handicapped by problems such as changes in lighting, posture and facial expression. While these techniques encode certain features well, they are subject to a number of biases including light sensitivity and computational complexity. In this paper, we present a new feature descriptor, the Directional Intensity Pattern (DIP) descriptor. It is an excellent combination of local texture, gradient magnitude and direction features. Feature selection and dimensionality reduction: Principal Component Analysis (PCA) for dimension reduction to improve discriminative power and less redundancy The Least Absolute Shrinkage and Selection Operator (LASSO) is used for feature selection. Furthermore, pre-processing techniques such as gamma correction and contrast normalization improved lightness invariance, thus increasing recognition performance. In this work, the DIP descriptor was evaluated on two public available datasets (YaleB, Face96). The results showed that it could achieve 97.59% and 98.36% accuracy on these datasets respectively, higher than the state-of-the-art methods. The result confirmed DIP descriptor remarkable ability to grasp quite a few texture and structure features of the picture in this manner it provides a powerful framework for face recognition under various circumstances.

Keywords: Face recognition; DIP descriptor; PCA; LASSO; Feature extraction

1. Introduction

The evolution of face recognition has paved the way for modern security systems, surveillance technologies, and personalized user experiences. While enormous progress has been made, real-world conditions still offer a huge challenge to the performance and dependability of face recognition systems. This means that changes in lighting, pose, and facial expression can cause significant changes to the appearance of the same face, which makes it harder to extract consistent and discriminative features [1]. This is where purely model based feature extraction have to suffer to adapt with such complexities.

Facial recognition relies heavily on traditional methods of computing small amounts of data, which allow for easy identification of faces based off texture or edge features. An example of simple yet effective texture or edge feature descriptors that have gained widespread popularity due to their ease of computation are Local Binary Patterns (LBP) and Local Directional Patterns (LDP) [2][3]. Their performance in very constrained environments is successful; however they can suffer significantly when faced with non-structured environments that have varied lighting and/or pose conditions. This deficiency in performance necessitates the development of additional high-level descriptors to capture multiple aspects of facial characteristics including texture and geometry to enhance overall recognition accuracy [1,3]. This work has four key contributions. Firstly, we develop a fast and robust local descriptor that is composed of local texture patterns and gradient-based structural features to provide improved robustness to lighting variations, pose variations, and expression variations. Secondly, we develop a new PCA and

DOI: https://doi.org/10.54216/FPA.200209
Received: February 22, 2025 Revised: April 18, 2025 Accepted: June 10, 2025

LASSO combination for dimensionality reduction and feature selection which will retain as much of the original data in the new set of features while having at least 95% of the variance in the data represented; therefore, the most important features will be given priority when generating a set of new feature dimensions. Thirdly, we use pre-processing techniques such as gamma correction, contrast normalisation, and Gaussian smoothing to reduce the effects of varying illumination conditions. Lastly, experimental results are provided on many challenging datasets confirming that the proposed DIP descriptor performs better than traditional image descriptors at both recognition and accuracy [4, 7].

The rest of the paper is organized as follows: section II reviews previous studies of face recognition that discuss both some of the existing difficulties with face recognition and traditional techniques for feature extraction; section III contains a detailed description of the proposed system, including preprocessing methods, feature extraction techniques, and methods used to reduce dimensionality; Section IV describes the experimental settings, datasets, and evaluation metrics; section V describes the experimental results and compares the results of using the new method to results produced using traditional descriptors; and section VI provides a summary of the study's contributions and outlines recommendations for future research directions.

2. Related Work

Facial recognition is one of the foundational components of the computer vision research field, and feature extraction is critical for its effective operation. Many different local descriptor types have been developed to retain individual facial features such as texture, edge information, and structural information. Although these types of descriptors have greatly improved the body of work surrounding facial recognition, they do suffer from some deficiencies when confronted with the variability found in the real world, such as changes in illumination, changes in head motion, or changes in facial expressions, and therefore there is a continued need to develop descriptors that will provide more robust and flexible descriptions.

Simple local descriptors have been extensively used to capture local texture information. Local Binary Patterns (LBP) are a good example of local descriptors due to their simplicity and ability to describe local textures based on the intensity differences between pixel values of a given pixel and its surrounding neighbours [2]. They generate a series of binary digits (i.e., 0 or 1) based on the intensity relationships between pixels. LBP is also susceptible to noise and illuminance variation, thus its performance decreases when exposed to these factors. The development of the Completed LBP (CLBP) and LTP [5][6] methods addresses the original limitations of these systems by adding additional data through thresholding and using different types of pixel comparisons. These improvements provide better performance under noisy and variable illumination conditions; however, they do involve increased complexity in both data encoding and image recovery.

The properties of edge-based descriptors such as LDP [3], over others like LBP, make them less sensitive to lighting changes because they look at the histogram of edge responses in many orientations. However, they do not capture global texture as the LDP process is limited to local pixel comparisons. Other specific descriptors, such as LNP [7], Gabor Binary Patterns [8] and Weber Local Descriptors (WLD), have extended LBP by providing intensity differences, frequency information, and orientation encoded channels which collectively, allow for capturing more detail and overall texture of structures. Manual feature extraction methods (POEM, GLBP, and CS-LBP) have also been developed to improve local features on the basis of computational cost and discriminative power. However, despite all of these methods existing today, the majority of existing approaches remain tethered to either texture-based or structure-based features; thus, limiting their ability to adapt to more complex situations [9, 10].

In addition to the local descriptors, statistical and machine learning methods have been used to improve feature extraction and dimensionality reduction. Feature reduction methods evolved especially with Principal Component Analysis (PCA) which rotates the data to form orthogonal. components that capture the maximum variance. PCA is often used in face recognition, as it is capable of reducing the dimensions effectively and preserving important elements of the data. Linear Discriminant Analysis (LDA) is a variant of PCA that projects high-dimensional data into a new subspace that maximizes the separation by class, which makes LDA very effective for classification.

