Fusion-Based 2.5D Face Recognition System

Min-Er Teo
Faculty of Information Science and Technology, Multimedia University, Malaysia

Lee-Ying Chong
Faculty of Information Science and Technology, Multimedia University, Malaysia

Siew-Chin Chong
Faculty of Information Science and Technology, Multimedia University, Malaysia

Abstract: Face recognition is the dominant biometrics system used to authenticate an individual's identity in various applications. Most commercial face recognition systems rely on 2D face images, but the changes in the environment lighting and a person's posture affect the accuracy of the 2D face recognition systems. Hence, the 2.5D face recognition system arises as the solution to eliminate the drawbacks of the 2D face recognition system. The depth feature in the 2.5D data (depth image) provides additional information that can help to improve the accuracy and robustness of 2.5D face recognition systems, particularly in challenging scenarios. This paper proposes a fusion-based approach for the 2.5D face recognition system to enhance the system's performance, where feature fusion involves the combination of features extracted from the depth image. Gabor-based Region Covariance Matrices (GRCMs) that serve as face identifiers combine the depth and texture images in the structure of a covariance matrix. Several experiments on different fusions have been conducted in the Face Recognition Grand Challenge version 2 (FRGC v2.0) database. This paper shows that the max-min fusion applied to the surface normal (y-direction) and the mean curvature has achieved the best accuracy rate of 93.66% among the other fusion approaches used.

Keywords: fusion-based approach, depth image, 2.5D data, Gabor-based Region Covariance Matrices, 2.5D face recognition

Introduction

Face recognition has attracted the interest of researchers, since it enables non-intrusive detection, identification, and authentication without seeking the individual's knowledge or consent. Researchers have recently been passionate about investigating the 2.5D face recognition system considering that, although the 2D face recognition system has made significant improvements in the last decade, the effectiveness of the findings is still heavily reliant on lighting conditions and a person's position. The disadvantages of 2D face recognition, such as lighting changes and difficulty with facial emotions, could potentially be overcome using 2.5D face recognition. Furthermore, by gathering less data and processing it more effectively, 2.5D face recognition boosts the reliability of 3D face recognition, which is time-consuming and costly. Each pixel — x and y on the camera's perspective — in a 2.5D image contains just one depth value — z — which shows its distance from the camera's scene. As a result, the 2.5D data accurately depicts the 3D structure, uninfluenced by variations in colour or lighting conditions (Chong *et al*., 2016).

The Gabor-based region covariance matrices (GRCMs) as face identifiers have gained interest and are frequently used in facial recognition systems to perform feature extraction. GRCM works in a way that integrates the Gabor feature with the face picture in the covariance matrix. The covariance matrix is an essential component of the region covariance matrices (RCMs) because it is able to gradually combine distinct image data while maintaining the connection among the features of the image (Chong *et al*., 2014). In this paper, GRCM is used to merge the depth and texture images, which are presented in sequential and direct addition methods, respectively.

In this work, a fusion-based 2.5D face recognition system is suggested to enhance the accuracy rate of the system. The typical approach for forming feature fusion in a face recognition system is by combining two features or merging two features using certain algorithms. The feature fusion technique, which generates more informative features, could strengthen the reliability of 2.5D face recognition. Several studies are carried out in order to determine the optimum fusion technique that produces the maximum accuracy rate and raises the system's efficiency. The experimental findings for each fusion strategy have been documented, along with the best and worst fusion techniques that were defined based on the results.

The objectives of this paper to overcome the constraints of the present 2.5D face recognition system are: (1) to investigate the different features of 2.5D data; (2) to study the fusion techniques of the 2.5D face recognition system; and (3) to conduct experiments in order to evaluate the system's effectiveness.

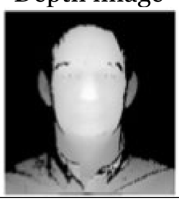
Literature Review

Face recognition

A face recognition system is a form of biometric security that uses facial biometric data and patterns to authenticate a person's identification. Two-dimensional (2D) face recognition represents a type of facial recognition technology that leverages the advantage of the two-dimensional geometry of a human face. The public widely accepts the 2D face recognition system, and the gadget used to capture the image is less expensive than a 3D face recognition system. The main drawback of 2D face recognition is that the accuracy of the output is strongly reliant on the illumination and postures of an individual in the captured picture or photo. With these constraints, three-dimensional (3D) face recognition is presented as a solution to the dilemma of 2D technology. A 3D face recognition system outperforms a 2D facial recognition system because it can gather additional data (surface normal, curvature) from a person's face and is less susceptible to lighting. As contrast, the primary technical limitation of 3D facial recognition techniques is the acquisition of a 3D picture, which normally demands a range camera, which is costly.

2.5D face recognition techniques have been developed to address the high-cost issue in 3D face recognition systems. 2.5D face recognition looks to be an effective technique that makes use of depth (range) face images. 2.5D data is a "digital image" created by 3D face scanning that represents a certain position's face look. To create a comprehensive 3D face model, a large amount of 2.5D data is collected from various views in 3D facial recognition technology. The 2.5D face recognition system requires just one use of the 2.5D data. 2.5D face data, which stores three-dimensional coordinates — *x, y,* and *z* — is perfect for the structure because it provides a depth value, the *z*-coordinate, which is a necessary component of a 3D model. Each 2.5D face data set is translated into a depth representation for computational efficiency. The depth value, which is the *z* coordinate, is kept in a 2D matrix structure known as a depth image (Chong *et al.*, 2019). Table 1 shows a comparison of all three forms of face recognition.

Table 1. Differences between three types of face recognition (2D, 2.5D and 3D)

Face Recognition Technology	**2D**	**2.5D**	**3D**
Image Format	Texture-based image 	Depth image 	3D facial model 
Pre-processing Element	Easy	Middle	Hard
Cost of Gadget	Cheap	Middle	Expensive
Coordinates Involved	Coordinates x, y	Coordinates x, y, z	Coordinates x, y, z
Limitations	Posture, facial expression, lighting	Posture, facial expression	Posture, facial expression

Feature fusion in face recognition

Bodla *et al.* (2017) presented a novel deep heterogeneous feature fusion network strategy for blueprint face recognition that employs complementary substance in features provided by several deep convolutional neural networks (DCNN). In addition, a feature fusion approach that merges two independent feature sources to represent a face image with the use of a Canonical Correlation Analysis (CCA) algorithm has been proposed by Nhat & Hoang (2019).

Dutta *et al.* (2021) suggested a complement components (CC) mathematical model for face elements based on depth points extracted from the depth image. Furthermore, an improved Region Covariance Matrix (RCM) for the 2.5D face identification system is implied to boost the recognition efficiency of the system by overcoming the vanilla RCM's shortcomings in obtaining distinctive features from face photos; and failure in recognition of faces is presented by Chong *et al.* (2017).

