

International Journal of Intelligent Systems Technologies and Applications

ISSN online: 1740-8873 - ISSN print: 1740-8865
<https://www.inderscience.com/ijista>

Face recognition using threshold string representation matching under unconstrained conditions

Manisha Kumari Meena, Hemant Kumar Meena, Srikanta Patnaik, Gurbinder Singh

DOI: [10.1504/IJISTA.2025.10073419](https://doi.org/10.1504/IJISTA.2025.10073419)

Article History:

Received:	14 June 2024
Last revised:	15 November 2024
Accepted:	16 December 2024
Published online:	30 September 2025

Face recognition using threshold string representation matching under unconstrained conditions

Manisha Kumari Meena* and
Hemant Kumar Meena

Department of Electrical Engineering,
MNIT Jaipur,
Jaipur, India
Email: 2019ree9078@mnit.ac.in
Email: hmeena.ee@mnit.ac.in
*Corresponding author

Srikanta Patnaik

Department of Computer Science Engineering,
IIMT,
Bhubaneswar, India
Email: patnaik_srikanta@yahoo.co.in

Gurpinder Singh

Department of Electrical Engineering,
MNIT Jaipur,
Jaipur, India
Email: 2019ree9142@mnit.ac.in

Abstract: This paper introduces a new approach called the threshold string representation matching (TSRM) algorithm, designed to enhance face identification and verification in scenarios by partial occlusions like sunglasses, scarves, and caps. The TSRM algorithm innovatively manipulates pixel features by calculating the difference between the centre pixel and its neighbours rather than relying on the traditional feature compression method (CSM-RL), where pixel difference is taken row-wise. This technique also improves the robustness of facial expression recognition by focusing on localised spatial relationships within the image patches. Our algorithm provides a novel feature-matching algorithm which effectively matches the prob image with the gallery image in the occlusion condition. We evaluated our proposed TSRM approach on two public face datasets: the AR dataset, which includes natural occlusion, and the ORL dataset, for synthetic occlusion.

Keywords: partial occlusion; face recognition; the difference between patches; k-means clustering; threshold string representation method; TSRM.

Reference to this paper should be made as follows: Meena, M.K., Meena, H.K., Patnaik, S. and Singh, G. (2025) 'Face recognition using threshold string representation matching under unconstrained conditions', *Int. J. Intelligent Systems Technologies and Applications*, Vol. 23, No. 3, pp.319–336.

Biographical notes: Manisha Kumari Meena received her BTech in Electrical Engineering from the Birla Institute of Technology (BIT) in 2014. She received her Master's in Power System from the National Institute of Technology, Kurukshetra, India, in 2017. Recently, she has been a research scholar at the Malaviya National Institute of Technology, Jaipur. Her research interests include artificial intelligence, soft computing, machine learning, digital image processing, and computer vision.

Hemant Kumar Meena is an Associate Professor in the Department of Electrical Engineering at the MNIT Jaipur. He received his BTech and MTech from the IIT Delhi. His research areas include machine learning application, signal and image processing, and power system optimisation.

Srikanta Patnaik is the Founder and the Director of IIMT, Bhubaneswar. He has completed his MBA from Sambalpur University in 1991 and received his PhD in Engineering on Computational Intelligence from Jadavpur University, India in 1999. He supervised 25 PhD theses and more than 60 Master theses in the area of computational intelligence, soft computing applications and re-engineering. He has published around 100 research papers in international journals and conference proceedings.

Gurpinder Singh received his BTech in Electrical Engineering from the IIT Dhanbad. He received his MTech in Power System from the MNIT Jaipur. Recently, he is a research scholar at MNIT Jaipur and his areas of interest are machine learning artificial intelligence data analysis in power system.

This paper is a revised and expanded version of a paper entitled 'Partial Face identification using local feature extraction algorithm on different classifiers' presented at 2023 International Conference on Computer, Electronics & Electrical Engineering & their Applications (IC2E3), NIT Uttarakhand, 9 June 2023.

1 Introduction

Face recognition is a biometric technology that utilises facial features to identify individuals. This biometric system is employed to verify personal identities (Zeng et al., 2021). Over recent years, face recognition has become integral to various applications, including forensic science, driving license and passport verification, locating missing persons, surveillance systems, social networks, and access control. Despite substantial advancements in computer vision research, face recognition rates face challenges due to common issues such as facial expressions (Wallace et al., 2006), pose variations, lighting changes, partial occlusion, ageing, and appearance alterations in unconstrained scenarios. In this paper, we mainly focus on the occlusion problem (Liu et al., 2023). In practical situations, human faces may be occluded by various items such as scarves, hats, sunglasses, masks or even body parts like hands and hair. Thus, managing occlusion is essential for secure and accurate analysis in surveillance applications. The sample images of faces with partial occlusion are illustrated in Figure 1. Addressing the issue of partial

occlusion requires an effective, reliable, and secure face recognition system (Kortli et al., 2020; Bommidi and Sundaramurthy, 2021).

Figure 1 Sample images of a person under occlusion condition (see online version for colours)



Effective face recognition in the presence of occlusions requires prior knowledge about the occlusion. Techniques robust to occlusion necessitate comprehensive training datasets and tests, complicating the recognition process. Over the past three decades, numerous facial recognition (FR) techniques have been developed, yielding high accuracy based on the face recognition pipeline, which includes holistic and part-based (Gupta et al., 2021) approaches. Holistic methods perform exceptionally well with frontal faces in optimal lighting conditions, without any obstructions or changes in pose. Nevertheless, these strategies struggle when objects like scarves, sunglasses, or other facial adornments partially hide faces. The main cause of reduced accuracy in these situations is the heightened similarity between different classes and the marked similarity within the same class due to occlusions. Conversely, part-based approaches effectively handle occlusion and variations by employing local matching methods to compare gallery and probe images (Wei et al., 2014).

Our proposed method focuses on robust feature extraction using region-based matching approaches. In this method, face images are divided into patches of the same size, and features are extracted by subtracting the neighbouring pixel from the centre pixel. The dimensionality of these components is reduced using the K-means clustering machine learning technique, creating approximately 250 cluster centres that are then represented as sequences of clusters. These cluster sequences are subjected to a threshold-matching representation method to align the gallery image with the probe image. Our proposed technique has been compared with advanced methods and demonstrates superior accuracy. Face recognition accuracy is significantly impacted by challenges such as the rate and density of occlusion. While various techniques provide good accuracy up to a 30% occlusion rate, their accuracy drastically declines, or they fail beyond this threshold. Our proposed method, however, can handle higher occlusion rates with improved results compared to existing techniques (Chen et al., 2023; Huang et al., 2012). The key contributions of this work are outlined as follows:

- 1 Introducing the threshold string representation matching (TSRM) face recognition technique demonstrates robustness against illumination, facial expression and occlusion.
- 2 Face features are extracted by measuring the distance between neighbouring and central pixels to detect occlusion effects. Unlike existing methods that use overlapping patches, this paper employs non-overlapping differences of pixels (DoPs), thereby avoiding an increase in vector size.

