




Article

A New Integrated Approach Based on the Iterative Super-Resolution Algorithm and Expectation Maximization for Face Hallucination

K. Lakshminarayanan ¹, R. Santhana Krishnan ², E. Golden Julie ³, Y. Harold Robinson ⁴ ,
Raghvendra Kumar ⁵, Le Hoang Son ⁶, Trinh Xuan Hung ⁷, Pijush Samui ^{8,9,*},
Phuong Thao Thi Ngo ¹⁰  and Dieu Tien Bui ¹¹ 

¹ Department of Electronics and Communication Engineering, Francis Xavier Engineering College, Tirunelveli, Tamil Nadu 627003, India; kyelyen@gmail.com

² Department of Electronics and Communication Engineering, SCAD College of Engineering and Technology, Tirunelveli, Tamil Nadu 627414, India; santhanakrishnan86@gmail.com

³ Department of Computer Science and Engineering, Anna University Regional Campus, Tirunelveli, Tamil Nadu 627007, India; goldenjuliephd@gmail.com

⁴ School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India; yhrobinphd@gmail.com

⁵ Computer Science and Engineering Department, GIET University, Gunupur 765022, Odisha, India; raghvendraagrawal7@gmail.com

⁶ VNU Information Technology Institute, Vietnam National University, Hanoi 100000, Vietnam; sonlh@vnu.edu.vn

⁷ Institute of Information Technology, Vietnam Academy of Science and Technology, Hanoi 100000, Vietnam; trxhung@gmail.com

⁸ Geographic Information Science Research Group, Ton Duc Thang University, Ho Chi Minh City 3467331, Vietnam

⁹ Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City 3467331, Vietnam

¹⁰ Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam; ngotphuongthao5@duytan.edu.vn

¹¹ GIS Group, Department of Business and IT, University of South-Eastern Norway, Gullbringvegen 36, N-3800 Bø i Telemark, Norway; Dieu.T.Bui@usn.no

* Correspondence: pijush.samui@tdtu.edu.vn

Received: 20 December 2019; Accepted: 16 January 2020; Published: 20 January 2020



Abstract: This paper proposed and verified a new integrated approach based on the iterative super-resolution algorithm and expectation-maximization for face hallucination, which is a process of converting a low-resolution face image to a high-resolution image. The current sparse representation for super resolving generic image patches is not suitable for global face images due to its lower accuracy and time-consumption. To solve this, in the new method, training global face sparse representation was used to reconstruct images with misalignment variations after the local geometric co-occurrence matrix. In the testing phase, we proposed a hybrid method, which is a combination of the sparse global representation and the local linear regression using the Expectation Maximization (EM) algorithm. Therefore, this work recovered the high-resolution image of a corresponding low-resolution image. Experimental validation suggested improvement of the overall accuracy of the proposed method with fast identification of high-resolution face images without misalignment.

Keywords: hallucination; global sparse representation; Expectation Maximization

1. Introduction

Super-resolution (SR) is the process of improving a low resolution (LR) image to a high resolution (HR) image without altering the originality of the image. This super-resolution is also called hallucination [1]. The process of converting a low-resolution image to a high-resolution image is necessary during image processing [2], and it has a wide range of applications. In satellite image processing, image restoration and image enhancement require super-resolution to remove distortions and for enhancement of satellite images [3]. Super-resolution has wide applications in medical image processing for improving the quality of medical images, such as MRI CT scan, which requires more contrast and fine enhancements [4]. Its multimedia applications are also increasing, wherein all multimedia, e.g., images, videos, and animation require high definition when the SR functionality is involved [5]. Face hallucination must have the following constraints:

1. Data constraints: The output image of face hallucination must be similar to the input after completion of smoothening and downsampling.
2. Global constraint: It must have the same features of the human face, e.g., mouth, nose, and eyes. The face features must be steady.
3. Local constraint: The final image must have the exact characteristics of a face image with the local feature.

Video surveillance cameras are used widely in many places, such as banks, stores, and parking lots, where intensive security is critical [6]. Details of facial features obtained from the surveillance video are essential in identifying personal identity [7]. However, in many cases, the images obtained from the surveillance cameras cannot be well identified due to the low resolution of facial images that cause the loss of facial features [8]. Thus, to obtain detailed facial features for personal recognition, it is necessary to infer a low-resolution (LR) facial image to a high-resolution (HR) using by face hallucination or face super-resolution [9].

Such techniques are applied in a variety of essential sectors, e.g., medical imaging, satellite imaging, surveillance system, image enlarging in web pages, and restoration of old historic photographs [10]. Due to limited information on image identification, reconstruction and expression analysis is a challenge to both humans and computers. Under some circumstances, it is impossible to obtain image sequences [11]. Several super-resolution reconstruction (SRR) researches have been proposed, relying on two approaches: reconstruction-based and learning-based approach [12]. The reconstruction-based approach employs multiple LR images of the same object as an input for reconstructing an HR image.

In contrast, the learning-based approach uses several training samples from the same domain with different objects to reconstruct the HR image [13]. An advantage of the learning-based approach is its ability to reconstruct the HR image from the single LR image. Learning-based super-resolution is applied to human facial images [14]. Several related facial hallucination methods have been proposed in recent years. Learning-based methods have acquired more considerable attention as they can achieve high magnification factors and produce positive super-resolved results compared to other methods [15].

The related facial hallucination methods may be used on position-patches to improve image quality [16]. Such methods perform a one-step facial hallucination based on the position-patch instead of neighbor-patch. A patch position is one of the learning-based approaches that utilize the facial image, as well as image features, to synthesize high-resolution facial images from low-resolution ones [17]. In comparison, neighbor patches are used widely in face hallucination. The reconstruction of a high-resolution facial image could be based on a set of high and low-resolution training image pairs [18]. The high-resolution image is generated using the same position image patches of each training image [19]. This method can be extended using bilateral patches [20]. The local pixel structure is learned from the nearest neighbors (KNN) faces. However, there are some uncontrollable problems regarding this method, i.e., the facial images captured by the camera (i.e. LR and facial hallucination)

are limited to frontal faces [21]. Therefore, it is practically significant to study how to create HR multi-viewed faces from LR non-frontal images.