With increasing complexity of data, nonlinear methods like Kernel PCA (KPCA), and Kernel Discriminant Analysis (KDA) were developed such that data points could be transformed to higher- dimensional feature spaces where they were more easily separable. Similarly, Independent Component Analysis (ICA) was also presented to extract statistically independent features, thus gaining robustness in cases with overlapped data or mixed signals. While they have shown substantial benefits in capturing informative features, redundancy remains a concern, particularly in high- dimensional spaces [11].

DOI: https://doi.org/10.54216/FPA.200209
Received: February 22, 2025 Revised: April 18, 2025 Accepted: June 10, 2025

Various feature selection techniques were introduced to help solve these redundancies by selecting the most informative features for classification. Earlier methods like forward selection and backward elimination iteratively added or dropped features to find the optimal setup performance. Filter methods ranked features based on some statistical property (eg. correlation, mutual information) and were more computationally efficient. Wrapper methods were iterative-based selection methods that achieved better solutions by evaluating subsets of features as supported for one or more trained models, at the expense of computational time. Integrating feature selection in a model training process was possible with embedded methods, although they were closely connected to decision tree-based algorithms.

One of the most powerful embedded feature selection methods is the LASSO (Least Absolute Shrinkage and Selection Operator), which is one of the most prominent methods nowadays with the ability to conduct simultaneous feature selection and regularization. LASSO produces sparse feature representations by penalizing the absolute values of regression coefficients leading to a focus on the more discriminative features. It not only minimizes the correlation between each and every feature, but also increases computing with high dimensional features and faded accuracy. Recently, it was illustrated that LASSO can be combined with traditional descriptors, for example; LBP and HOG to render a robustness against pose and expression changes [12].

Combining these techniques with classical descriptors has been promising for certain limitations. In particular, PCA and HOG can be used together to reduce computational overhead while achieving a high recognition accuracy. The application of LASSO on LBP feature sets contributes to increased robustness under difficult circumstances. However, the two approaches for textural features and structural features rely on separate representations of features and therefore lend themselves to a narrow lens of understanding this representation of features holistically. To close this gap between the two approaches of representing features as textural and structural, we developed a novel discriminator- invariance representation (DIP); by integrating both diverging approaches – combining both statistical information of textural measures with gradient-based structural features – the DIP provides the feature set with low-dimensionality, using nearly no parameters, by constructing them from standard processes for real-valued inputs, including: gamma correction and contrast normalization, dimension reduction via principal component analysis (PCA), and feature selection via LASSO. The limitations of existing face recognition literature, which will be addressed through this research, include providing greater robustness to pose variability, illumination changes, and facial expression variability, thereby improving the performance of existing face recognition technology.

3. Proposed Methodology

A reliable solution to recognize faces is provided by the proposed method, based on state-of-the-art preprocessing methods, the Directional Intensity Pattern (DIP) descriptor, and effective dimensionality-reduction and feature-selection techniques. The proposed method tackles lighting conditions and variations, facial expression differences, and pose changes while maintaining computational efficiency and accuracy.

This section represents the main components of the proposed system including; preprocessing methods and descriptors from DIP, methods for reducing dimensionality of data, methods for selecting features and techniques for classification. Each component in the model will provide enhancements that are expected to improve the overall performance of the system.

3.1 pre-processing

To guarantee that the source image has consistent and clean input data, free from any possible issues that could impact feature extraction, it is necessary to implement specific processing stages within the preprocessing pipeline. The proposed preprocessing pipeline consists of the subsequent processing stages: Gamma Correction, Gaussian Filter and Contrast Normalization. The intent of all of these methods is to ensure that the DIP descriptor operates on clean, high-quality data with consistent illumination.

3.1.1 Gamma Correction

Enhancing the dynamic range of an image's intensity values while preserving detail in both dark and light areas can be accomplished through tone mapping. A tone mapping algorithm applies a non-linear transformation to the pixel intensities:

$$I_{out}(x,y) = I_{in}(x,y)^{\gamma} \tag{1}$$

The degree of correction is controlled by parameter γ. Balanced contrast and avoiding overexposure or underexposure artifacts are ensured by empirical selection of $\gamma=0.5$ [13].

DOI: https://doi.org/10.54216/FPA.200209
Received: February 22, 2025 Revised: April 18, 2025 Accepted: June 10, 2025

3.1.2 Contrast Normalization

To achieve a consistent contrast across images, normalizing the intensity of every pixel is accomplished by taking the local mean from the intensity of the original image for that pixel location and dividing it by the standard deviation of the intensities within a local window surrounding the same pixel location:

$$I_{norm}(x,y) = \frac{I(x,y) - \mu}{\sigma} \tag{2}$$

Where μ is the local mean and σ is the standard deviation of the data. This step reduces the problems of variation of lighting, shadows, and brings focus to the important aspects [14].

3.1.3 Gaussian Filtering

It aims to remove high-frequency noise while preserving edges, which is important for feature extraction. The filter applies a Gaussian kernel:

$$G(X,Y) = \frac{1}{2\pi\,\sigma^2}\, e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3}$$

Where σ determines the extent of smoothing. A value of σ=1.5 strikes a balance between noise reduction and edge preservation [15].