- 3 The use of the K-means clustering algorithm to represent DoPs as cluster centres, effectively reducing the dimensions of the feature vectors.
- 4 Various training-based machine learning classifiers, including SVM, random forest (RF), and linear regression, need many images to train the data. The proposed technique classifies by applying a threshold to the normal distribution of cluster centre faces, effectively working on a single image in training data.

The article is organised as follows: Section 2 reviews the literature relevant to our study. Section 3 outlines the methodology we propose. Section 4 discusses the results of experiments conducted on various face databases. Finally, Section 5 concludes the paper.

2 Related work

Three primary strategies are employed for recognising faces with occlusions: identifying faces despite occlusions, detecting and removing occluded sections, and reconstructing occluded areas. The holistic and part-based approach focuses on recognising faces under occlusion.

Classic holistic and part-based methods such as eigenfaces (PCA) (Turk and Pentland, 1991), ICA (Yu and Yang, 2001), Fisherfaces (LDA) (Belhumeur et al., 1997; Bartlett et al., 2002), and hidden Markov models (HMM) (Nefian and Hayes, 1998) are particularly vulnerable to partial occlusions, which compromise their effectiveness in occluded face recognition scenarios. In response, several alternative strategies have been devised. Methods based on local features (Tan et al., 2005; Kanan et al., 2008; Tan et al., 2009; Jia and Martinez, 2009; Weng et al., 2013; Wei et al., 2014; Weng et al., 2016) exploit local characteristics of the face that remain unaffected by occlusions. Meanwhile, methods based on sparse representation classification (Wright et al., 2009; Zhou et al., 2009; Liao and Jain, 2011; Zhang et al., 2011; Jia et al., 2012; He et al., 2014; Yang et al., 2014; Vu and Monga, 2017; Su et al., 2018) address occlusions through a reconstructive process.

For instance, self-organising maps (SOM) (Tan et al., 2005) utilise SOM maps to recognise occluded faces. The AWPPZMA approach (Kanan et al., 2008) employs a patch-based strategy that adaptively weights patch pseudo-Zernike moment features to manage occlusions. The partial distance (PD) technique (Tan et al., 2009) detects similarities in partially visible faces, and partial support vector machines (PSVM) (Jia and Martinez, 2009) leverage local features to address occlusions. The metric learned extended robust point set matching (MLERPM) (Weng et al., 2013) employs local feature matching with SIFT and SURF features to handle partial occlusions. The dynamic image-to-class warping (DICW) (Wei et al., 2014) uses the dynamic time warping (DTW) algorithm for recognising faces with partial occlusions. Additionally, the robust point set matching method (RPSM) (Weng et al., 2016) integrates SIFT, SURF, and scale-invariant LBP features for identifying faces in partially visible images.

Reconstruction-oriented approaches like the SRC (Wright et al., 2009) apply sparse representation theory for face recognition but necessitate pixel-perfect alignment of face images. To address this issue, several models have been introduced. The SRC-MRF model (Zhou et al., 2009) was designed to handle continuous occlusions effectively. Meanwhile, the MKD-SRC technique (Liao and Jain, 2011) offers an alignment-free method for facial representation by integrating multi-task sparse representation with

multi-keypoint descriptors. The CRC-RLS method (Zhang et al., 2011) employs L2-norm regularisation in its classification framework. The lstruct system (Jia et al., 2012) specifically aims to tackle challenges related to occlusions, shadows, and alignment discrepancies. The HQM model (He et al., 2014) addresses robust sparse representation problems. The FDDL approach (Yang et al., 2014) leverages Fisher discrimination for facial image classification within an SRC-based framework. Additionally, the LRSDL model (Vu and Monga, 2017) presents a fast, low-rank shared dictionary learning technique. Lastly, the SCIP method (Su et al., 2018) has demonstrated improved capabilities in recognising faces with occlusions. Lastly, the group sparse representation-based (GSR) (Fritz et al., 2019) method proposes an optimisation algorithm for managing occlusions within an SRC framework. Despite these innovations, SRC-based techniques often require a substantial number of training samples per individual, which limits their utility in real-time face recognition applications where training data is scarce.

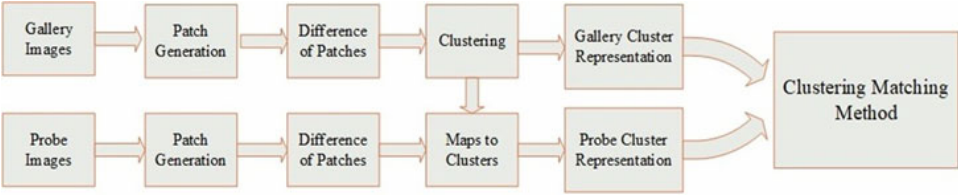
Many face recognition technologies (Chen and Gao, 2013, 2010; Gao and Leung, 2002; Mario, 2014) utilise string matching techniques. The ensemble string matching (Chen and Gao, 2013, 2010) conceptualises the face as a collection of attributed strings, each corresponding to a facial component, and performs string-to-string comparisons to recognise partially occluded faces. Another method involves representing human face profiles (Gao and Leung, 2002) as a series of strings, each depicting an ordered line segment of a face profile. However, this approach is less effective when the profile itself is occluded.

The approximate string matching method (Mario, 2014) captures facial features like eye width and distances between eyes and nose and encodes them in a sequence of strings for recognition purposes. This paper introduces the threshold string representation method (TSRM), which matches gallery and probe images by applying a threshold on the string sequence generated by replacing cluster centres with features depicted by DoPs. This method is compared with CSM-RL, which uses pixel differences within the same patch row-wise. Our approach, however, calculates the pixel difference between neighbouring and central pixels within the same patches.

3 Proposed work

The proposed work is outlined in a flow chart shown in Figure 2 and divided into three main parts. Image processing, feature extraction and feature matching. Image pre-processing is essential for converting raw face data into an appropriate format for analysis. This process includes steps such as face detection, normalisation, image scaling, and face alignment. Initially, the dataset is split into two sets: gallery images and probe images. Each image in these sets is then divided into blocks of equal length, called patches. The differences between the centre and neighbour pixels of the same patches are considered the feature vectors for both the gallery and probe images. The K-means clustering algorithm is applied to the DoPs, replacing these clusters with their centres. After that, a feature selection algorithm is used for person classification. A detailed analysis of the proposed flow chart is provided in Figure 2.