As a sequence, a method called the multi-viewed facial hallucination method based on the position-patch was developed in [22]. It is a simple face transformation method that converts an LR face image to a global image, predicting the LR multiple views of that given LR image. Based on the synthesized LR faces, facial details are incorporated using the local position-patch. Meanwhile, the traditional locally linear embedding (LLE) [23] technique, when applied in such a hallucination method still faces a problem related to the determination of the optimal weights. Such weights are defined using a fixed number of neighbors for every point [24]. This is not practical for real-world data because the number of neighbors in each point is not equal to the other point. In that study, multi-view face hallucination using an adaptive locally linear embedding technique was proposed for efficiently reconstructing high-resolution face images from a single low-resolution image [25]. The optimal weights determination was applied to manipulate the non-frontal facial details. By feeding a single LR face in one of up, down, left, or right views to the proposed method, the HR images are generated in all views [26].

The critical value of the mapping coefficient in LR to HR is computed in the TRNR [27]. Accuracy is maintained in the training vector using subspace matching, and data with dissimilar scales are computed for representation. The weighted coding methodology is used to evaluate the noisy image as input. Generalized adversarial related networks are used to evaluate the quality of the image using the super-resolution strategy [28]. Dissimilar types of global priority related methodologies prevent the data from segregating into tiny decomposed elements from the input image [29]. The neighbor embedding [30] methodology was used to resolve the natural images, according to the geometry LR patch space construction. The LR training set may be applied to the HR space to demonstrate the image patch. The least-square regression methodology was used to implement location-based patches to regularize coefficient representation to get the most accuracy [31].

The drawbacks of the related methods include not being suitable for global face images while performing super resolving generic image patches. They have minimized accuracy and are time-consuming. While processing the minimized resolution image, the guaranteed output is not achieved. Therefore, the main objectives of that proposed method are:

- (a) Using a hybrid methodology, where low-resolution images can be converted into high-resolution based images.
- (b) Reducing misalignment variations using the local geometric co-occurrence matrix model.
- (c) Every image with random transformation for each pixel will minimize the computational complexity and minimize the loss to avoid degenerate solutions.
- (d) Maintaining a high amount of accuracy.

Therefore, this paper proposed a new integrated approach based on the iterative super-resolution algorithm and expectation-maximization for face hallucination. Training global face sparse representation was used to reconstruct images with misalignment variations after a local geometric co-occurrence matrix. In the testing phase, we propose a hybrid method, which is a combination of the global sparse representation and local linear regression using the Expectation Maximization (EM) algorithm. Therefore, this work recovers the high-resolution image of a corresponding low-resolution image. Simulation results proved that the proposed work performed well in various performance metrics compared to related works.

The paper is organized as follows: The proposed system with modules such as global face model learning, local geometric co-occurrence model learning, and the expectation-maximization algorithm is introduced in Section 2. The results and discussions are in Section 3, and finally, the conclusion and the future enhancements are given in Section 4.

2. Proposed Method

2.1. Main Ideas

Sparse representation is the existing method for super-resolution, which is not appropriate for global face images, as we need to explore non-linear information hidden in the data. Moreover, the available Eigenface model and Markov network model have very low accuracy. The processing step starts when the low-resolution image is interpolated to the size of the target high-resolution image [11]. The low-resolution interpolated image is a blurred image that is deficient in high-frequency information, and it is used as a preliminary estimation of the target HR image. Then, the image will be divided further into non-overlapping image patches that are not separate. The framework uses the generated patches to identify the image, which is most matched by searching a training data set of low-resolution and high-resolution image patches. The selected high-resolution image is generated to study high-resolution information. Finally, the trained high-resolution image and the interpolated images are combined to estimate the high-resolution target image. The embedded space is learned for producing solutions with representation, distinct solutions for the same functionality, and picking the group of embedding vectors. Therefore, mapping within the embedding spaces is restricted. Figure 1 demonstrates the block diagram of SR based face hallucination.

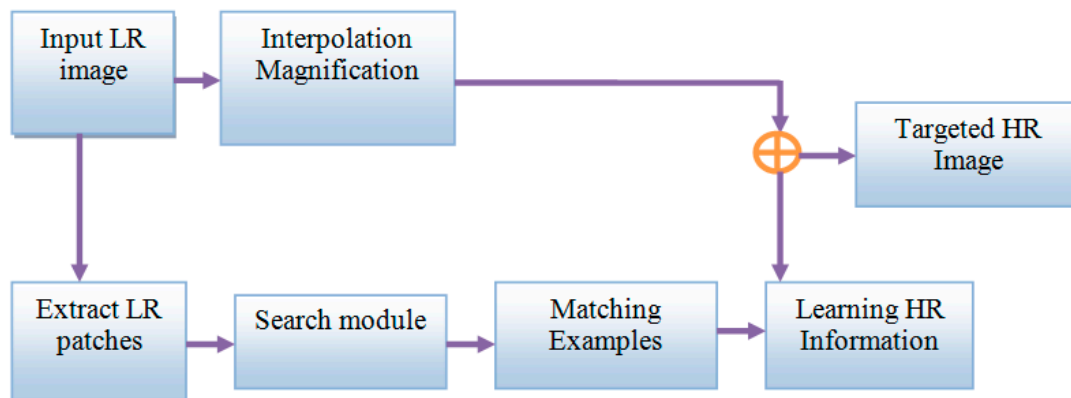


Figure 1. Block diagram of super-resolution (SR) based face hallucination.