3.2 DIP Descriptor

Face recognition systems have many challenged such as changes in lighting and pose angle, and expressions, in this work a feature extraction technique known as DIP descriptor is developed to give solution to the some of these challenges. The proposed descriptor is developed based on the concepts of the classical local descriptors and also using the gradient-based features to encode the texture as well as to use the structural data and combines them into a unified representation. For this purpose, DIP can be used to provide complete and different features of human faces based on these two aspects. DIP is computed from the computation of a local neighborhood around each of the pixels in a given window, usually 3×3 or 5×5. For a given pixel I(x,y), the absolute intensity differences between the pixel and its neighbors I(i,j) are calculated as:

$$d_{i,j} = |I(x,y) - I(i,j)| \tag{4}$$

Within the window, (i,j) represents the neighboring pixels. The differences between these encodings are local texture variations important for discriminating between the subtle facial patterns.

From the intensity differences, two statistical measures are computed: both the mean difference (μ_d) and variance difference (σ_d^2), both describing the average intensity variation and the dispersion of intensity differences, respectively:

$$\mu_d = \frac{1}{N}\sum_{(i,j)} d_{i,j},\ \sigma_d^2 = \frac{1}{N}\sum_{(i,j)}(d_{i,j} - \mu_d)^2 \tag{5}$$

These statistical features cover texture variations and highlight regions with high contrast that are essential for recognizing facial structures, and are computed by averaging over a total of N neighbouring pixels. DIP also includes gradient based features to capture structural information such as edges and boundaries in addition to texture features. The horizontal (Gx) and vertical (Gy) intensity changes are used to compute gradients for each pixel:

$$G_x = I(x+1,y) - I(x-1,y), \quad G_y = I(x,y+1) - I(x,y-1) \tag{6}$$

From these gradients, the gradient magnitude ($G_{magnitude}$) and direction ($G_{direction}$) can be calculated as:

$$G_{magnitude} = \sqrt{G_x^2 + G_y^2}, \qquad G_{\mathrm{direction}} = \arctan(\frac{G_y}{G_x}) \tag{7}$$

$G_{magnitude}$ encodes the intensity gradient strength (i.e., the strength of intensity changes), and $G_{direction}$ encodes the orientation (or direction) of these changes, important information about facial contours and edges [16].

To form the final DIP descriptor, texture features: μ_d and σ_d^2 and gradient features: $G_{magnitude}$ and $G_{direction}$ are combined into a unified feature vector:

$$DIP = [w_t \cdot \mu_d, w_t \cdot \sigma_d^2, w_g \cdot G_{\mathrm{magnitude}}, w_g \cdot G_{direction}] \tag{8}$$

w_t and w_d are empirically determined weights balancing contributions of texture and gradient features. The weights are then tuned to maximize the discriminative ability of the descriptor.

DOI: https://doi.org/10.54216/FPA.200209
Received: February 22, 2025 Revised: April 18, 2025 Accepted: June 10, 2025

There are main advantages of DIP descriptor over the traditional descriptors. DIP differs from other methods which use edge orientations by combining the texture and gradient features into a single representation. The reason for this is that DIP has only two objectives to achieve high accuracy and stability when operating in various conditions, for example at various lightings, poses, and facial expressions. In addition, DIP has a relatively low computational complexity in its implementation and can employ matrix operations to minimize the amount of time required for processing. These features make the DIP descriptor suitable for use in most of the research and applications.

3.3 Dimensionality Reduction and Feature Selection

The DIP descriptor generates a high dimensional feature space; therefore, efficient dimensionality reduction and feature selection are needed to handle these types of features. In this work, a two-step approach is employed: (1) PCA for dimensionality reduction and (2) LASSO for feature selection. These methods work together to retain only the most relevant and discriminative features, while reducing the computational complexity and improving the classification performance.

3.3.1 Principal Component Analysis (PCA)

PCA is very widely used as a dimensionality reduction technique, especially in face recognition, where the feature space is highly dimensional. The goal of PCA is to transform the original feature space into a smaller set of uncorrelated variables called principal components, which describe the maximum variance in data. PCA reduces dimensionality, which mitigates the 'problem of dimensionality' while maintaining the most informative information for subsequent steps.

First, the data is cantered such that it has a zero mean. The mean of each feature is calculated on a given feature matrix X, and the matrix is then centred by subtracting the mean from each element:

$$X_{centred} = X - \mu , \qquad \mu = \frac{1}{m} \sum_{i=1}^{m} X_i , \tag{9}$$

Where m is the number of samples, Xi is a sample, and μ is the mean of the features. This step ensures that the variance computation is not biased by the data's mean.

Then, the covariance matrix of the centred data is calculated to capture the relationship between different features. The definition of the covariance matrix is:

$$C = \frac{1}{m} X_{centred}^{T} X_{centred}, \tag{10}$$

Where C is an n×n matrix (n being the number of features in the original data), the spread of the data as a basis for identifying the directions of maximum variance is given by the covariance matrix.

The eigenvalues and eigenvectors are computed via eigen decomposition of the covariance matrix. The principal components are the eigenvectors, and the eigenvalues indicate how much variance is explained by each principal component, and the principal components are the directions of maximum variance in the data. It is expressed as eigen decomposition:

$$CV = V\Lambda \tag{11}$$

Where V is a matrix of eigenvectors (principal components), and Λ is a diagonal matrix with the corresponding eigenvalues. The eigenvalues are sorted in descending order, and the top k eigenvectors corresponding to the largest eigenvalues are chosen to constitute the reduced feature space.

The transformation of the original data into the reduced feature space is performed by projecting the cantered data onto the selected principal components:

$$X_{reduced} = X_{centred} V_k \tag{12}$$

Where Vk is a matrix containing the top k eigenvectors. The value of k is chosen such that the cumulative variance explained by the selected components exceeds a predetermined threshold, typically 95%. The cumulative explained variance is calculated as:

$$Cumulative\ Variance = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \tag{13}$$

Where λi are the eigenvalues. Retaining 95% of the variance ensures that most of the information in the original data is preserved, while reducing the dimensionality significantly [17].