Figure 2 Proposed workflow chart (see online version for colours)



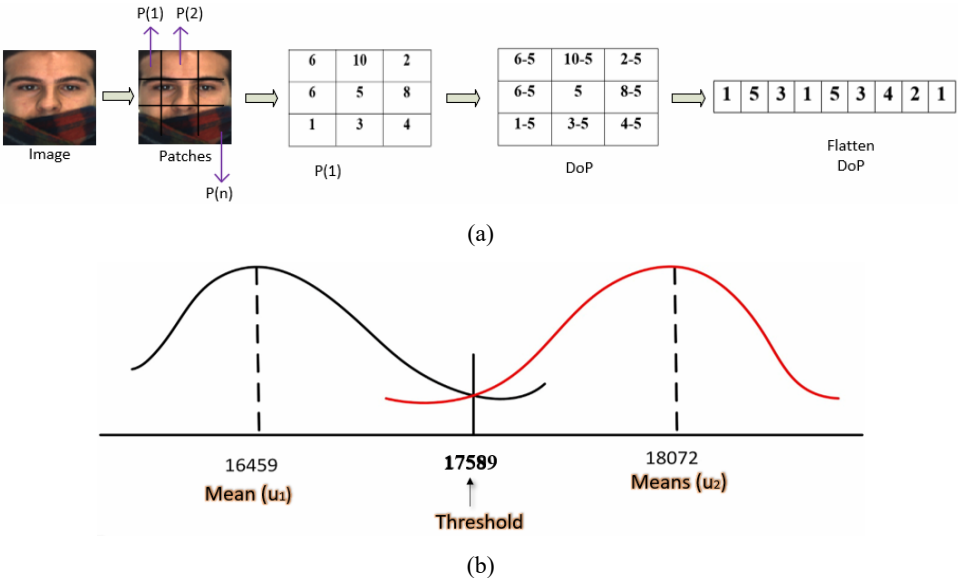
3.1 Feature extraction

Feature extraction is conducted in two main steps: calculating the DoPs and performing clustering. Each face image, sized $m \times m$, is divided into $n \times n$ patches of equal size. The features are extracted by determining the contrast within these patches. The number of pixels in each patch constitutes the facial features of an image. After computing the differences within the same patches, unique attributes are retained while redundant parts are discarded. These differences are then represented as cluster centres using the K-means clustering algorithm, which reduces dimensionality and forms string sequences.

3.1.1 Difference of pixels

The process of creating DoPs is illustrated in Figure 3(a). Patch $P(1)$ is a 3×3 matrix containing nine pixel values, representing one image patch. The central pixel of this patch is compared with its neighbouring pixels. Similarly, another patch, $P(2)$, with different pixel values but the same matrix size as $P(1)$, is used. The DoPs are generated by calculating the difference of pixels of each patch from $P(1)$ to $P(n)$.

Figure 3 (a) Features are extracted by taking the difference between the centre pixel and the neighbour pixel within the same pattern, (b) Represent the classification threshold value (see online version for colours)



3.1.2 Clustering

DoPs capture the variations across image patches. These patches are converted into a linear sequence by arranging pixels from left to right and top to bottom.

Algorithm 1 K-mean clustering

Input: $Y = \{Y_1, Y_2, Y_3, \dots, Y_N\}$ be the set of gallery image DOP vectors. $C = \{C_1, C_2, C_3, \dots, C_s\}$ be the set of cluster centres. Here, n denotes the number of DOPs in a gallery, and s represents the number of cluster centres.

Output: cluster centres with r gallery image DOPs.

- Assign s clusters with related DOPs.
 - Find the local distance between each image DOP vector and cluster centres.
 - Select the cluster centre with the minimum local distance and assign it to DOP.
 - Compute the new cluster centre using the $c_i = \left(\frac{1}{R_i}\right) * \text{sum}(y_i)$, where y_i are DOPs assigned to a specific cluster centre, R represents the number of DOPs in the i^{th} cluster.
 - Find the distance difference between each image DOP vector and new cluster centres.
 - Stop the process when the supposed number of iterations is reached.
-

Following this transformation, the K-means clustering algorithm is utilised to organise these DoP sequences into distinct cluster centres. The process of employing K-means clustering is represented in Algorithm 1. After the clustering algorithm is applied, the DoP sequences are transformed into clusters. The cluster sequence for the training images is referred to as the gallery cluster/string sequence. The string sequence for the probe images is termed the probe string/cluster sequence. The algorithm of the clustering centre replaced by string sequence is represented in Algorithm 2.

Algorithm 2 Cluster representation of gallery/probe images

Input: Cluster centres and gallery/probe image of S and K difference within the patches, respectively.

Output: Cluster representation of a gallery/probe image.

Procedure: Cluster representation

- Retrieve the cluster labels given to a gallery image's DoPs.
- Represent a gallery image as a single sequence using the cluster labels obtained in step 1.
- Find the local distance $E(i, j)$ between probe DoPs and the cluster centres.
- Select the cluster centres with minimum distance ($\min E(i, j)$).
- Get the labels of selected cluster centres.
- Represent a probe image as a cluster sequence using the labels of cluster centres obtained in step 5.

end procedure

3.2 Feature matching

When the difference of patch vectors replaces the cluster centres, this new vector form is called cluster sequence. The cluster sequence for gallery images is considered a gallery

cluster sequence set (GCS); probe images are regarded as a probe cluster sequence set (PCS). The distance function is applied between each image in GCS and PCS, and each person's mean and standard deviation is found. The normal distribution algorithm is modified to find the mean threshold between similar and dissimilar images. The normal distribution formula (Ma et al., 2022) is shown below:

$$y_i = \frac{e^{\frac{1}{2}\left(\frac{x-\mu_i}{\sigma}\right)^2}}{\sigma_i\sqrt{2\pi}} \quad (1)$$

Equation (1) shows the basic formula of normal distribution. This paper y_1 is considered the normal distribution of similar persons and y_2 is considered the normal distribution of different persons. The threshold is applied between the string sequences of the gallery image and the probing picture sequence. When:

$$\frac{e^{\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}}{\sigma_1\sqrt{2\pi}} = \frac{e^{\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}}{\sigma_2\sqrt{2\pi}} \quad (2)$$

where μ_1 and μ_2 shows the mean of the gallery and test images' string representation. σ_1 and σ_2 shows the standard deviation string representation of the gallery and probe images, and x represents the threshold value.