Region covariance matrix approaches

Tuzel *et al.* (2006) suggested the Region Covariance Matrix (RCM) approach, which gives a new area classifier and demonstrates how it could address two problems: recognition of an object and texture categorising. This approach is represented as a covariance matrix to combine numerous picture numbers produced within a sole image region. The RCM is implemented to define the characteristics of the region. However, RCM shows poor results when implemented in a 2D face recognition technique, proving that it is unsuitable to be employed as a face classifier in 2D technology.

As a result, Pang *et al.* (2008) recommended Gabor-based area covariance matrices (GRCMs) as face descriptors to identify a human face. The Gabor features include more due to their strong spatial localisation, size, and position consistency. By including Gabor characteristics in the calculation of area covariance, the RCM's descriptiveness, along with its differentiating ability, may be significantly improved. Thus, the proposed GRCM approach has the potential to generate excellent face recognition results.

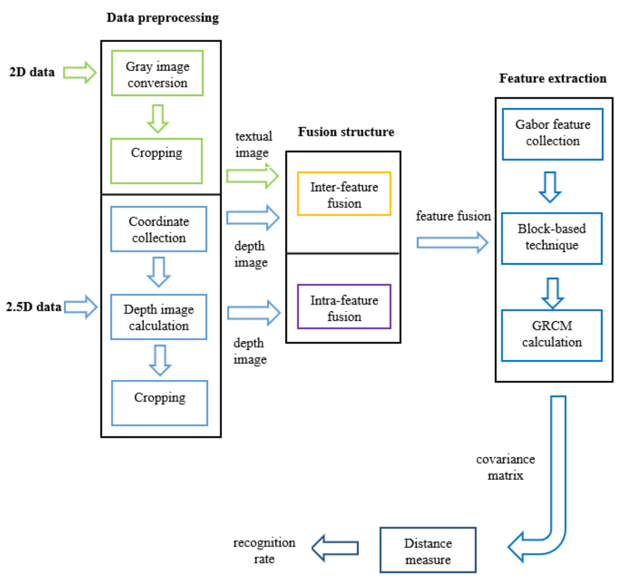


Figure 1. The proposed 2.5D fusion-based face recognition system

Research Methodology

Based on Figure 1, the suggested approach consists of four phases: the data preprocessing stage, feature fusion stage, feature extraction stage, and distance measure stage. Note that the 2D and 2.5D partial data are pre-processed before being combined to generate the inter-feature fusion.

Pre-processing of data

To get started, the partial 2.5D and 2D face images need to be pre-processed to standardise the size, minimise noise, and establish zero mean normalisation of the image. The pre-processing processes of 2.5D data involve the collection of coordinates, computation of depth image, cropping, normalising, and standardising of faces in order to build a normalised depth picture. The x, y, and z coordinates are collected from each data point in 2.5D data. To extract a face region from the partial 2.5D data, the unnecessary background setting is removed. This depth image is generated by interpolating the z-coordinate retrieved from the rectangular grid in the x-y plane. Cropping the picture according to the location of both the eyes and the mouth yields a canonical depth image. The standard depth image is normalised via a typical standard adjustment, which rescales the data to ensure the variance and mean are set to zero. Figure 2 shows the procedure of the pre-processing steps for 2.5D data.

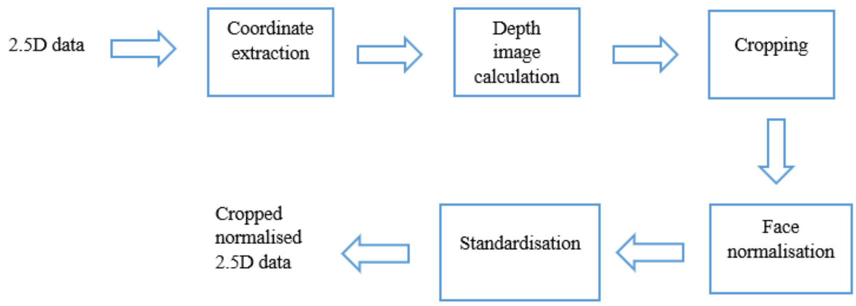


Figure 2: The procedure for the pre-processing of 2.5D data

The 2D pre-processing steps include grayscale image computing, image cropping, and normalising along with standardising faces. The 2D face image data is converted into a grayscale image. The textual data is then manually cropped based on the middle of the lips and both eyes. Following that, the standard normal transformation is employed to normalise the canonical texture image depending on the locations of the mouth and eyes. The standardisation phase aims to achieve zero mean and unit variance. Figure 3 illustrates the process of pre-processing for 2D data.

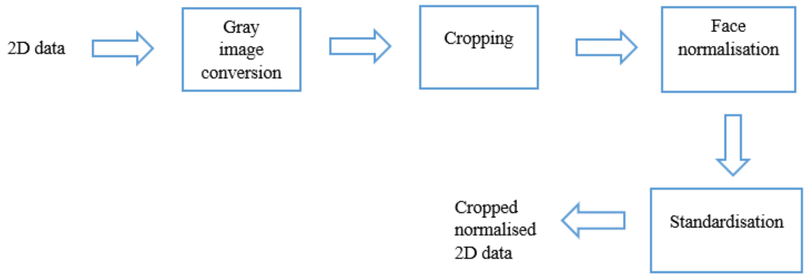


Figure 3. The procedure for the pre-processing of 2D data

Feature fusion

Feature fusion represents the integration of features from different parts or phases of an entity. In a face recognition system, feature fusion involves combining two distinct techniques; merging and mixing the methods improves the system's accuracy and dependability. Different facial traits could represent various parts of a person's facial characteristics, and employing these unique characteristics may enhance one's facial recognition. Facial recognition techniques extensively employ feature fusion to boost the system's recognition accuracy rate (Talab *et al.*, 2022). In this section, different feature fusion methods including inter-feature fusion, intra-feature fusion, and fusion integration methods have been used to fuse features.

Inter-feature fusion is known as one of the feature fusion methods in a face recognition system. The process of inter-feature fusion includes merging data from several different characteristics or features collected from various face modalities or portions. This fusion approach attempts to enhance the overall effectiveness of the facial recognition system by using complementary information from diverse sources. For instance, this method merges different types of information, such as a 2D image with a depth image, using various fusion techniques, such as max fusion, min fusion, and many more.

By contrast, intra-feature fusion differs from inter-feature fusion in the way that it aims to integrate data through a sole feature description. In order to gather more accurate and reliable information, it requires acquiring several descriptions from a single area of the face. This fusion method seeks to improve the description of a single feature through various orientations and sizes. For example, some properties can be extracted using the depth image's comprehensive collection of face topological data, such as the curvature feature.