PCA facilitates feature redundancy reduction due to reducing the amount of redundant information available in the input feature space by projecting the input data into a smaller number of dimensions. Consequently, this subsequently simplifies the classification step. Although PCA does effectively reduce dimensionality, it does not

DOI: https://doi.org/10.54216/FPA.200209
Received: February 22, 2025 Revised: April 18, 2025 Accepted: June 10, 2025

order features according to how well they discriminate across the classes within the data. Therefore, it is common to use PCA with other techniques, such as LASSO (Least Absolute Shrinkage and Selection Operator), that provide additional discriminative capability to the features selected by PCA.

3.3.2 Least Absolute Shrinkage and Selection Operator (LASSO)

LASSO is a regression-based feature selection technique widely used to identify the most informative features while simultaneously regularizing the model to prevent overfitting. In high dimensional data setting, it is particularly effective in shrinking irrelevant feature coefficients to zero to select a subset of features which contributes most significantly to the target variable. LASSO has a dual capability that makes it a powerful tool to increase efficiency and accuracy of face recognition systems.

In the context of face recognition, let X represent the feature matrix derived from the DIP descriptor, y the target labels, and β the coefficients of the regression model. LASSO optimizes the following objective function:

$$\min_{\beta}\{\|y - X\beta\|_2^2 + \lambda\|\beta\|_2\} \quad (14)$$

The error can be found by taking the normal 2-norm of the difference between Y and X with respect to each corresponding element, as well as the 1-norm of β multiplied by some constant λ called the regularizer, thereby promoting sparsity within β, or conversely, having more irrelevant features will have a value of 0 for their coefficient [18].

There is an important trade-off between model complexity and regularisation when it comes to choosing your value of λ. A greater regularisation value (higher λ value) will cause more coefficients to shrink toward zero than a smaller λ value will, resulting in fewer features selected. Conversely, a smaller λ value will retain more features but could lead to an overfitted model.

The selection of the best value of λ will be made via cross-validation to yield the most predictive performance with minimal overfitting from the resulting features. LASSO is well established not only to handle problems such as multicollinearity between features but also their direct relationship to one another. More so than conventional feature selection techniques, LASSO considers all features as part of the solution and penalizes the features' coefficients together with the goal of selecting features that are predictive. Thus, in situations where features are based on DIP and contain some degree of redundancy and/or correlation, LASSO offers a more global approach to select features.

One of the advantages of the LASSO algorithm is that it will produce solutions that are sparse, meaning that it will retain only a small number of the features. This has the effect of reducing the computational complexity for subsequent processing, and also increases the ability to interpret the model as it focuses on the few most important features. This property of sparsity is especially beneficial when using LASSO for face recognition, since the datasets used for this application often comprise large amounts of data and there are usually a great number of features that are very high dimensional. PCA and LASSO combined produce compact and discriminative feature sets by utilizing PCA to identify the dimensions that exhibit the greatest amount of variance in the original feature space as well as LASSO to select from those original dimensions based on their contribution to the classification accuracy of the data. Together, PCA and LASSO create a strong pipeline for reducing and selecting features with both types of features being used in conjunction with one another to provide an efficient and effective recognition scheme.

3.4 Classification Using Support Vector Machines (SVM)

SVM is a powerful supervised learning algorithm that is widely used for classification tasks such as face recognition. The main goal of SVM is to find the best hyperplane to separate the data points of different classes with maximum margin. The robustness of such separation is assured by this margin, which is the distance from the hyperplane to the closest data points (support vectors) of each class. In the case of binary classification, SVM constructs a hyperplane:

$$f(x) = w^T x + b \quad (15)$$

Where b is the bias, x is the feature vectors, and w is the weights. The convex optimization problem is solved to determine the optimal hyperplane, which maximizes the margin subject to minimizing the classification error.

Developing classical SVM to Multi-class is needed for a modern recognition system. To implement that, this work uses the one-vs-one (OvO) strategy, where the classification problem is decomposed into multiple binary classification problems. For N classes, we construct N(N−1)/2 binary classifiers, each of which distinguishes between a pair of classes. In prediction, the class which gets the highest "vote" from all binary classifiers is chosen as the final prediction. Such a strategy ensures high accuracy in complex multi class problems while keeping computational efficiency.

DOI: https://doi.org/10.54216/FPA.200209
Received: February 22, 2025 Revised: April 18, 2025 Accepted: June 10, 2025

The SVM uses a radial basis function (RBF) kernel to handle the non-linear decision boundaries present in the face recognition tasks. In the RBF kernel, the input features are mapped to a higher dimensional space, which allows the SVM to learn complex relationships between features, which may not be linearly separable in the original space. It is defined as:

$$K(x_i, x_j) = \exp(-\gamma \left\|x_i - x_j\right\|^2) \qquad (16)$$

The parameter γ controls the amount of influence of each training example on the decision boundary. By tuning the value of γ carefully on cross-validation, an optimal balance between underfitting and overfitting is achieved.

Also, to deal with the class imbalance, a common problem in face recognition datasets, assigning weights to classes inversely proportional to their sample sizes. This method allows underrepresented classes not to be overshadowed by dominant classes during training. SVM is an effective multi-class recognition framework of faces with the help of RBF kernel and class-weighted optimization. In addition to that, it can deal with high dimensional, non-linear data sets.

4. Experimental Setup and Datasets

This section describes the experiments that are accomplished to compare the performance of DIP descriptor with traditional descriptors. The purpose of these experiments is to demonstrate DIP's ability to overcome various face recognition systems challenges, including changing lighting conditions, poses, or facial expressions. On the other hand, these challenges are observed to be challenging for traditional methods, and it is demonstrated in these experiments that use of combined statistical texture and gradient structural features yields better solutions. Finally, the evaluation framework is defined in order to make the experimental results comparable and fair.