The proposed algorithm was employed to discover the non-linear structure in the data. The algorithm was developed under the assumption that any object is “nearly” flat on small scales. The initial object was to map data from one space globally to another space by embedding the locations of the neighbors for each point. For face hallucination, it was about LR and HR face image spaces. The main idea of the proposed methodology was minimizing the reconstruction error of the set of all local neighborhoods in the data set. The reconstruction weights were computed by reducing the reconstruction error using the cost function in Equation (1).

$$\delta_w = \sum_{i=1}^n \left(x_i - \sum_{j=1}^k w_{i,j} x_{N(j)} \right)^2 \quad (1)$$

For face hallucination, such reconstruction weights are computed in the LR face image space and then applied to the HR space to reconstruct the HR face image that corresponds to the LR input. The proposed method normally contains the weight matrix W computed from the neighborhood in the same space as the training samples. Similar to LLE, the ALLE (Adaptive locally linear embedding) computes the weights from the space of training samples, which is built adaptively from only the neighborhood of each input, not all training samples. Since some information contained in all training

samples can make the optimizer misleading to another optimal value, the threshold of similarity θ is defined for building each subspace in Equation (2).

$$\frac{\sum_{i=1}^k d_i}{\sum_{i=1}^n d_i} \leq \theta \quad (2)$$

Face image is represented as a column vector of all pixel values. Based on structural similarity, the face image can be synthesized using the linear combination of training objects. In other words, a face image at an unfixed view can also be reconstructed using a linear combination of other objects in the same view aligned as computed using Equation (3).

$$I_p = W_p L_p \quad (3)$$

There is no information to determine the construction coefficients at other views. After obtaining the LR in all views, these LR images are hallucinated to the HR face images of all views using the position-patch methodology. The proposed framework for multi-view face hallucination is to replace the linear combinations in all processes to improve the performance of hallucinating face images in multiple views. The framework means the number of the neighbors can be adapted for each input patches, and this would relieve the deviation from optimal values. The symbols used to construct the proposed framework are demonstrated in Table 1.

Table 1. Symbol description.

Symbol	Description
δ_w	cost function
$w_{i,j}$	reconstruction weights
x_i	linearly reconstructed
$x_{N(j)}$	linearly reconstructed for the neighbors
W	weight matrix
θ	threshold of similarity
d_i	distance within the input sample and first training sample
I_p	the given LR face at view p
W_p	the construction coefficients at view p
L_p	the training faces at view p

2.2. AIHEM Algorithm

Step 1: Consider a set of starting parameters in incomplete data.

Step 2: (E-step): Using the observed data from the data set approximates the values of the lost data.

Step 3: (M-Step): the complete data generated after the E-step is used for updating the parameters.

Step 4: Iterate Steps 2 and 3 until convergence.

The proposed model consists of 3 phases:

- Global face model learning.
- Local geometric co-occurrence model learning
- Iterative sparse representation optimization.

The training phase consists of two parts: The first part is the global face model learning and local geometric co-occurrence model learning. We use a Gaussian mixture model-based estimator for expectation-maximization, as the Gaussian mixture model uses a combination of probability distribution and estimates the mean and standard deviation parameters, as demonstrated in Figure 2.

Step 1: Consider a set of starting parameters in the incomplete data, with a hypothesis that the observed data is generated from a precise model.

Step 2: E-step—Using the observed data from the data set, approximate the values of the lost data, which are used to update the variables.

Step 3: M-Step—the complete data generated after the E-step is used to update the parameters, i.e., for updating the hypothesis.

Step 4: Checks whether the values are converged, where the concept of convergence is an intuition based on probabilities. When the probability of the variables has a tiny difference, then we say it has converged, i.e., the values are matched. If it is matched, the process stops, otherwise Steps 3 and 4 are iterated until convergence.

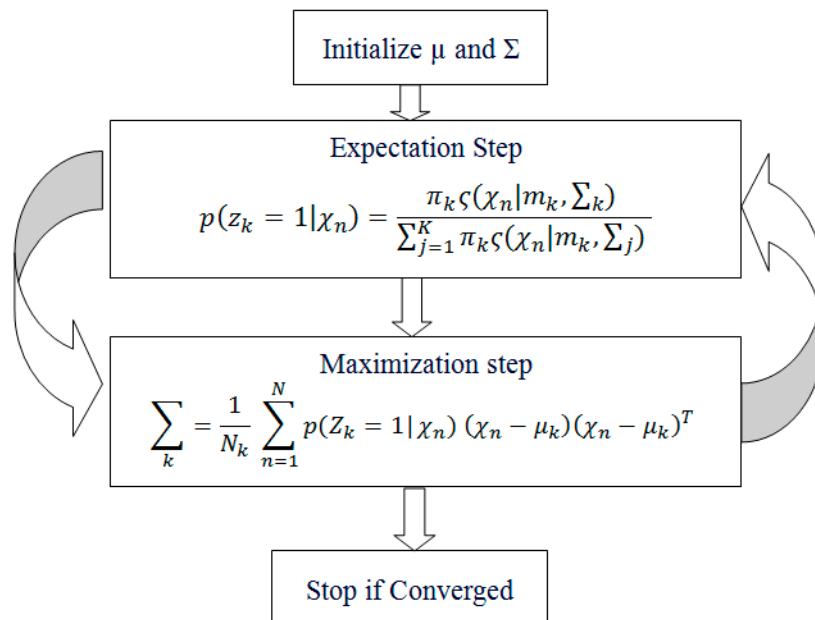


Figure 2. Gaussian mixture formation.

2.3. Global Face Model Learning

First, we provide the input HR face images with misalignment and apply the online dictionary to create the dictionary H for the HR face images. Figure 3 demonstrates the global face model learning. To construct a dictionary to obtain face image “sparsity prior,” face images are preferentially represented by the basis vectors of the same misalignment variation so that the super-resolved HR face images are constrained in a subspace of certain misalignment variation.

HR face images are always highly dimensional (the dimension is 127×158 for an HR image and 50×62 for an LR image in this paper), and the training set needs to be large enough for training the redundant dictionary. However, in practical applications, the computational complexity of the sparse problem is very high with an extensive dictionary. Generally speaking, the computational cost is proportional to the size of the dictionary H, i.e., a size of $\frac{1}{4} nm$, where n is the dimension of the HR face image, and m is the number of basis vectors in the dictionary. Since the solutions are usually an iterative method dealing with a large dictionary matrix, the storage requirements and computational cost are significant. In our case, if all images with different misalignment variations were selected to compose a dictionary, it would be difficult to solve the super-resolution problem. Thus, we considered dictionary learning to compress a dictionary to reduce the operational burden. Batch-based methods have problems dealing with an extensive training matrix effectively. This paper employed an online dictionary learning method to process the images, one at a time, with low memory consumption and lower computational costs. The number of basis vectors in the dictionary was 1024. The following steps were required to perform online dictionary learning.