Furthermore, a specific kind of fusion technique that combines previously fused data once again is referred to as the fusion integration approach. This fusion technique aims to combine and gather more discriminative features from the fused data to boost the system's performance. For instance, this fusion method is generated by integrating the intra-feature fusion with another intra-feature fusion using different fusion approaches, such as the sum

http://doi.org/10.18080/jtde.v12n1.770

rule, max-min fusion, and others. The comparison between inter-feature fusion, intra-feature fusion, and fusion integration is exhibited in Table 2.

Table 2. Comparison between three types of feature fusions

	Inter-Feature Fusion	**Intra-Feature Fusion**	**Fusion Integration**
Structure	Combination of the depth image with 2D image	Features produced from depth image, such as mean curvature, surface normal etc.	Combination of the intra-feature fusion with another intra-fusion feature
Number of fusions occurring	One	One	Two
Involves fusion technique?	Yes	No	Yes

Inter-feature fusion

Inter-feature fusion combines features extracted from several raw images to boost the system's accuracy. A novel fusion dataset is developed by combining the two distinct characteristics provided by the textual data and the depth data. Minimum fusion is known as one of the fusion strategies. This technique creates a fused image by selecting the pixels with the minimum data point (Kaur *et al.*, 2021). For example, Equation (1) shows the use of minimum fusion (*Mn*) with the depth image (A) and the 2D image (B).

$$Mn = Min(A, B) \tag{1}$$

Additionally, maximum fusion (*Mx*) is another fusion approach that separates high-intensity pixels from images to form a combined image as illustrated in Equation (2) (Kaur *et al.*, 2021).

$$Mx = Max(A, B) \tag{2}$$

The Max-Min fusion (*M*) represents one of the fusion techniques that utilises the values retrieved from both the maximum and minimum fusion. The fused picture is created by identifying the average scores of the components with the lowest and highest scores within the entire input picture (Kaur *et al.*, 2021).

$$M = Max(A, B) - Min(A, B) \tag{3}$$

The sum rule can also be applied in the system as one of the fusion approaches to increase the system's accuracy. The sum rule method adds up the features to form a new feature.

Intra-feature fusion

Intra-feature fusion represents another form of feature fusion approach to increase the reliability of the system. The depth image itself collects an extensive amount of face topological information, from which these characteristics may be retrieved for intra-feature fusion.

http://doi.org/10.18080/jtde.v12n1.770

Surface normal represents one of the intra-feature fusion approaches to boost the accuracy rate of the face recognition system. Three sorts of data from axes x, y, and z could be extracted from an individual's face, since the 2.5D face photos often collected in a point-cloud form are made up of x, y, and z coordinates for a face. The vectors with the values Nx, Ny, and Nz represent the three separate data sets collected from the face (Vezzetti & Marcolin, 2012). To generate a new equation, the sum rule of the surface normal (SN) is employed for these acquired different surface normal points, as defined in Equation (4).

$$SN = Nx + Ny + Nz \tag{4}$$

Besides surface normal, an example from among the intra-feature fusion methods used to improve the effectiveness of the face recognition system is curvature as a 3D feature. The minimum and maximum curvatures are the two basic curvatures used to characterise the local shape of a surface. The parameter k_1 is defined as the maximum curvature, while k_2 is defined as the minimum curvature. Then, mean curvature can be computed after getting the maximum and minimum curvature values (Vezzetti *et al.*, 2014). Mean curvature (H) is the average of the minimum (k_2) and maximum (k_1) values, where $k_1 > k_2$ as shown in Equation (5).

$$H = \frac{1}{2}(k_1 + k_2) \tag{5}$$

Furthermore, the Gaussian and mean curvature represents the intrinsic and extrinsic geometric characteristics of the surface. Gaussian curvature (K) is determined from the surface's differential geometry; it remains local, intrinsic, and preserved by affine translation (Vezzetti *et al.*, 2014). It is the product of the two essential curvatures, as displayed in Equation (6).

$$K = k_1 k_2 \tag{6}$$

Fusion integration approach

The fusion integration approach is a type of fusion method that fuses the fused data again. Fusion integration in this section is generated by combining the intra-feature fusion with another intra-feature fusion retrieved from the previous section. For instance, the best recognition rate, which is achieved by the surface normal (y-direction) in the intra-feature fusion, is combined with the second highest recognition rate, which is the surface normal (average), to form a new fusion (fusion integration) and it has been examined through various fusion methods such as using Sum Rule, minimum fusion, and others. There are many types of fusion approaches that can be used in order to increase the system's efficiency and boost the system's performance.

http://doi.org/10.18080/jtde.v12n1.770

Feature extraction

Face recognition necessitates feature extraction that involves segmenting facial pictures, generating images, and scaling faces. Gabor-based region covariance matrices (GRCMs) are frequently employed as face descriptors in facial recognition systems nowadays. Gabor features have more spatial localisation, dimension, and alignment precision than first- and second-order grades, and hence involve more information. By including Gabor characteristics in the calculation of covariance area, the RCM's descriptiveness, along with its differentiating ability, may be significantly improved (Pang *et al.*, 2008). Consequently, the proposed Gabor-based RCM approach yields satisfactory face recognition scores. The region covariance matrix is generated by combining the feature mapping function via Gabor features, as indicated in Equation (7), where R is the fusion technique applied (Chong *et al.*, 2016).

$$\phi(R, x, y) = [x\ y\ g_{00}(x, y), g_{01}(x, y), \dots, g_{74}(x, y)\ ] \tag{7}$$

The collection of Gabor features, the use of a block-based strategy, and the calculation of the GRCM are the three processes that make up this phase. Equation (8) describes the Gabor-based region covariance matrix (GRCM). The Gabor (G) dimensions are determined by the size of the covariance matrix (42×42 dimension), which is formed by multiplying the x and y coordinates of the pixels using a 40 Gabor wavelet and adding up the results (Chong *et al.*, 2016). Figure 4 shows the process of generating the GRCM.

$$G \in \Re^{42\times42} \tag{8}$$

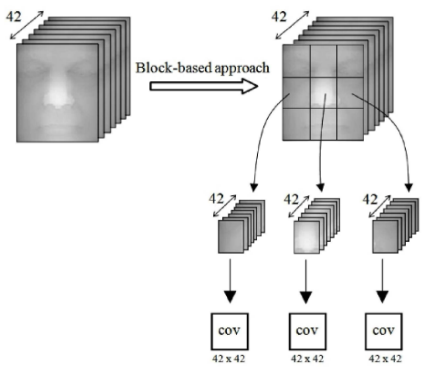


Figure 4. The generation of GRCM map

Distance measure

The distinctive features of a person's face, such as whether the person is male or female, whether the person is wearing glasses or not, and many other things, could be determined and categorised using distance measure metrics. The Euclidean distance is employed to calculate the distance between two coordinates in 2D space and to determine the exact distance between parameters in the space of N dimensions along a straight line. Therefore, in facial recognition, smaller values indicate a higher level of similarity between two faces. The Euclidean distance computation equation is shown in Equation (9), assuming that the image's face characteristic is to be represented as $P = (x_1, x_2, \ldots, x_{128})$ and the training sample's facial traits are to be identified as $Q = (y_1, y_2, \ldots, y_{128})$:

$$PQ = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_{128} - y_{128})^2} \tag{9}$$

The Euclidian distance measurement, unfortunately, is only precise on a straight line. The recognition rate is expected to be low when employing the Euclidian distance technique, since the human face involves numerous curve lines.