$$\ln\left(\frac{e^{\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}}{\sigma_1\sqrt{2\pi}}\right) = \ln\left(\frac{e^{\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}}{\sigma_2\sqrt{2\pi}}\right) \quad (3)$$

$$-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \ln(\sigma_1\sqrt{2\pi}) = -\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2 + \ln(\sigma_2\sqrt{2\pi}) \quad (4)$$

$$\begin{aligned} -\frac{1}{2}\left(\frac{x^2}{\sigma_1^2} + \frac{\mu_1^2}{\sigma_1^2} - \frac{2x\mu_1}{\sigma_1^2} - \frac{x^2}{\sigma_2^2} - \frac{\mu_2^2}{\sigma_2^2} + \frac{2x\mu_2}{\sigma_2^2}\right) &= -2\ln\left(\frac{\sigma_1\sqrt{2\pi}}{\sigma_2\sqrt{2\pi}}\right) \\ x^2\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right) + 2x\left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2}\right) &+ \left(\frac{\mu_1^2}{\sigma_1^2} + \frac{\mu_2^2}{\sigma_2^2}\right) + 2\ln\left(\frac{\sigma_1\sqrt{2\pi}}{\sigma_2\sqrt{2\pi}}\right) = 0 \end{aligned} \quad (5)$$

The modified equation of normal distribution for finding the threshold is shown below:

$$\begin{aligned} x^2\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right) + 2x\left(\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2}\right) + \ln\left(\frac{\sigma_2^2}{\sigma_1^2}\right) + \left(\frac{\mu_1^2}{\sigma_1^2} + \frac{\mu_2^2}{\sigma_2^2}\right) &= 0 \\ x &= b \pm \frac{\sqrt{(b^2 - 4ac)}}{2a} \end{aligned}$$

Here

$$b = \left(\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2}\right), a = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right), c = \ln\left(\frac{\sigma_2^2}{\sigma_1^2}\right) + \left(\frac{\mu_1^2}{\sigma_1^2} + \frac{\mu_2^2}{\sigma_2^2}\right) \quad (7)$$

In equation (7), the value of x represents the classification threshold.

In Figure 3(b), the red line indicates the distance values for different individuals, while the black line represents the distance values for the same individual. The intersection of these lines denotes the classification threshold. An evaluation matrix is used to assess the accuracy determined by this threshold value. If the distance value for the same individual is less than the threshold, it is classified as a true positive; otherwise, it is a true negative. Conversely, if the distance value for a different individual is less than the threshold, it is also considered a true positive; otherwise, it is a true negative. Using these criteria, results such as the confusion matrix, F1-score, and accuracy are evaluated. The proposed work is illustrated in Algorithm 3.

Algorithm 3 Detailed procedure for proposed method

Step 1	Segment training images into patches of uniform size.
Step 2	Assess the difference between adjacent and central pixels in each patch.
Step 3	Linearise the patch pixels, arranging them from left to right and top to bottom to form a continuous sequence.
Step 4	Implement K-means clustering to establish 250 cluster centres for each individual, using feature vectors, and compute the average of data points near each cluster centre.
Step 5	Transform the linear pixel sequences into cluster sequences via K-means clustering, referred to as the gallery cluster sequence.
Step 6	Apply steps 1, 2 and 3 to the test images.
Step 7	Align cluster centres from the training images with those of the test images, creating what is known as the probe cluster sequence.
Step 8	Compute the mean values for cases of similar and dissimilar individuals.
Step 9	Establish a threshold to differentiate between similar and dissimilar persons.
Step 10	Evaluate the degree of similarity between gallery and probe cluster sequences.

4 Experiment setup

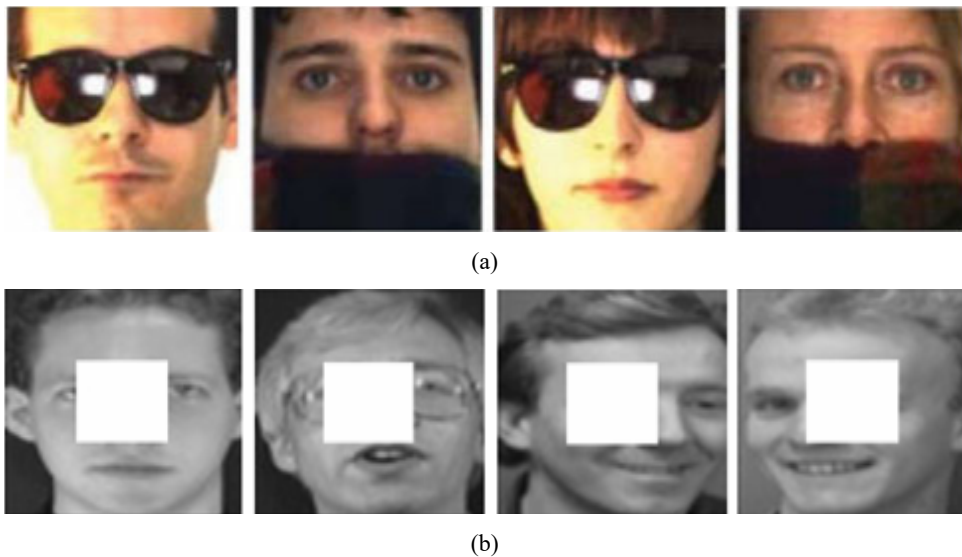
In this part, we perform experimental evaluations using different facial databases, particularly the Aleix Martinez and Robert Benavente (AR) and Olivetti Research Laboratory (ORL) datasets, to evaluate the performance of our proposed model under uncontrolled conditions. These conditions considered illumination, facial expressions, and partial face occlusion. The AR dataset includes natural occlusions like sunglasses and scarves, while the ORL dataset includes synthetic occlusions like back or rectangular boxes. In the ORL dataset, the synthetic occlusion percentage varies from 10% to 50%, as depicted in Figures 4(a) and 4(b). The experimental settings for both datasets are consistent with that used in existing work (Liu et al., 2023). Our proposed method is applied to these datasets to enhance the performance of FR systems compared to existing methods. Our proposed TSRM method was evaluated for both natural and synthetic occlusion using the AR and ORL datasets, respectively. For the comparison with the existing method, we resized the image to 64×64 size and considered less than six images in the gallery set.

4.1 Datasets

The AR dataset, as described in Gross (2005), includes approximately 4,000 colour face images from 126 individuals, comprising 70 men and 56 women. Each individual has 26 face images in the dataset, with 14 images depicting various facial expressions and lighting conditions without occlusions and the remaining 12 images showcasing two types of occlusions: sunglasses and scarves. Figure 3 presents sample images from the AR dataset.