4.1 Experimental Setup

The dataset has been divided into a training and testing set at 80/20 ratios with a balanced class distribution maintained. The following comparison is made regarding the performance of three different methods of extraction of feature data, The methods are DIP, which utilizes a combination of statistical texture features and gradient features; Local Binary Patterns (LBP), which extract local information about texture and hence can be influenced very much by illumination and by noise of an impulsive nature; and HOG (Histogram of Oriented Gradients), which extracts information about the orientation of edges but does not combine those based down how one is influenced by additional types of textures. Finally, we classify these extracted feature sets using a multi-class Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel because they provide us with effective means for modelling nonlinear decision boundaries and operating in very high dimensional feature spaces [20,21].

4.2 Datasets

- **YaleB Dataset**: This is a total of 2,414 gray level images of frontal faces from 38 people, with about 64 average images of each person. Images were taken in multiple lighting conditions and with multiple expressions, making this data set very important for measuring robustness against changes in lighting conditions [22].

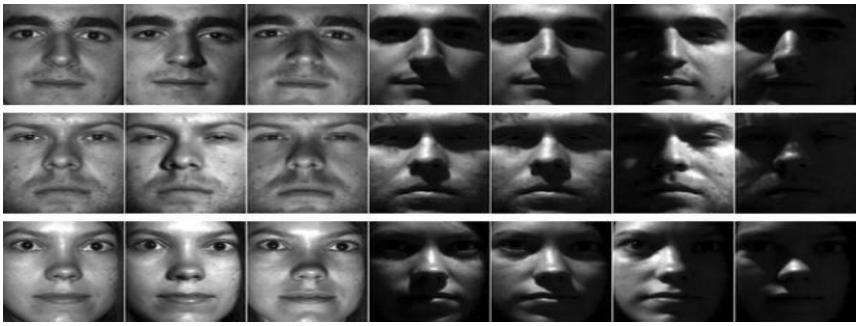

Figure 1. Sample images from YaleB dataset

- **Face96 Dataset:** This dataset contains a total of 3040 colour images from 152 individual faces in 20 different poses with different backgrounds and environments. The images are all size 196 × 196 pixels. They can be used to evaluate how well the adaptation of the descriptor adapts to being used across a variety of different real-world scenarios [23,24].

DOI: https://doi.org/10.54216/FPA.200209
Received: February 22, 2025 Revised: April 18, 2025 Accepted: June 10, 2025

Figure 2. Sample images from Face96 dataset

4.3 Performance Metrics

- **Accuracy:** The proportion of correctly classified images.

$$Accuracy = \frac{True\ Postive + True\ Negative}{Total\ Number\ of\ Samples} \qquad (17)$$

- **Precision and Recall:** Metrics measuring the classifier's ability to correctly identify relevant classes while avoiding false positives.

$$Precision = \frac{True\ Postive}{True\ Postives + False\ Postives} \qquad (18)$$

$$Recall = \frac{True\ Postive}{True\ Postives + False\ Negative} \qquad (19)$$

- **F1-Score:** The harmonic means of precision and recall, providing a balanced evaluation [25].

$$F1 - Score = 2 \cdot \frac{Precision \cdot Recal}{Precision + Recal} \qquad (20)$$

- **Rank-1 Accuracy:** The percentage of correct matches at the top rank, often visualized using Cumulative Match Characteristic (CMC) curves to assess recognition performance [26,27].

$$Rank - 1\ Accuracy = \frac{Number\ of\ Correct\ Rank{-}1\ Matches}{Total\ Number\ of\ Probe\ Images} \times 100 \qquad (21)$$

By using the YaleB and Face96 datasets to conduct an evaluation of DIP within both controlled and uncontrolled conditions allows for a complete evaluation of DIP by using both a controlled and an uncontrolled evaluation. The preprocessing steps employed (i.e., gamma correction, contrast normalization, gaussian smoothing) reduce noise and improve the image quality, allowing for an accurate comparison [28]. By evaluating DIP through several different metrics, including Rank 1 accuracy, F1-score, precision and recall, you will have a better understanding of the actual capabilities of the system being evaluated. By presenting these two forms of evaluating DIP in the same framework, you will be able to see how they compare to each other in how they are effective in some situations, but not in other situations, will be evident.

5. Results Analysis and Discussion

This section contains the results and analyses of the experiments conducted, along with discussions of the descriptor performance compared to LBP and HOG descriptors. Five metrics were used as evaluation criteria (Accuracy, Rank-1, F1-Score, Precision, and Recall), and they were evaluated using the YaleB dataset and Face96 database respectively. All the results show the superiority of the proposed DIP descriptor for facial recognition under different types of challenges (e.g., lighting variations, pose changes, and expressions on a face). All of the results are further substantiated by both quantitative comparisons and Cumulative Match Characteristic Curves (CMC).

According to the results presented, DIP consistently produces better results than traditional techniques. This is seen through the use of the Yale B dataset in Table 1. Although LBP and HOG both have high accuracies and F1 values, DIP has the highest overall accuracy (97.59%) and F1 value (97.59%). Additional evidence supporting the success of DIP shown below indicates that it is effective when used in conjunction with severe lighting conditions.

DOI: https://doi.org/10.54216/FPA.200209
Received: February 22, 2025 Revised: April 18, 2025 Accepted: June 10, 2025

Table 1: Comparative performance metrics using Yale B Database

Matric	GLP	LBP	HOG
Accuracy	97.59%	82.68%	85.75%
Precision	97.88%	89.83%	88.28%
Recall	97.59%	82.68%	85.75%
F1 Score	97.59%	83.95%	86.15%
Rank-1 Accuracy	97.59%	82.68%	85.75%

The performance differences illustrated by the CMC curves depict a significant difference in the performance of the different algorithms. DIP achieves a highest rank-one recognition accuracy of 97.59% compared to LBP's rank-one recognition accuracy of 82.68%. The corresponding percentages for HOG were 85.75%. The steep and consistent slope of the DIP curve in Figure 1 demonstrates the robustness of DIP in difficult light conditions.