Tensor Manifold is a non-Euclidean space that is flat locally but curved globally. Each covariance matrix associated with GRCM remains uniquely equal to a single location on the Tensor manifold. The Euclidean distance measures the distance between two points (array-based data) in Euclidean space, while the particular distance measure calculates the distance between two covariance matrices (matrix-based data) within the Tensor manifold (Chong *et al.*, 2014). The space between two GRCMs must take into account the geometric qualities of the manifold, a measurement termed geodesic distance, which represents the shortest distance between two GRCMs. The tensor manifold yields greater accuracy than Euclidean distance since it does not reside in Euclidean space (Chong *et al.*, 2016).

Cholesky Distance (CHOL) is recognised as a re-parameterisation measurement that divides each GRCM evenly into a mixture of a lower matrix with a triangle form and its transpose (Chong *et al.*, 2014):

$$(P, Q) = \parallel L_P - L_Q \parallel_F \tag{10}$$

LogDet Divergence distance (LD) represents a collection of information-theoretic approaches. It is a kind of matrix difference that determines how far away two GRCMs are (Chong *et al.*, 2014):

$$(P, Q) = log \left|\frac{P-Q}{2}\right| - \frac{1}{2} log|PQ| \tag{11}$$

where $|.|$ is the determinant of the matrix.

Another of the metrics used to determine distance is the Affine Invariant Riemannian Metric (AIRM). This approach employs a similarity measure based on the tensor manifold, which includes eigenvalue decomposition, exponentials, logarithms, and square roots (Chong *et al.*, 2014).

$$(P, Q) = \sqrt{\sum_{i=1}^{5} ln^2 \lambda_i (P, Q)} \tag{12}$$

http://doi.org/10.18080/jtde.v12n1.770

where the eigenvalues of P and Q are represented by $\lambda_1, \dots, \lambda_5$.

Additionally, the Log-Euclidean Riemann Metric (LERM) is another method for calculating distance. This method uses Euclidean metrics logarithms' space to compute the distance between two GRCMs. LERM represents a standard measurement that is employed within the tensor manifold in a face recognition system (Chong *et al.*, 2014):

$$(P, Q) = \| Log(P) - Log(Q) \|_F \tag{13}$$

where Log(.) indicates the logarithm of the matrix and $\|\cdot\|_F$ represents the Frobenius norm.

Recognition rate

Every biometrics system must determine the accuracy rate of its method because it illustrates the system's accuracy and reliability. The true positive (TP) is the number of times the technique correctly identifies the same individual in two different images. The true negative (TN) denotes the number of times the algorithm properly distinguishes between two distinct persons in the images. By dividing the total number of correctly identified images (TP and TN) with the total number of images, the recognition rate is determined:

$$Recogniton\ rate = \frac{(TP + TN)}{(total\ no.of\ images)} \times 100\% \tag{14}$$

Furthermore, one method for evaluating the efficiency of the system is to measure the amount of time in seconds that the system utilised to complete the recognition procedure of the individual from the database. A more rapid computation time used by the system results in an increase in the effectiveness of the system.

Experiments and Discussion

Face dataset

The Face Recognition Grand Challenge version 2 (FRGC v2.0) is being used in this paper. Every time a person's biometric information is collected, a subject session is taken, including four controlled still images, two uncontrolled still shots, along with a 3D picture of an individual. The collection of data in FRGC v2.0 involves 466 individuals, 4,007 subject sessions, and 32,056 recordings.

Experimental setup

A part of FRGC v2.0 is implemented, which consists of a total of 254 subjects. A total of 16 images, 8 images from both the 2.5D partial data and 2D data, were selected randomly for each individual, corresponding to a total of 4,046 images. Each image has been resized and

http://doi.org/10.18080/jtde.v12n1.770

standardised to 73×61 pixels. The pixel intensity of the picture has been adjusted to a zero mean and zero unit variance. The GRCM's Gabor kernel is set up with the following parameters: $k_{max} = \frac{\pi}{2}$, $\sigma = \pi$, $f_v = \sqrt{2}$. When the pixel position (x, y) information is added, a Gabor wavelet with the dimensions 40×40 becomes 42×42. Two distinct fusion schemes, such as intra-feature fusion and inter-feature fusion, along with fusion integration methods, including max fusion, min fusion, max-min fusion, and sum rule, have been examined. Besides, the 2D image, depth image as well as feature fusions use the block-based GRCM in the feature extraction stage.

Results and analysis

Results of inter-feature fusion

Figure 5 exhibits the recognition rate of the several inter-feature fusions based on sum rule, min fusion, max fusion and max-min fusion, using difference distance measures. Based on Figure 5, sum rule with the combination of both the depth image and the 2D image scores the most significant recognition rate using LERM at 93.31%. The sum rule outperforms the single depth image because it incorporates characteristics from both the depth and 2D images. In addition to the sum rule, methods like min fusion, max fusion, and max-min fusion achieve better performance than a depth image. This demonstrates that inter-feature fusion approaches outperform a sole depth image for most distance measures.