The ORL database, detailed in Boiman et al. (2008), contains images of 41 individuals, each considered a subject or class, with ten greyscale images per subject, a total of 410 face images. These images exhibit different pose variations from left to right with minor facial expressions. We introduced synthetic occlusions by randomly generating rectangular blocks on each face image, varying the block size to test the model's effectiveness, with occlusion ratios ranging from 20% to 50%. This dataset also includes pose variations along with the random rectangular block occlusions, as shown in Figure 4(b).

Figure 4 Sample images of AR and ORL dataset, (a) AR dataset (Gross, 2008) (b) ORL dataset (Boiman et al., 2008) (see online version for colours)



4.2 Results on partial occlusion

Most existing studies present results only when the gallery set contains non-occluded face images. To thoroughly assess model efficiency in both natural and synthetic occlusion scenarios, our results are categorised into three conditions:

- 1 When gallery images are non-occluded, and probe images are occluded.
- 2 When gallery images are occluded, and probe images are non-occluded.

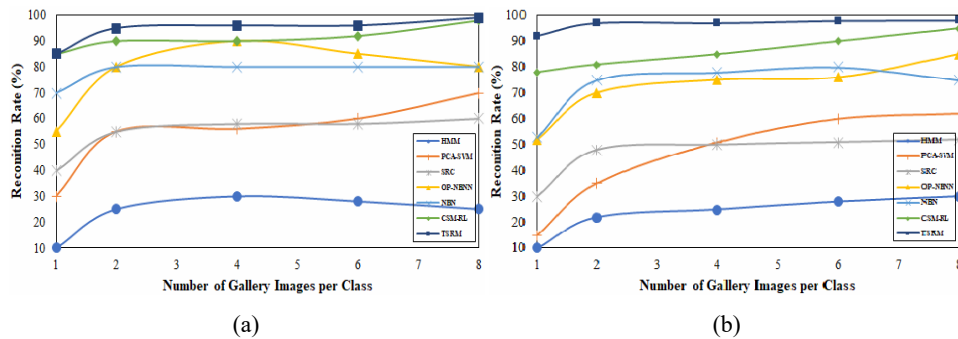
3 When both gallery and probe images are occluded.

These conditions allow us to evaluate the robustness of our proposed TSRM method under different occlusion scenarios. The performance metrics and accuracy across different scenarios are further elaborated in the following sections. These conditions allow us to evaluate the robustness of our proposed TSRM method under different occlusion scenarios. The performance metrics and accuracy across different scenarios are further elaborated in the following sections.

4.2.1 When gallery images are non-occluded, and probe images are occluded

Our proposed method was compared with existing state-of-the-art techniques, including HMMs (Nefian and Hayes, 1998), principal component analysis with support vector machine (PCA-SVM) (Boiman et al., 2008), sparse representation classifier (SRC) (Wright et al., 2009), original patch Naive Bayes nearest neighbour (OP-NBNN) (Chang and Lin, 2011), Naive Bayes nearest neighbour (NBNN) (Chang and Lin, 2011), and dynamic image class warping (DICW) (Chen et al., 2023), for AR and ORL datasets.

Figure 5 Accuracy comparison with existing state-of-the-art methods for the AR dataset, (a) sunglasses occlusion (b) scarf occlusion (see online version for colours)



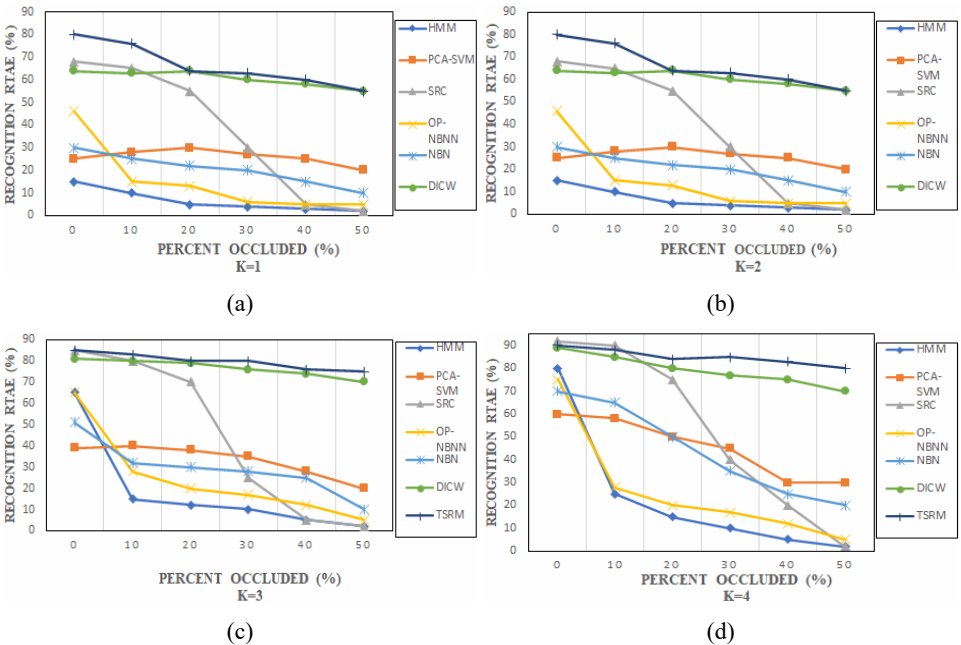
For AR dataset: as shown in Figures 5(a) and 5(b), our TSRM method provides better performance than the existing methods, such as HMM, PCA-SVM, SRC, NBN and OP-NBN. The HMM perform very well with zero percent occlusion and when there are many images in the gallery set. The HMM method provides the worst output under occlusion. However, NBNN and DICW methods, which are based on the difference within patches, do not consider pixel differences. SRC is a reconstruction-based method and accuracy degrades when the occlusion ratio is large. Our TSRM shows better performance compared to reconstruction-based methods and other state-of-the-art methods. Table 1 represents the comparison of accuracy with other existing methods, such as CSM-RL, DICW, MLERPM, Istruct, CRC-RLS, P-CRC-RLS, LRC, P-SRC, SCIP and RRC-I2. Our TSRm provide the best accuracy compare with other existing state-of-the-art methods. In all tables, bold values represent the highest accuracy.