1. Assume the training set, such as the first dictionary and number of iterations.
2. Reset the past information.

- Find the sparse coding using LARS (least angle regression coefficient).

$$\delta_t = \min_{\delta \in \mathbb{R}^k} \frac{1}{2} \|\chi_t - D_{t-1} \delta\|_2^2 + \vartheta \|\delta\|_1 \tag{4}$$

Here δ is dictionary coefficient

ϑ is Lagrange's multiplier.

χ is an Original data.

D_i is the created dictionary.

Suffix 2 is a normalization factor.

- Compute the learned dictionary,

$$D_{i_t} = \min_{D_i \in C} \frac{1}{t} \sum_{i=1}^t \left(\frac{1}{2} \|\chi_i - D_i \delta_i\|_2^2 + \vartheta \|\delta_i\|_1 \right) \tag{5}$$

Here i is loop iteration.

- Get the updated dictionary.

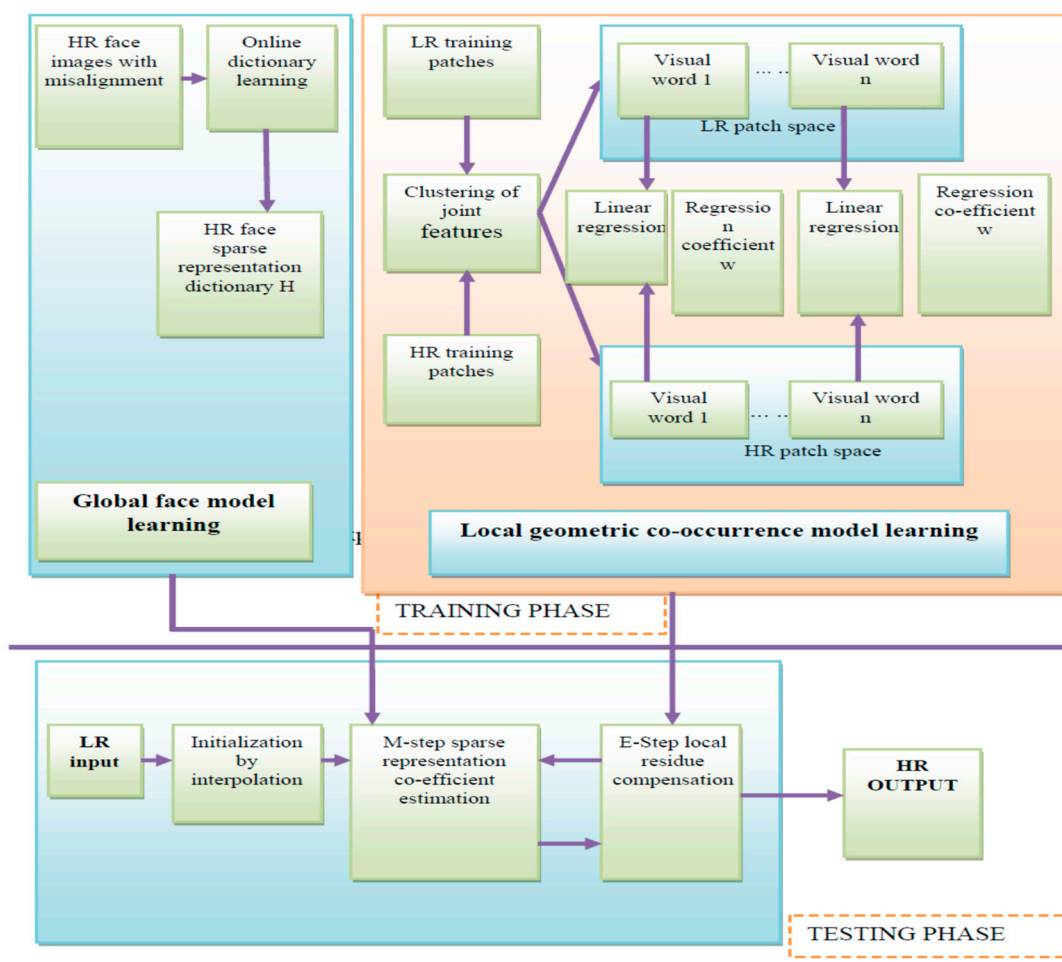


Figure 3. Global face model learning.

2.4. Local Geometric Co-Occurrence Model Learning

The following steps are required to perform the local geometric co-occurrence model learning:

1. Consider the training set LR and HR face residue.
2. Get the local geometric feature representation for each patch.
3. Connect the HR and the corresponding LR geometric feature vectors for jointly learning the HR and LR visual vocabularies.
4. Apply the affinity propagation clustering algorithm.
5. All data points are considered simultaneously as candidate cluster centers.
6. Messages are transmitted between the data points until an appropriate set of cluster centers and corresponding clusters emerge.
7. The number of clusters also depends on a prior specification of how preferable each data point is as a cluster center.
8. Consider all data points that have equal potential to be the cluster centers, so the preferences of all data points are set to the same value that can be controlled to produce different numbers of clusters.

In the testing phase, there were three steps to be followed. In the first step, we chose the LR image. Then, the initialization and the optimization process were performed. The LR image was given as input and it was initialized by interpolation. That image was compared to the global face image and processed to the M-step. Hence, the initialized image was compared to the local geometric co-occurrence model image, where that image was a process in the E-Step. Finally, the M-step and the E-step image was optimized. The geometrical feature that was defined in this section represented the high-frequency structures that are due to various intensity changes.

2.5. Demonstration of Execution Steps of the Proposed Method

Figure 4 demonstrates the execution steps for the proposed method.

Step 1: The modules are selected from the graphic user interface. In the training phase, there is global face model learning and local geometric co-occurrence model learning. Here, two modules can be selected. The first module is the training phase, and the second module is the testing phase. Figure 5 illustrates the GUI (Graphical User Interface) of Module Selection.

Step 2: Selection of training phase, the second page opens, which is the GUI of the Training Phase, which contains global face model learning and local geometric co-occurrence model learning. Figure 6 demonstrates the GUI of the training phase.

Step 3: Selection of the global face model learning. Two directories can be chosen as the high-resolution directory, and they create the dictionary. Figure 7 illustrates the GUI of the global face model learning phase.