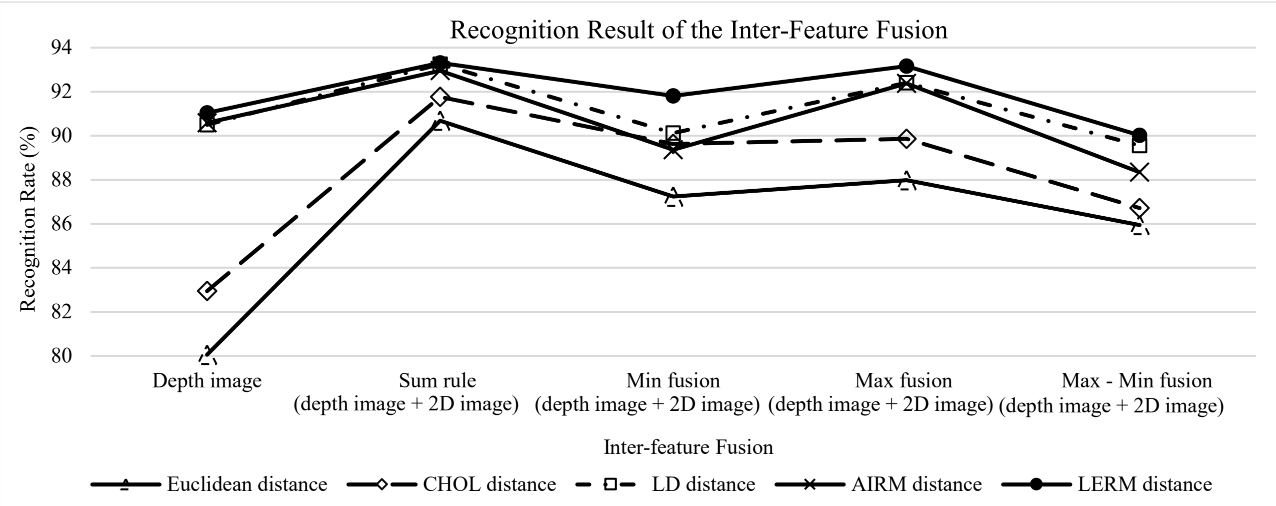


Figure 5. Recognition rate (%) of the inter-feature fusion

The computation time for each inter-feature fusion is shown in Figure 6. It can be seen that every inter-feature fusion utilises the shortest amount of time to compute the recognition rate using Euclidean distance. However, the recognition rate for each inter-feature fusion using Euclidean distance is the worst compared to the other distance measures, as shown in Figure 5. In contrast, although the computation time of the LERM distance to calculate the recognition rate is the longest among the other distance measures as shown in Figure 6, but the recognition rate of the majority of the inter-feature fusion scores the highest performance in LERM distance, as illustrated in Figure 5.

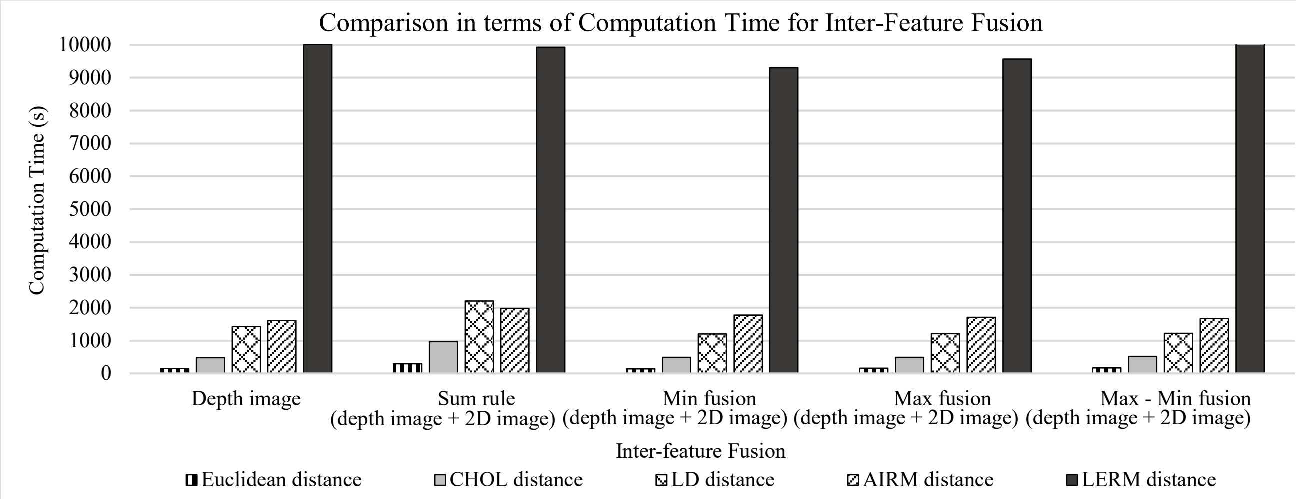


Figure 6. Computation times for inter-feature fusion

Results of intra-feature fusion

Figure 7 demonstrates the recognition rate of the intra-feature fusions. It shows that the majority of intra-feature fusions perform better than the depth image. The y-direction surface normal holds the highest intra-feature fusion performance, especially in the LD distance measure, where it obtained the accuracy rate of 93.32%. Contrarily, Gaussian curvature yields the poorest performance when contrasted with the other intra-feature fusions. Hence, it can be concluded that Gaussian curvature is not able to perform well in 2.5D face recognition.

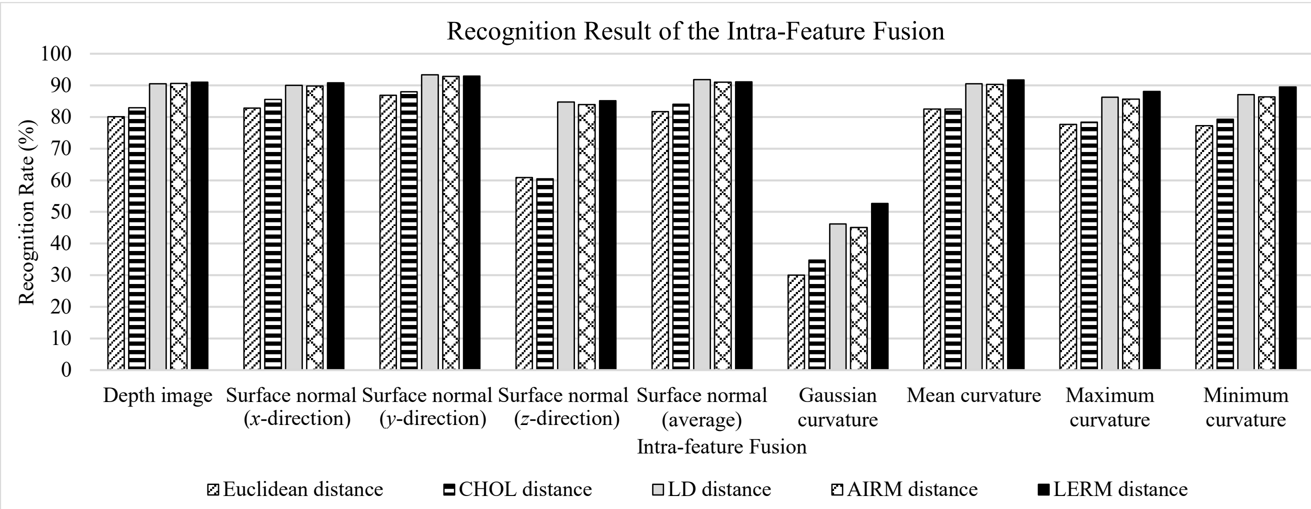


Figure 7. Recognition rate (%) of the intra-feature fusion

Additionally, the recognition rate for the surface normal in the x and y directions is almost identical. However, the accuracy rate for the surface normal in the z direction appeared to be the lowest among the three directions in the surface normal. Figure 7 has proven that summing the surface normal in the x, y, and z directions produces the best results. Moreover, single curvature (curvature minimum or curvature maximum) is not able to compete with the mean

http://doi.org/10.18080/jtde.v12n1.770

curvature (combination of curvature minimum and curvature maximum) as it contains the feature integration of these two curvatures.