Table 1 Comparison of accuracy (%) on different methods for sunglasses and scarves occlusion

Methods	Sunglasses	Scarves
TSRM	100	99.68
CSM-RL (Bommidi and Sundaramurthy, 2021)	98.5	96.6
DICW (Wei et al., 2014)	99.5	97.5
MLERPM (Weng et al., 2013)	98.0	97.0
Istruct (Jia et al., 2012)	99.5	87.5
CRC-RLS (Zhang et al., 2011)	68.5	90.5
P-CRC-RLS (Zhang et al., 2011)	91.5	95.0
LRC (Naseem et al., 2010)	96.0	26.0
P-SRC (Wright et al., 2009)	97.5	93.5
SCIP (Su et al., 2018)	100	99.0
RRC-I2 (Yang et al., 2013)	99.5	96.5

Our proposed method is compared with the other existing state-of-the-art methods, such as Naseem et al. (2010), Wen (2012), Zhang et al. (2010), Lu et al. (2010) and BoW (Zhao et al., 2015) for the ORL dataset as represented in Table 2. Our TSRM method provides the best accuracy for all the methods but not for the BoW method.

Figure 6 Recognition Rate when the number of images in gallery set changes, (a) $K = 1$, one image in gallery set (b) $K = 2$, two image in gallery set (c) $K = 3$, three image in gallery set (d) $K = 4$, four image in gallery set (see online version for colours)



For ORL dataset: the ORL dataset contains the synthetic occlusion where the occlusion ratio can be varied from 10% to 50% as shown in Figure 4(b). The ORL dataset generates synthetic occlusion on the human face by imposing a black-and-white rectangular block. In this paper, we are imposing white rectangular boxes on human faces. Most of the existing methods have accuracy degradation when the occlusion size increases or if there is only one image in the gallery set. In contrast, our proposed method performs very well with a single image per class, even with a high occlusion ratio. Figure 6 shows the accuracy comparison with the existing methods when changing the number of images (k) in the training set. The gallery set or training set includes a minimum of one image ($K = 1$) and a maximum of eight images ($K = 8$) per class. Here, K represents the number of images in the gallery set.

4.2.2 When gallery images are occluded and probe images are non-occluded

In most of the literature survey, authors do not consider the part where the training set contains the occluded images when features are extracted using traditional methods and deep learning models. However, deep learning-based methods need a large number of images in the training set to train the occluded image in the training set. Here, our model efficiently works when the number of images in the training set is less, around 1 to 5 only as well as occluded present in the training set

Table 2 Comparison of our proposed method with state-of-the-art methods on ORL dataset in terms of rectangular black box occlusion

Methods	Recognition rate (%)
TSRM	97.66
Yang et al. (2013)	95.5
Chang and Lin (2011)	97.25
Naseem et al. (2010)	Not evaluated
Yang et al. (2013)	92.02
BoW (Zhao et al., 2015)	99.50

Figure 7 Accuracy comparison with the existing methods when gallery images are occluded, and probe image is unoccluded for AR and ORL dataset, (a) for AR dataset (b) for AR dataset (see online version for colours)

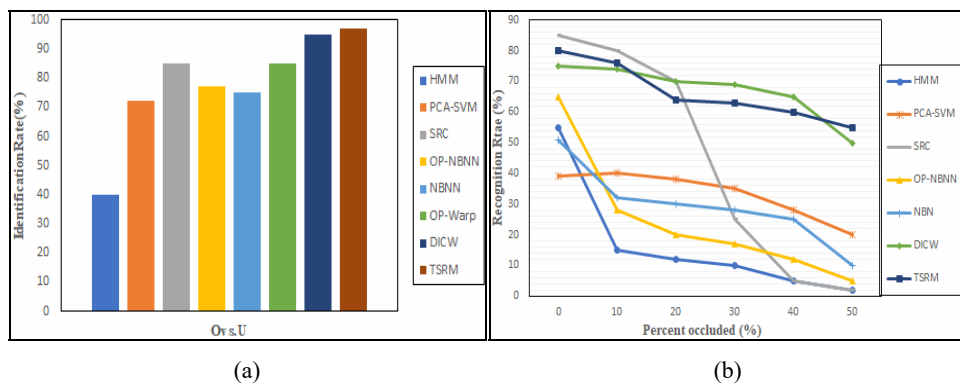


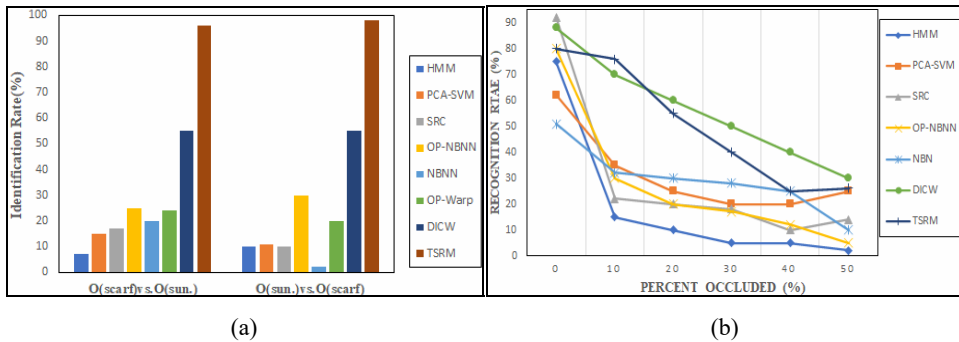
Figure 7 displays the accuracy comparison with the existing and our TSRM methods for the AR and ORL datasets. Figure 7(a) represents the recognition rate of our method. TSRM (brown colour in the bar chart) provides the best results (97%) for the AR dataset. Similarly, Figure 7(b) shows the accuracy comparison between our TSRM and the existing method on the different occlusion ratios. These comparisons highlight the effectiveness of our proposed method under different occlusion ratios, showcasing its robustness and improved performance compared to existing methods.

4.2.3 When both gallery and probe images are occluded

Face recognition accuracy degrades too much when both gallery and probe images are occluded, especially when the occluded areas differ between the two sets of images. Most existing methods, such as HMM and NBNN, need to recognise faces under these conditions. Our model efficiency is tested with gallery and probe images that do not share the same occlusion part.

As shown in Figure 8, our TSRM method accuracy is much larger than the existing method. However, our model provides the best performance when both gallery and probe images are occluded for AR and ORL datasets. This is very important in the recent scenarios where the occluded face is captured in surveillance applications such as malls, airports and traffic light signals.