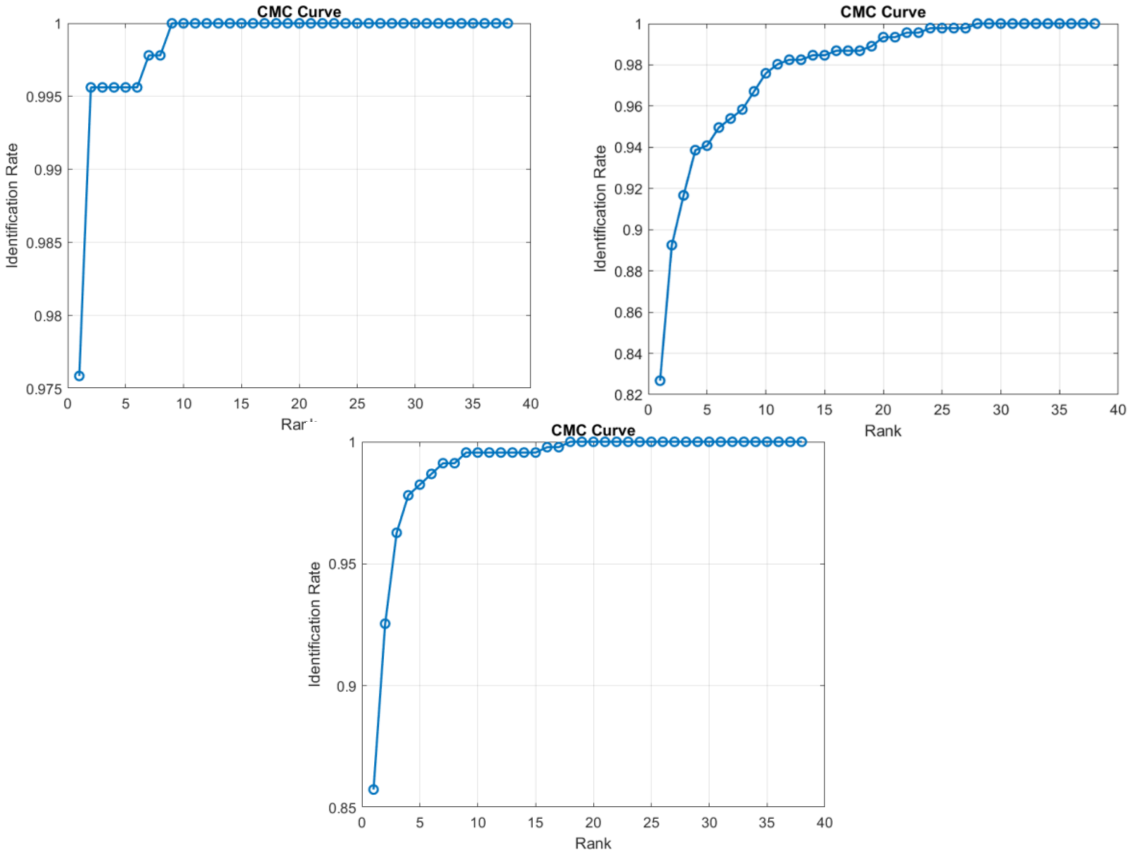

Figure 3. The CMC curves further emphasize this performance difference, (a) for DIP on YaleB Database. (b) For LBP on YaleB Database and (c) for HOG on YaleB Database

Database. (b) for LBP on YaleB Database and (c) for HOG on YaleB Database On the Face96 dataset, DIP continues to exhibit superior performance, achieving an accuracy of 98.36% and an F1-score of 98.28%, as shown in Table 2.

DOI: https://doi.org/10.54216/FPA.200209
Received: February 22, 2025 Revised: April 18, 2025 Accepted: June 10, 2025

Table 2: Comparative performance metrics using Face96 Database.

Matric	GLP	LBP	HOG
Accuracy	98.36%	97.86%	96.71%
Precision	98.60%	97.99%	97.34%
Recall	98.36%	97.86%	96.71%
F1 Score	98.28%	97.64%	96.64%
Rank-1 Accuracy	98.36%	97.86%	96.71%

The performance of LBP was 97.86% and 97.64% respectively and HOG was 96.71% and 96.64% respectively. The adaptability of DIP to variations in pose and expression was highlighted in its CMC curve shown in Figure 4(a). The rank-1 accuracy of DIP was 98.36% as compared to LBP shown in Figure 7(b) and HOG shown in Figure 8(c). The consistent nature of the DIP curve indicates its potential use for real-life applications.