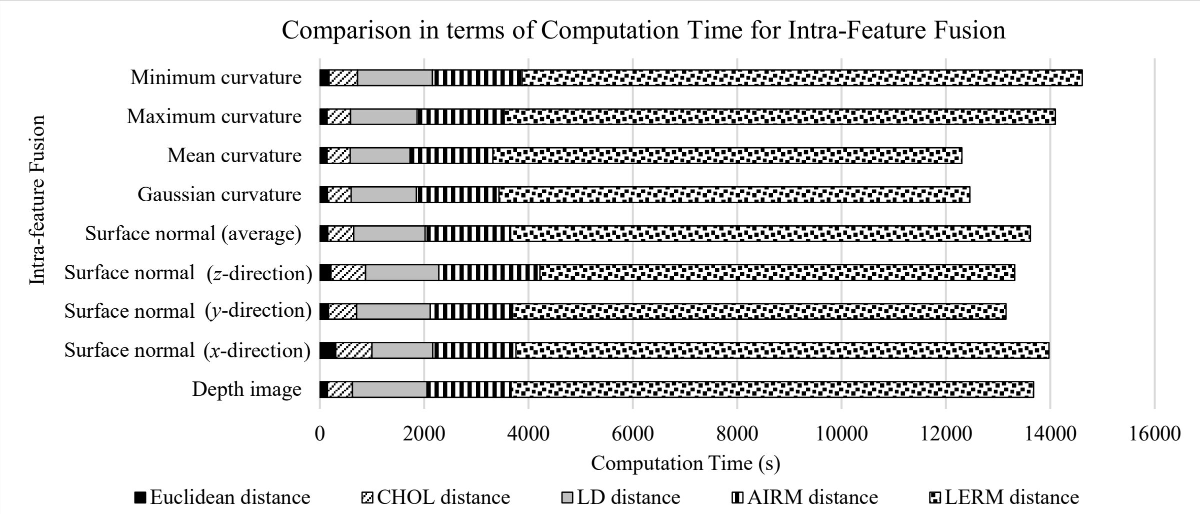


Figure 8 Computation times for intra-feature fusion

Figure 8 displays the computation time for each intra-feature fusion. The total computation time of each distance measure used to calculate the recognition rate in intra-feature fusion is almost identical with the inter-feature fusion. Figure 8 has proven that every intra-feature fusion utilised the shortest amount of time to compute the recognition rate in Euclidean distance. Still, unfortunately, the recognition rate for each intra-feature fusion is the lowest compared to the other distance measures, as shown in Figure 7. On the other hand, the computation time of the LERM distance to calculate the recognition rate is the longest among the other distance measures. However, the recognition rate of the majority of inter-feature fusions achieves superior performance in LERM distance, as illustrated in Figure 7.

Results of fusion integration

The recognition rate for each fusion integration is presented in Table 3. There are five different combinations of fusion integration that have been tested in this section.

Table 3. Recognition result of the fusion integration

Fusion Integration	Recognition Rate (%)				
	Euclidean	CHOL	LD	AIRM	LERM
Depth image	80.05	82.94	90.5	90.58	91.03
Max fusion:					
depth image + surface normal (y-direction)	84.57	86.24	90.83	89.84	90.69
depth image + mean curvature	83.7	84.29	90.63	90.36	90.72
surface normal (y-direction) + mean curvature	85.35	86.61	**93.41**	92.83	91.88
surface normal (y-direction) + surface normal (average)	85.35	87.52	93.08	93.08	91.88
surface normal (average) + mean curvature	81.43	82.54	89.48	89.24	90.36
Min fusion:					
depth image + surface normal (y-direction)	68.92	75.05	86.56	85.29	85.7

Fusion Integration	Recognition Rate (%)				
	Euclidean	**CHOL**	**LD**	**AIRM**	**LERM**
depth image + mean curvature	63.4	71.36	83.81	82.86	83.8
surface normal (y-direction) + mean curvature	84.57	85.14	**92.78**	92.12	92.74
surface normal (y-direction) + surface normal (average)	81.98	84.12	91.76	91.12	91.25
surface normal (average) + mean curvature	83.35	85.44	91.47	91.09	91.19
Max – Min fusion:					
depth image + surface normal (y-direction)	79.49	83.62	88.33	87.09	88.78
depth image + mean curvature	75.67	81.53	86.78	86.41	87.31
surface normal (y-direction) + mean curvature	89.68	89.37	**93.66**	93.35	93.51
surface normal (y-direction) + surface normal (average)	83.28	84.49	90.93	89.99	89.71
surface normal (average) + mean curvature	85.31	85.55	91.85	91.36	92.13
Sum Rule fusion:					
depth image + surface normal (y-direction)	80.16	83.17	91.07	90.72	90.6
depth image + mean curvature	80.96	83.43	90.68	90.77	90.86
surface normal (y-direction) + mean curvature	86.16	86.55	92.91	92.44	93.11
surface normal (y-direction) + surface normal (average)	84.70	85.51	**93.56**	92.84	92.80
surface normal (average) + mean curvature	83.20	84.83	91.22	91.07	91.36

Max fusion is produced by contrasting and obtaining the most outstanding values between the two aspects. The reason for choosing max fusion as one of the fusion methods in this section is its effectiveness in selecting the highest recognition rate among the data which is resistant to alterations and can obtain a specific facial trait. Table 3 illustrates that max fusion achieves the highest score with 93.41% in LD distance when it is utilised on the surface normal (y-direction) and mean curvature.

In addition, min fusion is created by contrasting and taking the lowest values between the two traits. Min fusion is selected in this section as it is simple to employ and is useful in retrieving the minimum score among the data to minimise the false positives. Compared to other fusion techniques, the min fusion method, which combines the mean curvature and the surface normal (y-direction), has the highest score of 92.78 %.

Furthermore, max-min fusion is created by subtracting the max and min feature fusions. Due to the effectiveness of max-min fusion, which merges the advantages of both the max and min fusion, max-min fusion is chosen in this section. The system can maintain reliability and accuracy by using max-min fusion, lowering the likelihood of false positives and increasing the system's resistance to noisy data. Among the different distance measures in Table 3, max-min fusion applied to the surface normal (y-direction) and mean curvature achieved the highest recognition rate, with 93.66% in LD.