Step 4: Choosing a directory for training the GUI for every image. Three poses are created so that it will be easy to compare the images with the testing phase. Figure 8 illustrates the GUI for directory selection.

Step 5: All images with different misalignment variations are selected to compose a dictionary. This output shows that the dictionaries are created for each directory using online dictionary learning. Figure 9 demonstrates the GUI of the HR face sparse representation dictionary.

Step 6: Steps in local geometric co-occurrence model learning. The first step chooses the LR images and extracts features from the LR patches, and the second step chooses the HR images and extracts the features of the HR patches. Figure 10 illustrates the GUI of global face model learning.

Step 7: The given global face input image is divided into LR (Low resolution) and HR (High Resolution) patches that include individual parts such as the eye and nose region. Figure 11 demonstrates the selection of the LR image. Figure 12 illustrates feature extraction from the LR image. Figure 13 demonstrates the selection of the HR image. Figure 14 illustrates feature extraction from the HR Image.

Step 8: On extraction of the HR and LR features from the input image, visual words are created for both the LR and HR, separately. Figure 15 demonstrates the creation of visual words.

Step 9: Applying linear regression to the HR and LR visual words, the regression coefficients are determined. The regression coefficients are calculated by finding the mean value of the RGB (Red-Green-Blue) of the HR and LR image gradient features. Figure 16 demonstrates the application of linear regression on visual words.

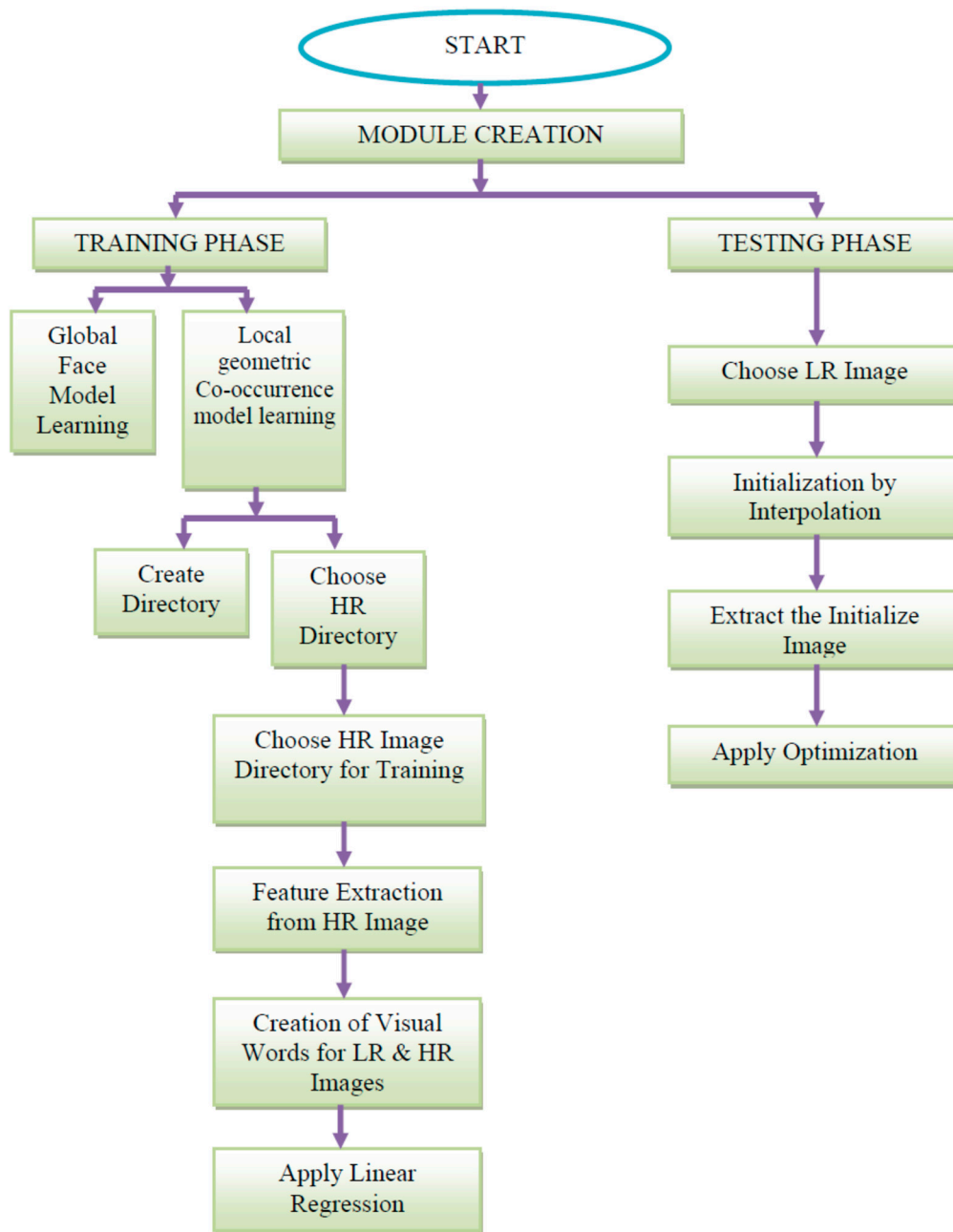


Figure 4. Execution steps of the proposed method.

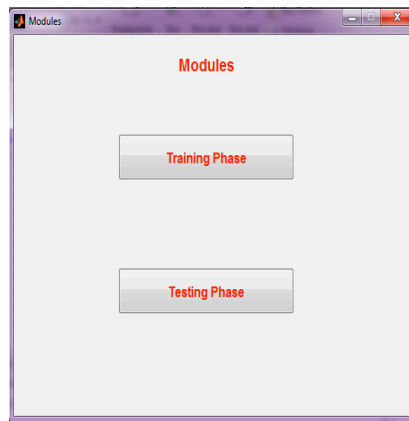


Figure 5. GUI of Module Selection.