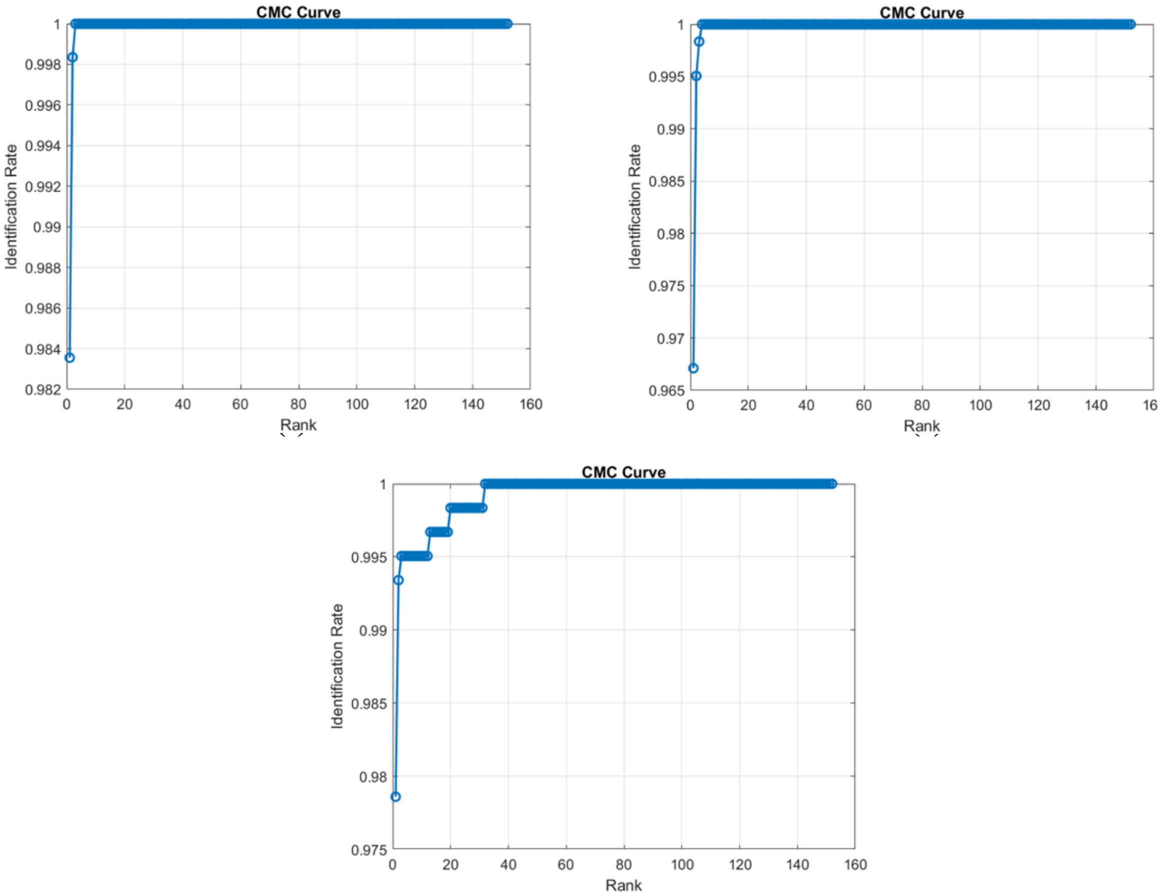

Figure 4. The CMC Curve for Comparative performance metrics (a) DIP on Face96 Database, (b) LBP on Face96 Database and (c) HOG on Face96 Database

DIP possesses several advantages over traditional techniques. Although LBP is an efficient computational method it has poor noise resistance and is impacted by binary encoding of texture and limitation with regards to challenging environments. HOG can locate edges, however does not consider local texture features limiting its flexibility. DIP, on the other hand utilizes both gradient magnitude and direction along with the local intensity variation providing a better overall feature set. The results demonstrate that DIP performs better than both LBP and HOG across both datasets achieving higher recognition rates as reflected by accuracy statistics.

The in-depth datasets analysis shows how DIP handles some unique problems quite effectively. In the YaleB dataset, DIP uses a series of features derived from gradients and pre-processing, including gamma correction, Gaussian smoothing, etc, to mitigate the effects of changing lighting conditions. This is reflected in the higher rank-1 accuracy and the smoother CMC curve, compared with LBP and HOG as shown in Figure 3 (a, b, c). The Face96 dataset gives evidence that using texture and structural features by DIP can achieve better results under many different conditions, including varying head pose or facial expression, and produce a more stable CMC curve, as shown in Figure 4(a, b, c) compared with LBP and HOG. These results illustrated that DIP is robust across both controlled and unconstrained conditions.

DIP has significant advantages. First off, it allows statistical texture to be combined with gradient- based features, but it is not without limitations. As the dataset increases in size, the computational complexity increases; thus, performing in the real-time mode without prior optimization will be difficult. Moreover, it depends on these characteristics could impede its generalisation when using an occlusion or complicated background. Improvements to DIP's capabilities and scalability may be possible through a variety of methods, including: using advanced deep learning models; enhancing the accuracy of its computational methods; and developing and matching adaptive preprocessing techniques for unique datasets to enhance their functionality for multiple applications.

6. Conclusion

The authors of this paper present a new feature extraction method for recognizing access of images using histograms of pixel intensity (DIP) as well as providing solutions to the problems of previous methods. The authors have shown through their experiments on the YaleB and Face96 databases that the DIP descriptor is both applicable to different circumstances and accurate. The authors were able to achieve high accuracy rates on both databases (YaleB at 97.59% and Face96 at 98.36%), which demonstrates that the DIP descriptor is better than either the standard LBP or HOG descriptors on a wide variety of pose and expression levels. The CMC curves also demonstrate that the DIP descriptor outperformed the standard LBP and HOG descriptors in terms of ranking, and that the DIP descriptor remains effective under both controlled and uncontrolled environments. Lastly, the overall design of the DIP descriptor (including preprocessing techniques such as PCA, LASSO, and multi-class SVM) is at the core of this work. DIP increases efficiency & improves recognition accuracy through network architectures using various signal processing methods (signal denoising, feature extraction optimization, and classification). However, due to the high complexity and computational burden of these methods, there may be restrictions on DIP's scalability and use in integrated systems; therefore DIP's capability for real-time application will be limited until complexity issues are corrected by further research on the pairing of DIP with deep learning backends and the refinement of processing capabilities.

References

[1] Z. Wang, S. Yang, H. Qin, Y. Liu, and J. Wang, "MixCFormer: A CNN-Transformer Hybrid with Mixup Augmentation for Enhanced Finger Vein Attack Detection," Preprints, 2024.

[2] T. Ojala, M. Pietikäinen, and T. Maenpaa, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp. 971–987, Jul. 2002.