Figure 8 Accuracy comparison with the existing methods when both gallery and probe images are occluded for AR and ORL dataset, (a) for AR dataset (b) for ORL dataset (see online version for colours)



4.3 Results of facial expression

The AR dataset evaluated facial expression recognition, while the ORL dataset is used to assess illumination recognition. Our TSRM method is not able to identify the facial expression, as mentioned in Table 3 for the AR dataset. However, our TSRM provides patch distribution by taking the patch difference across the image, while facial expression recognition needs local patch differences in the specific portion of the image. Table 4 checks the model efficiency in illumination conditions for the ORL dataset. Similarly, in the case of illumination recognition, our model cannot meet the BOW method, as shown in Table 4. Our proposed method does not outperform other existing methods in these scenarios. However, our method can identify a person's facial expression with

approximately 83% accuracy and illumination with 96% accuracy, which is a notable achievement given the challenging conditions.

Table 3 Accuracy comparison of our proposed method with state-of-the-art methods on AR dataset for facial expression recognition

Methods	Recognition rate (%)
TSRM	83.5
HMM (Yu and Yang, 2001)	66.5
PCA-SVM (Mario, 2014)	89.7
SRC (Weng et al., 2013)	92.0
OP-NBNN (Ogasawara, 2021)	89.3
NBNN (Ma et al., 2022)	85.3
DICW (Bommidi and Sundaramurthy, 2021)	96.7

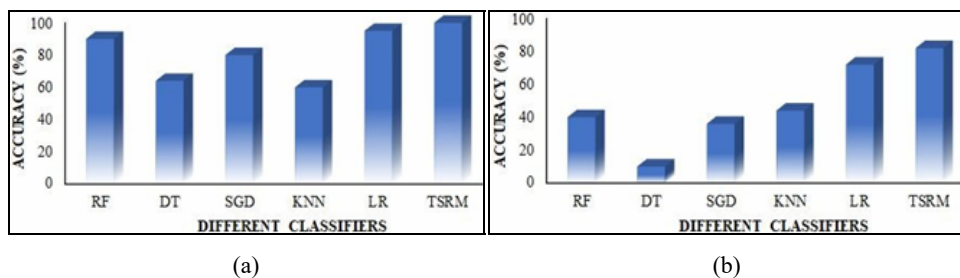
4.4 Comparison with other training and testing-based methods

Many existing training and testing-based techniques are used for image classification, such as support vector machine (SVM), RF, decision tree (DT), stochastic gradient descent (SGD), k-nearest neighbouring (KNN), and logistic regression (LR). Our TSRM method is based on normal distribution, where we do not need to train and test the images for classification. Our method performs better than the existing classification methods on AR and ORL datasets.

Table 4 Comparison of our proposed method with state-of-the-art methods on ORL dataset in terms of illumination

Methods	Recognition rate (%)
TSRM	96.66
Naseem et al. (2010)	95.8
Wen (2012)	94.10
Zhang et al. (2010)	Not evaluated
Lu et al. (2010)	90.5
BoW (Zhao et al., 2015)	95.0

Figure 9 Accuracy comparison of different classifier models for AR and ORL datasets, (a) for AR dataset (b) for ORL dataset (see online version for colours)



As shown in Figures 9(a) and 9(b), DT classifiers perform worse than all the classifiers on the AR and ORL datasets, where our TSRM provides the best performance compared to other training-based classifiers. Our proposed work offers excellent performance on AR and the ORL dataset.

5 Conclusions

Based on local feature-matching techniques, we propose a threshold string representation model for reliable face identification under occlusion. While traditional classification methods like SVM and RFs are employed, our proposed classification method outperforms other techniques in terms of performance. This approach, called TSRM, combines feature extraction and feature matching algorithms. The K-means clustering method generates features by taking the differences between patches and converting these features into clustering or string sequences. Components are matched by applying a threshold to a string sequence of gallery and probe images. Our proposed method offers improved identification accuracy compared to the CSM-RL technique. We evaluated our methodology using the AR and ORL databases to check the model's performance under natural and synthetic occlusion conditions. Moreover, our model outperforms other state-of-the-art approaches in robust face identification across various scenarios, including different illuminations, facial expressions, and face occlusions.

While the TSRM algorithm demonstrates promising performance in handling occlusions like sunglasses, scarves, and caps, there are some limitations to consider. One limitation is the method's reliance on image segmentation into patches, which requires careful tuning of patch size and configuration to achieve optimal results. This can make the algorithm sensitive to parameter selection, potentially impacting its performance on different datasets or in varying real-world scenarios. Additionally, the TSRM method has primarily been tested on controlled datasets like AR and ORL, which may not fully represent the diversity and complexity of real-world occlusions. Further evaluation on larger and more diverse datasets is necessary to confirm its generalisability.

References

- Bartlett, M.S., Movellan, J.R. and Sejnowski, T.J. (2002) 'Face recognition by independent component analysis', *IEEE Transactions on Neural Networks*, Vol. 13, No. 6, pp.1450–1464.
- Belhumeur, P.N., Hespanha, J.P. and Kriegman, D.J. (1997) 'Eigenfaces vs. fisherfaces: recognition using class specific linear projection', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp.711–720.
- Boiman, O., Shechtman, E. and Irani, M. (2008) 'In defense of nearest-neighbor based image classification', in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp.1–8.
- Bommidi, K. and Sundaramurthy, S. (2021) 'A compressed string matching algorithm for face recognition with partial occlusion', *Multimed Syst.*, April, Vol. 27, pp.191–203.
- Chang, C-C. and Lin, C-J. (2011) 'LibSVM: a library for support vector machines', *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol. 2, No. 3, pp.1–27.
- Chen, S., Wu, G., Yang, Y. and Guo, Z. (2023) 'A simple and effective patch-based method for frame-level face anti-spoofing', *Pattern Recognition Letters*, Vol. 171, pp.1–7.