http://doi.org/10.18080/jtde.v12n1.770

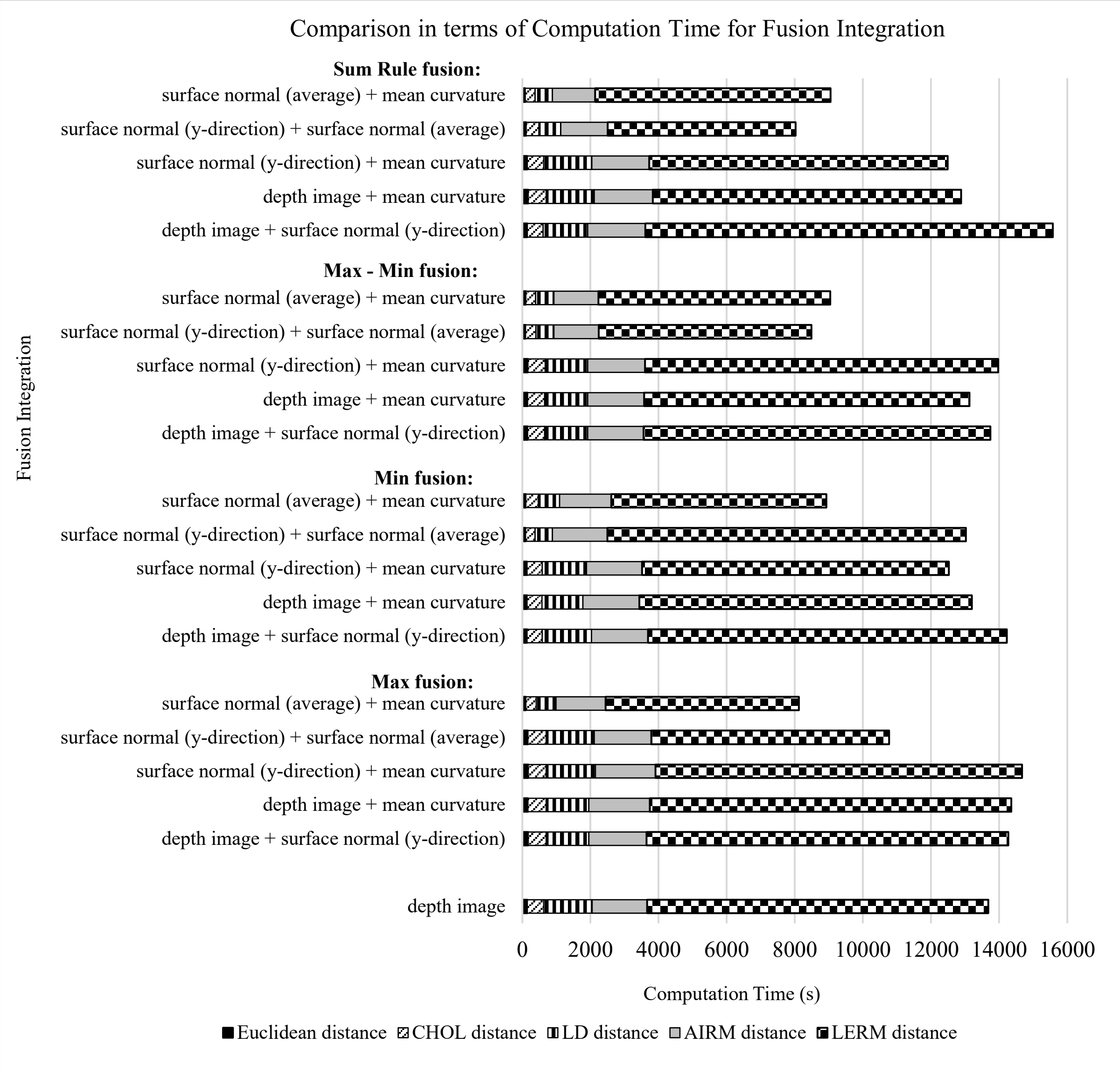


Figure 9. Computation times for fusion integration

Moreover, the sum rule, also referred to as sum fusion, is produced by integrating two distinct characteristics from different sources. The summing of the matching scores from different sources enhances the system's recognition rate, as it contains various facial characteristics from the sources. Due to these advantages, the sum rule is selected in this section. As can be observed from Table 3, contrasted to other distance measures, the sum rule that was implemented to the surface normal (y-direction) along with the surface normal (average) scored the second-best result in the LD distance measure with a score of 93.56%.

In short, it can be concluded that the surface normal (y-direction) + mean curvature, which is the combination of the first and second placed intra-feature fusions, achieves the highest scores in almost every type of fusion method including the max fusion, min fusion and max-min fusion. Besides that, most of the max, min, max-min fusion, and sum rule approaches outperform a single depth image in terms of the recognition rate. By combining two distinct characteristics as a new fusion data (fusion integration), rather than using just one feature, the efficiency of 2.5D face recognition can be improved.

The effectiveness of the fusion integration, including max, min, max-min, and sum rule fusion is shown in Figure 9 via the computation of time in seconds. The total computation time of each distance measure used to evaluate the accuracy rate in fusion integration is almost identical with the inter-feature fusion and the intra-feature fusion times. Figure 9 shows that, for every fusion integration, the recognition rate in Euclidean distance was computed in the least amount of time. However, as shown in Figure 9, the recognition rate for each fusion integration is the poorest contrasted with the other distance metrics. By contrast, the processing period of the LERM distance to compute the recognition rate is the longest among the distance measurements. Nevertheless, the accuracy rate of most of the fusion integrations obtains higher performance in LERM distance.

Comparison with other state-of-the-art methods

Table 4 compares the performance of the state-of-the-art results with the proposed method. From Table 4, our proposed method shows excellent performance compared to the state of the art.

Table 4. Comparison between the state of the art and the proposed method

Authors	Method used	Recognition Rate (%)
Kamencay *et al.* (2014)	CCA-PCA fusion	85%
Chong *et al.* (2014)	Textual image + depth image	80%
Chong *et al.* (2016)	Intra-feature fusion	90.87%
Liu *et al.* (2020)	Echo State Network (ESN) fusion	90%
The proposed method	Fusion integration	93.66%

In this paper, the proposed method is inspired by the work of Chong *et al.* (2014; 2016), which uses various fusion methods to increase the system accuracy rate. With this inspiration, the proposed method recommends using the fusion integration approach, which merges fused data once again to gain more powerful features to improve the system's performance. According to Table 4, the proposed method slightly outperforms the state of the art with an accuracy rate of 93.66%, proving the method's efficiency by using the fusion integration method.

Discussion

Figure 5, Figure 7, and Table 3 show the recognition results for the single depth image, inter-feature fusion, intra-feature fusion, and fusion integration (max fusion, min fusion, max-min fusion, and sum rule fusion). In this study, the max-min fusion used on the surface normal (y-direction) and mean curvature produced the highest accuracy rate of 93.66% compared to the

http://doi.org/10.18080/jtde.v12n1.770

other methods assessed. The depth image is utilised in this experiment as a baseline system for comparing the outcome regarding the feature fusion techniques employed. As can be observed from the experimental results, the accuracy rate of a single depth image is not optimistic, and almost all feature fusion techniques outperform a sole depth image. In summary, this experiment indicates that feature fusion outperforms a sole depth image in terms of recognition performance.