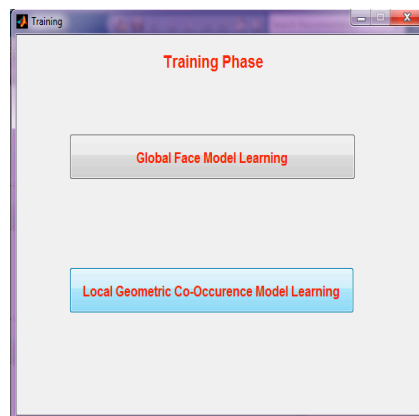


Figure 6. GUI of the training phase.

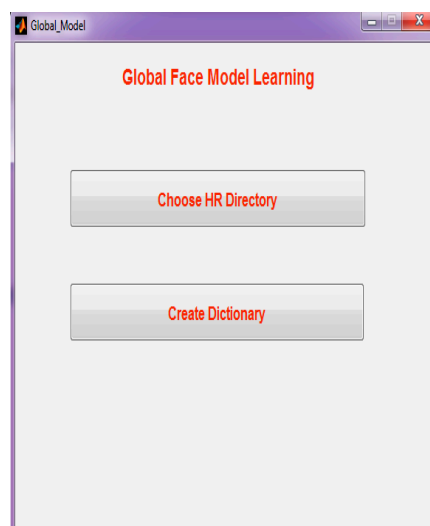


Figure 7. GUI of global model learning.

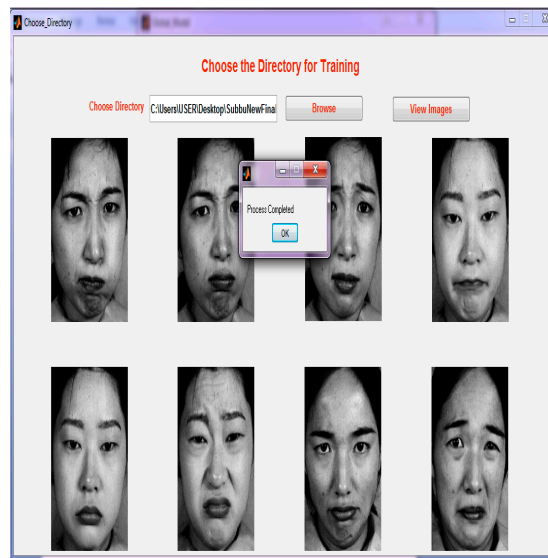


Figure 8. GUI of directory selection.

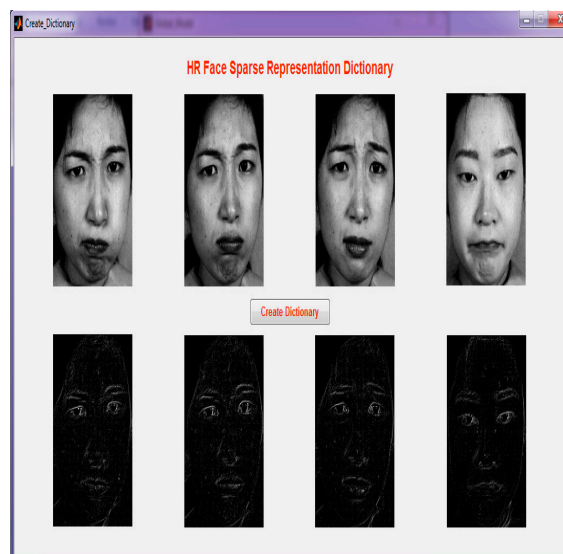


Figure 9. GUI of the high-resolution (HR) face sparse representation dictionary.

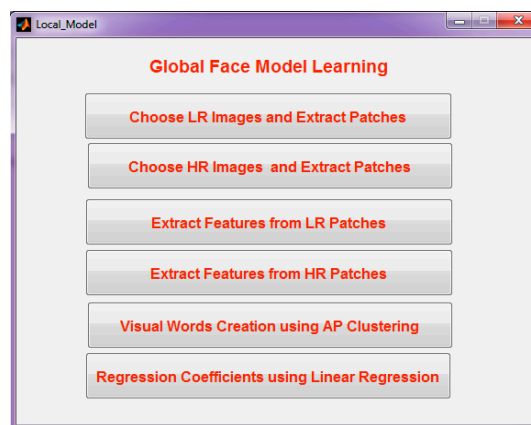


Figure 10. GUI of global face model learning.

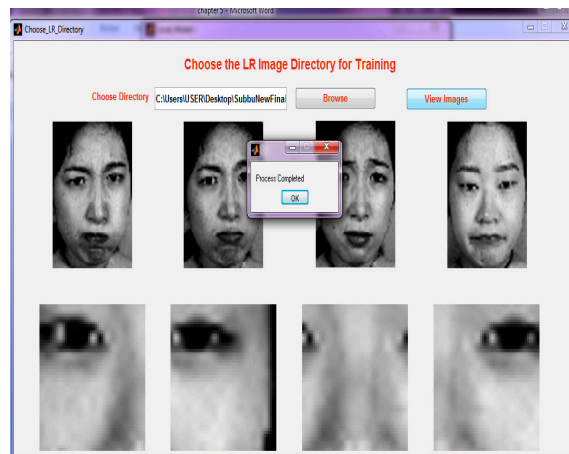


Figure 11. Selection of the low-resolution (LR) image.

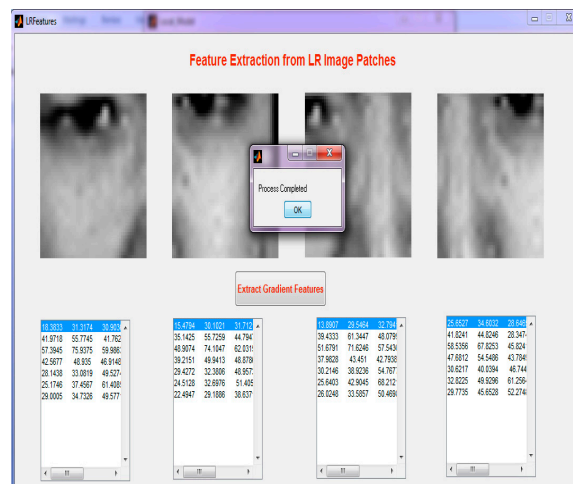


Figure 12. Feature extraction from the (LR) image.