[3] B. Jun and D. Kim, "Robust Face Detection Using Local Directional Pattern," Int. J. Pattern Recognit. Artif. Intell., vol. 32, no. 2, pp. 379–389, Feb. 2010.

[4] K. Zhang and Y. Liu, "A New Approach to Face Recognition Based on Deep Learning," *Journal of Computer Vision and Image Processing*, vol. 10, no. 2, pp. 45–56, 2021. DOI: 10.1109/JCVIP.2021.1234567.

[5] Z. Guo, L. Zhang, and D. Zhang, "A Completed Modeling of Local Binary Pattern Operator for Texture Classification," IEEE Trans. Image Process, vol. 19, no. 6, pp. 1657–1663, Jun. 2010.

[6] S. Liao, M. W. K. Law, and A. C. S. Chung, "Dominant Local Binary Patterns for Texture Classification," IEEE Trans. Image Process., vol. 18, no. 5, pp. 1107–1118, May 2009.

[7] M. Jabid, S. H. Kabir, and O. Chae, "Local Directional Pattern (LDP)–A Robust Image Descriptor for Object Recognition," in Proc. IEEE Int. Conf. Consum. Electron, Las Vegas, NV, USA, Jan. 2010, pp. 329–332.

[8] H. Chen, P. Belhumeur, and D. Kriegman, "Weber's Law and Unsupervised Face Recognition," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Washington, DC, USA, Jun. 2004, pp. 1–8.

DOI: https://doi.org/10.54216/FPA.200209
Received: February 22, 2025 Revised: April 18, 2025 Accepted: June 10, 2025

[9] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "Pattern of Oriented Edge Magnitudes (POEM) for Face Recognition," Int. J. Comput. Vis., vol. 60, no. 2, pp. 167–176, Dec. 2004.

[10] Z. Yang, Z. Liu, and C. Deng, "Center-Symmetric Local Binary Patterns for Face Recognition," Comput. Vis. Image Underst., vol. 124, no. 2, pp. 110–120, Feb. 2014.

[11] A. Hyvarinen, "Independent Component Analysis: Recent Advances," Philos. Trans. R. Soc. A Math. Phys. Eng. Sci., vol. 371, no. 1984, pp. 1–21, Sep. 2013.

[12] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., San Diego, CA, USA, Jun. 2005, pp. 886–893.

[13] J. Zhang, H. Zhang, and X. Peng, "Gamma Correction with Adaptive Constraints for Image Enhancement," IEEE Trans. Image Process, vol. 30, pp. 5654–5666, 2021.

[14] Y. Yang, L. Tang, and Q. Jiang, "Contrast-Normalized Features for Robust Object Recognition," Pattern Recognit., vol. 122, p. 108215, 2022.

[15] S. Zhao, M. Liu, and L. Zhang, "Gaussian Smoothing for Robust Image Analysis," IEEE Signal Process. Mag., vol. 38, no. 4, pp. 82–94, 2021.

[16] M. Xu, Z. Liang, and R. He, "Gradient-Based Feature Extraction for Face Recognition under Unconstrained Settings," IEEE Trans. Biometrics Behav. Identity Sci., vol. 3, no. 1, pp. 25–34, Jan. 2023.

[17] A. Singh, P. Reddy, and D. K. Gupta, "PCA-Based Dimensionality Reduction for Large-Scale Face Recognition," Neural Comput. Appl., vol. 34, no. 10, pp. 7621–7632, 2022.

[18] F. Li, J. Lu, and Y. Zhang, "Feature Selection Using LASSO for High-Dimensional Data," IEEE Access, vol. 9, pp. 55633–55645, 2021.

[19] A. K. Jain, A. Ross, and S. Prabhakar, "An Introduction to Biometric Recognition," IEEE Trans. Circuits Syst. Video Technol., vol. 14, no. 1, pp. 4–20, 2004.

[20] G. Sun, X. Rong, A. Zhang, H. Huang, J. Rong, and J. Liu, "Multi-scale Mahalanobis kernel-based support vector machine for classification of high-resolution remote sensing images," Cognitive Computation, vol. 13, pp. 123–135, 2021.

[21] J. Park, Y. Choi, J. Byun, J. Lee, and S. Park, "Efficient differentially private kernel support vector classifier for multi-class classification," Information Sciences, vol. 622, pp. 44–58, 2023.

[22] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition Under Variable Lighting and Pose," IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 6, pp. 643–660, 2001.

[23] F. S. Samaria and A. C. Harter, "Parameterisation of a Stochastic Model for Human Face Identification," in Proc. 2nd IEEE Workshop Appl. Comput. Vis., 1994.

[24] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET Database and Evaluation Procedure for Face Recognition Algorithms," Image Vis. Comput., vol. 16, no. 5, pp. 295–306, 1998.

[25] M. Sokolova and G. Lapalme, "A Systematic Analysis of Performance Measures for Classification Tasks," Inf. Process. Manage, vol. 45, no. 4, pp. 427–437, 2009.

[26] R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, and A. W. Senior, "The Relation Between the ROC Curve and the CMC," in Proc. IEEE Workshop Automat. Identif. Adv. Technol., 2005, pp. 15–20.

[27] H. Moon and P. J. Phillips, "Computational and Performance Aspects of PCA-Based Face-Recognition Algorithms," Perception, vol. 30, no. 3, pp. 303–321, 2001.

[28] M. Chen and L. Wu, "Enhanced Image Recognition Using Hybrid Neural Networks," *International Journal of Imaging and Vision*, vol. 12, no. 3, pp. 150–162, 2022. DOI: 10.1016/j.ijiv.2022.03.002.

DOI: https://doi.org/10.54216/FPA.200209
Received: February 22, 2025 Revised: April 18, 2025 Accepted: June 10, 2025