Although almost all feature fusion techniques can improve the performance of the 2.5D face recognition system, each of them still has some limitations. For example, the inter-feature fusion method that combines the features obtained from the 2D and depth data can perform well compared to a sole depth image. However, this feature fusion method that integrates multiple features increases the dimensions and complexity of the feature space, which requires more resources and time for processing.

On the other hand, the other features extracted from the depth image are called intra-feature fusion. Although most of the features derived from the depth image perform better than the single depth image itself, the recognition rate using the derived features is not satisfactory compared to the inter-feature fusion and fusion integration methods. In addition, the fusion integration method, in which the fused data is fused again, achieves the highest recognition rate among the other fusion methods. However, the performance of the fusion integration approach strongly depends on the quality of the features, and features containing noise may affect the effectiveness of the method.

Conclusions

The 2.5D face recognition system with the use of feature fusion approach is proposed in this paper. In the proposed method, the use of more than one type of feature by merging and mixing them up helps to boost the recognition rate of the system. Based on the experimental findings, in comparison to all the fusion techniques, this study proves that the max-min fusion employed on the mean curvature and surface normal (y-direction) has obtained the best accuracy rate at 93.66%. Moreover, nearly all fusion techniques outperform a single depth image in terms of accuracy rate. The computation time of the feature fusion methods relying on several geodesic distance measurements has also been studied in this paper. In the experiment, the best recognition rate was achieved by LD and LERM distance. In future work, more different fusion strategies will be studied and applied to enhance the 2.5D face recognition system.

Acknowledgements

The work presented in this paper was supported by Multimedia University through the IR Fund 2021 (MMUI/210029). A version of this paper was presented at the third International Conference on Computer, Information Technology and Intelligent Computing, CITIC 2023, held in Malaysia on 26–28 July 2023.

References

Bodla, N., Zheng, J., Xu, H., Chen, J.-C., Castillo, C., & Chellappa, R. (2017). Deep Heterogeneous Feature Fusion for Template-Based Face Recognition. *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 586–595. https://doi.org/10.1109/WACV.2017.71

Chong, L. Y., Ong, T. S., & Teoh, A. B. J. (2019). Feature fusions for 2.5D face recognition in Random Maxout Extreme Learning Machine. *Applied Soft Computing*, *75*, 358–372. https://doi.org/10.1016/j.asoc.2018.11.024

Chong, L. Y., Teoh, A. B. J., & Ong, T.-S. (2016). Range Image Derivatives for GRCM on 2.5D Face Recognition. In K. J. Kim & N. Joukov (Eds), *Information Science and Applications (ICISA) 2016* (Vol. 376, pp. 753–763). Springer Singapore. https://doi.org/10.1007/978-981-10-0557-2_73

Chong, L. Y., Teoh, A. B. J., & Ong, T.-S. (2017). PCA filter based covariance descriptor for 2.5D face recognition. *Proceedings of the 2017 International Conference on Biometrics Engineering and Application - ICBEA '17*, pp. 13–20. https://doi.org/10.1145/3077829.3077832

Chong, L. Y., Teoh, A. B. J., Ong, T.-S., & Chong, S.-C. (2014). 2.5D Face Recognition under Tensor Manifold Metrics. In C. K. Loo, K. S. Yap, K. W. Wong, A. T. Beng Jin & K. Huang (Eds), *Neural Information Processing* (Vol. 8836, pp. 653–660). Springer International Publishing. https://doi.org/10.1007/978-3-319-12643-2_79

Dutta, K., Bhattacharjee, D., Nasipuri, M., & Krejcar, O. (2021). Complement component face space for 3D face recognition from range images. *Applied Intelligence*, *51*(4), 2500–2517. https://doi.org/10.1007/s10489-020-02012-8

Fang, L., He, N., Li, S., Plaza, A. J., & Plaza, J. (2018). A New Spatial–Spectral Feature Extraction Method for Hyperspectral Images Using Local Covariance Matrix Representation. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(6), 3534–3546. https://doi.org/10.1109/TGRS.2018.2801387

Kamencay, P., Hudec, R., Benco, M., & Zachariasova, M. (2014). 2D-3D Face Recognition Method Basedon a Modified CCA-PCA Algorithm. *International Journal of Advanced Robotic Systems*, *11*(3), 36. https://doi.org/10.5772/58251

Kaur, H., Koundal, D., & Kadyan, V. (2021). Image Fusion Techniques: A Survey. *Archives of Computational Methods in Engineering*, *28*(7), 4425–4447. https://doi.org/10.1007/s11831-021-09540-7

Liu, S., Chen, D., Chen, Z., Ru, C., & Pang, M. (2020). *Research on Face Recognition Technology Based on ESN Multi Feature Fusion*. https://doi.org/10.3966/160792642020092105029

Nhat, H. T. M., & Hoang, V. T. (2019). Feature fusion by using LBP, HOG, GIST descriptors and Canonical Correlation Analysis for face recognition. *2019 26th International Conference on Telecommunications (ICT)*, 371–375. https://doi.org/10.1109/ICT.2019.8798816

Pang, Y., Yuan, Y., & Li, X. (2008). Gabor-Based Region Covariance Matrices for Face Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, *18*(7), 989–993. https://doi.org/10.1109/TCSVT.2008.924108

Phillips, P. J., Flynn, P. J., Scruggs, T., Bowyer, K. W., Chang, J., Hoffman, K., Marques, J., Min, J., & Worek, W. (2005). Overview of the Face Recognition Grand Challenge. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, *1*, 947–954. https://doi.org/10.1109/CVPR.2005.268

Talab, M. A., Qahraman, N. A., Aftan, M. M., Mohammed, A. H., & Ansari, M. D. (2022). Local Feature Methods Based Facial Recognition. *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 1–5. https://doi.org/10.1109/HORA55278.2022.9799910

Tosranon, P., Sanpanich, A., Bunluechokchai, C., & Pintavirooj, C. (2009). Gaussian curvature-based geometric invariance. *2009 6th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, 1124–1127. https://doi.org/10.1109/ECTICON.2009.5137242

Tuzel, O., Porikli, F., & Meer, P. (2006). Region Covariance: A Fast Descriptor for Detection and Classification. In A. Leonardis, H. Bischof, & A. Pinz (Eds), *Computer Vision – ECCV 2006* (Vol. 3952, pp. 589–600). Springer Berlin Heidelberg. https://doi.org/10.1007/11744047_45

Vezzetti, E., & Marcolin, F. (2012). 3D human face description: Landmarks measures and geometrical features. *Image and Vision Computing*, *30*(10), 698–712. https://doi.org/10.1016/j.imavis.2012.02.007

Vezzetti, E., Marcolin, F., & Fracastoro, G. (2014). 3D face recognition: An automatic strategy based on geometrical descriptors and landmarks. *Robotics and Autonomous Systems*, *62*(12), 1768–1776. https://doi.org/10.1016/j.robot.2014.07.009