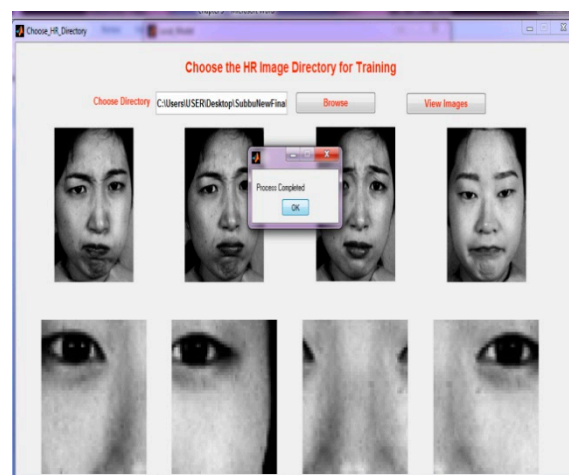


Figure 13. Selection of the HR image.

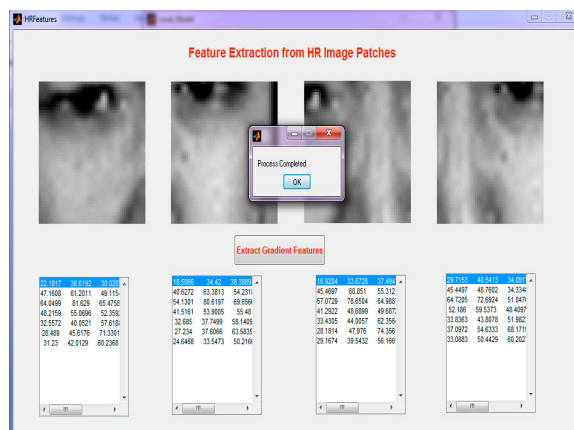


Figure 14. Feature extraction from the (HR) image.

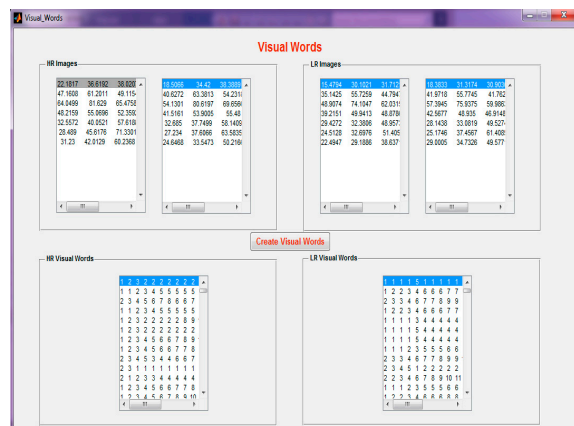


Figure 15. Creation of visual words.

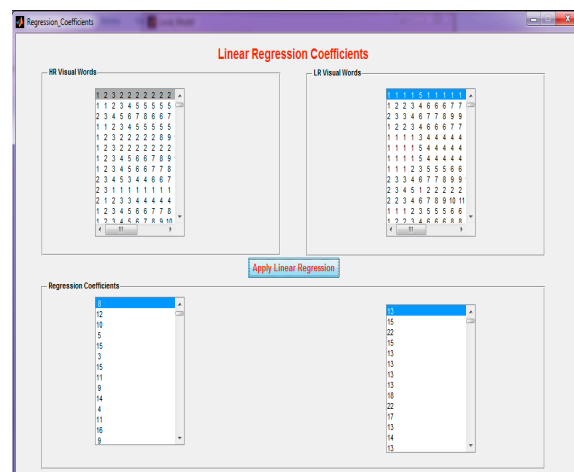


Figure 16. Linear regression on visual words.

Testing Phase:

- Step 1:** Choosing an LR Image. Figure 17 demonstrates choosing the LR image from the directory.
- Step 2:** The LR image should be applied for initialization by interpolation for resizing the LR image, which is easy to compute using the global face model. Figure 18 demonstrates initialization for interpolation. Figure 19 illustrates applying interpolation. After applying the initialization, the same

LR image is split into patches. Figure 20 demonstrates initial image extraction. The LR image should extract the features and give the gradient of the features. Then, the feature is compared to the local geometric co-occurrence model of the training phase.

Step 3: Optimization.



Figure 17. Choosing an LR image from the directory.

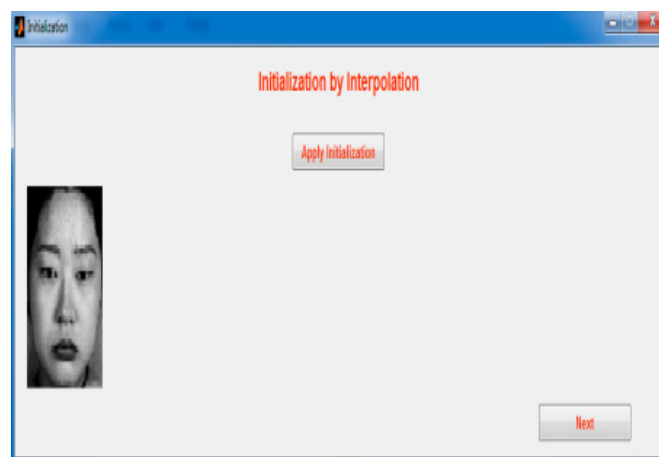


Figure 18. Initialization for interpolation.

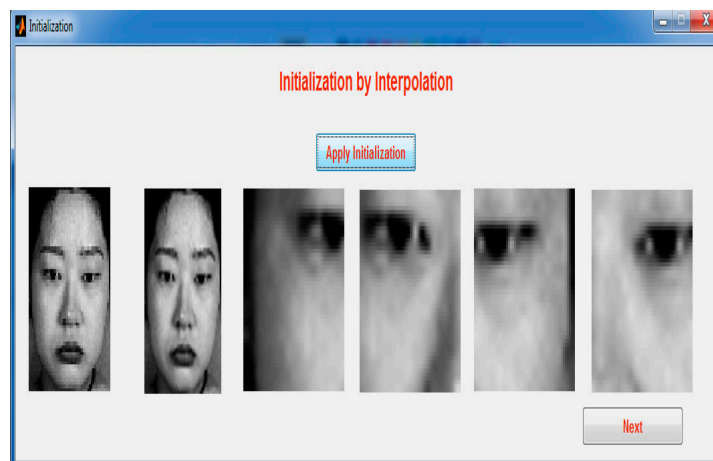


Figure 19. Applying interpolation.