- Chen, W. and Gao, Y. (2010) 'Recognizing partially occluded faces from a single sample per class using string-based matching', in *Proceedings of the European Conference on Computer Vision (ECCV)*, Vol. 3, pp.496–509.
- Chen, W. and Gao, Y. (2013) 'Face recognition using ensemble string matching', *IEEE Transactions on Image Processing*, Vol. 22, No. 12, pp.4798–4808.
- Fritz, K., Damiana, L. and Serena, M. (2019) 'A robust group sparse representation variational method with applications to face recognition', *IEEE Transactions on Image Processing*, Vol. 28, No. 6, pp.2785–2798.
- Gao, Y. and Leung, M. (2022) 'Human face profile recognition using attributed string', *Pattern Recognition*, Vol. 35, No. 2, pp.353–360.
- Gross, R. (2005) 'Face databases', *Handbook of Face Recognition*, pp.301–327, Springer, New York, USA.
- Gupta, S., Thakur, K. and Kumar, M. (2021) '2D-human face recognition using sift and surf descriptors of face's feature regions', *Visual Computer*, March, Vol. 37, pp.447–456.
- He, R., Zheng, W.S., Tan, T. and Sun, Z. (2014) 'Half-quadratic-based iterative minimization for robust sparse representation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 36, No. 2, pp.261–275.
- Huang, D., Ardabilian, M., Wang, Y. and Chen, L. (2012) '3-D face recognition using ELBP-based facial description and local feature hybrid matching', *IEEE Transactions on Information Forensics and Security*, Vol. 7, No. 5, pp.1551–1565.
- Jia, H. and Martinez, A.M. (2009) 'Support vector machines in face recognition with occlusions', in *Proceedings of the IEEE 10th Scientific World Journal Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR)*, pp.136–141.
- Jia, K., Chan, T.H. and Ma, Y. (2012) 'Robust and practical face recognition via structured sparsity', in *Proceedings of the European Conference on Computer Vision (ECCV)*, Vol. 7575, pp.331–344.
- Kanan, H.R., Faez, K. and Gao, Y. (2008) 'Face recognition using adaptively weighted patch PZM array from a single exemplar image per person', *Pattern Recognition*, Vol. 41, No. 12, pp.3799–3812.
- Kortli, Y., Jridi, M., Falou, A.A. and Atri, M. (2020) 'Face recognition systems: a survey', *Sensors*, January, Vol. 20, No. 2, p.342, Switzerland.
- Liao, S. and Jain, A.K. (2011) 'Partial face recognition: an alignment-free approach', in *Proceedings of the International Joint Conference on Biometrics*.
- Liu, F., Chen, D., Wang, F., Li, Z. and Xu, F. (2023) 'Deep learning based single sample face recognition: a survey', *Artificial Intelligence Review*, Vol. 56, No. 3, pp.2723–2748.
- Lu, G.F., Lin, Z. and Jin, Z. (2010) 'Face recognition using discriminant locality preserving projections based on maximum margin criterion', *Pattern Recognit.*, Vol. 43, No. 10, pp.3572–3579.
- Ma, X., Ma, Q., Ma, Q. and Han, X. (2022) 'Robust face recognition for occluded real-world images using constrained probabilistic sparse network', *IET Image Processing*, Vol. 16, No. 5, pp.1359–1375.
- Mario, F. (2014) 'Face recognition using approximate string matching', *Program Studi Teknik Informatika, Sekolah Teknik Elektro dan Informatika ITB, Corpus*.
- Naseem, I., Togneri, R. and Bennamoun, M. (2010) 'Linear regression for face recognition', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 11, pp.2106–2112.
- Nefan, A.V. and Hayes, M.H. (1998) 'Hidden Markov models for face recognition', in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 5, pp.2721–2724.
- Ogasawara, H. (2021) 'A non-recursive formula for various moments of the multivariate normal distribution with sectional truncation', *Journal of Multivariate Analysis*, Vol. 183, p.104729.

- Su, Y., Zhe, L. and Mengyao, W. (2018) 'Sparse representation-based face recognition against expression and illumination', *IET Image Processing*, Vol. 12, No. 5, pp.826–832.
- Tan, X., Chen, S., Zhou, Z.H. and Liu, J. (2009) 'Face recognition under occlusions and variant expressions with partial similarity', *IEEE Transactions on Information Forensics and Security*, Vol. 4, No. 2, pp.217–230.
- Tan, X., Chen, S., Zhou, Z.H. et al. (2005) 'Recognizing partially occluded, expression variant faces from a single training image per person with SOM and soft K-NN ensemble', *IEEE Transactions on Neural Networks*, Vol. 16, No. 4, pp.875–886.
- Turk, M.A. and Pentland, A.P. (1991) 'Face recognition using eigenfaces', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.586–591.
- Vu, T. and Monga, V. (2017) 'Fast low-rank shared dictionary learning for image classification', *IEEE Transactions on Image Processing*, Vol. 26, No. 11, pp.5160–5175.
- Wallace, M., Ioannou, S., Raouzaoui, A., Karpouzis, K. and Kollias, S. (2006) 'Dealing with feature uncertainty in facial expression recognition', *International Journal of Intelligent Systems Technologies and Applications*, Vol. 1, Nos. 3–4, pp.409–429.
- Wei, X., Li, C.T., Lei, Z. et al. (2014) 'Dynamic image-to-class warping for occluded face recognition', *IEEE Transactions on Information Forensics and Security*, Vol. 9, No. 12, pp.2035–2050.
- Wen, Y. (2012) 'An improved discriminative common vectors and support vector machine based face recognition approach', *Expert Systems with Applications*, Vol. 39, No. 4, pp.4628–4632.
- Weng, R., Lu, J. and Tan, Y.P. (2016) 'Robust point set matching for partial face recognition', *IEEE Transactions on Image Processing*, Vol. 25, No. 3, pp.1163–1176.
- Weng, R., Lu, J., Hu, J., Yang, G. and Tan, Y.P. (2013) 'Robust feature set matching for partial face recognition', in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pp.601–608.
- Wright, J., Yang, A.Y. and Ganesh, A. (2009) 'Robust face recognition via sparse representation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 2, pp.210–227.
- Yang, M., Zhang, L., Feng, X. et al. (2014) 'Sparse representation based fisher discrimination dictionary learning for image classification', *International Journal of Computer Vision*, Vol. 109, No. 3, pp.209–232.
- Yang, M., Zhang, L., Yang, J. and Zhang, D. (2013) 'Regularized robust coding for face recognition', *IEEE Transactions on Image Processing*, Vol. 22, No. 5, pp.1753–1766.
- Yu, H. and Yang, J. (2001) 'A direct LDA algorithm for high-dimensional data with application to face recognition', *Pattern Recognition*, Vol. 34, No. 10, pp.2067–2070.
- Zeng, D., Veldhuis, R. and Spreuwers, L. (2021) 'A survey of face recognition techniques under occlusion', *IET Biometrics*, November, Vol. 10, pp.581–606.
- Zhang, B., Gao, Y., Zhao, S. and Liu, J. (2010) 'Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor', *IEEE Transactions on Image Processing*, Vol. 19, No. 2, pp.533–544.
- Zhang, D., Yang, M. and Feng, X. (2011) 'Sparse representation or collaborative representation: which helps face recognition?', in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pp.471–478.
- Zhao, C., Li, X. and Cang, Y. (2015) 'Bisecting k-means clustering based face recognition using block-based bag of words model', *Optik (Stuttgart)*, Vol. 126, No. 19, pp.1761–1766.
- Zhou, Z., Wagner, A., Mobahi, H. et al. (2009) 'Face recognition with contiguous occlusion using Markov random fields', in *IEEE International Conference on Computer Vision (ICCV)*, pp.1050–1057.