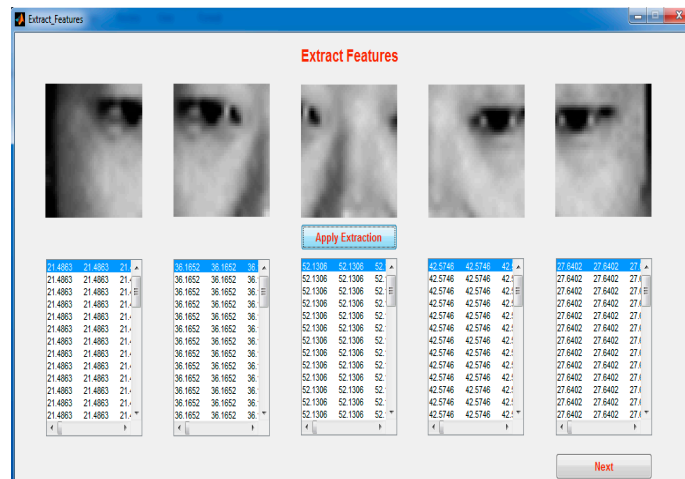


Figure 20. Initial image extraction.

The input LR image is initialized, and then the E-step is applied. In this step, the LR image is given as the input and it is initialized by interpolation. The image is compared to the global face image, and then it is processed to the M-step. Hence, the initialized image is compared to the local geometric co-occurrence model image, and that image is processed to the E-Step. Finally, the M-step and the E-step image is optimized. Figure 21 demonstrates the optimized image of the input LR image. The technique works perfectly with an RGB image like the black and white image shown in Figure 22.



Figure 21. Optimized image of the input LR image.

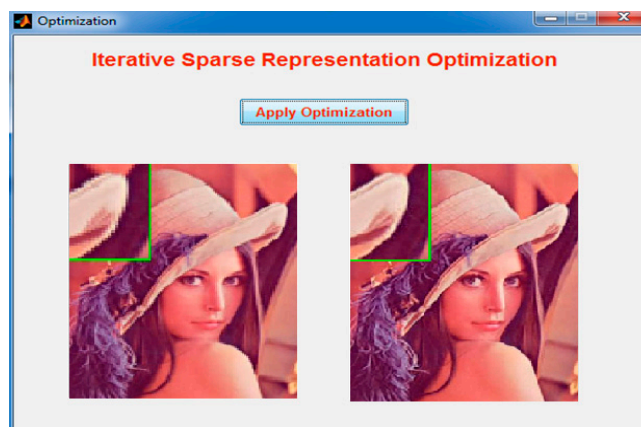


Figure 22. Iterative sparse representation.

This method consists of training testing phases. The training phase consists of two parts. The first part is global face sparse representation learning in which an online dictionary is used to learn the face prototype dictionary. The second part is a local geometric co-occurrence before learning, where we jointly clustered the HR and LR image residue patches and learned the regression coefficients from the LR to HR visual vocabulary.

In the testing phase, when the LR face image is presented, an initial image was generated by simple interpolation. An iterative sparse representation model using the learned dictionary was proposed to generate an HR image. The representation coefficients were iteratively optimized using the EM algorithm, in which local residue compensated to the global face using linear interpolation.

The training phase consisted of two parts. The first part was the global face model learning, and the second was the local geometric co-occurrence model learning. In global face model learning, two directories choose the high-resolution directory and create the dictionary. The images are selected from the directory for training. In the database for every image, three poses were created so that it would be easy to compare the images with the testing phase. All images with different misalignment variations were selected to compose a dictionary. This output showed that dictionaries were created for each directory using online dictionary learning. In the local geometric co-occurrence model learning, the first step was to choose the LR images and extract the features from the LR patches. The second step involved choosing the HR images and extracting the features of the HR patches.

The given global face input image was split into LR (low-resolution) patches that included unique parts such as the eye and nose region. It showed the features of the LR image and that the images should extract the gradient features. The given global face input image was split into HR (high-resolution) patches that included unique parts such as the eye and nose region. It showed the features of the HR image and that images should be extracted from the gradient features. The visual word was the collection of all local geometric features. The visual words were created for both HR and LR images separately. HR and corresponding LR geometric features were joined together for learning the HR and LR visual words. Regression coefficients were determined by applying linear regression to the HR and LR visual words. The regression coefficients were calculated by finding the mean value of the HR and LR image gradient features. In the testing phase, the global and local geometric features are involved. During the testing phase, three steps should be followed. The first step is choosing the LR image. Then the initialization and the optimization processes are performed.

The LR image should be applied for the initialization by interpolation, for resizing the LR image. It is used to compute the global face model. After applying the initialization, that same LR image is split into patches. The LR image should be extracted from the features, and it gives the gradient of the features. Then the feature is compared to the local geometric co-occurrence model of the training phase. The input LR image is initialized, and then the E-step is applied. In this step, the LR image is given as input, and that is initialized by interpolation. Then the image is compared to the global face image, after that the image is processed to the M-step. Hence, the initialized image is compared to the local geometric co-occurrence model image, where that image is processed to the E-Step. Finally, the M-step and the E-step image are optimized. The Kalman filter is used during the image optimization process. It is very effective in performing computational operations with linear filter modeling. It can produce an estimation of the current states of the system to remove the noise in the image. The main goal of utilizing the Kalman filter is to predict the position of a particular area of an image to be evaluated during the image optimization process. The expected position measures the prediction for the identifying area and it is identified using the variance and the confidence level. Building robust dictionaries is required to recover the high-resolution images and remove the dark regions of the images.

3. Performance Evaluation

The implementation requires an i3 processor with a 4GB RAM on MATLAB R2018b using the Windows 7 operating system. The proposed method was compared to related methods, e.g.,

SRGAN [28], TRNR [27], and LSR [26]. Here, the time taken analysis is denoted by the training phase image value. Each individual has five different views (left, right, up, down, and frontal views) under the same light conditions. These face images were aligned manually using the locations of three points: centers of the left and right eyeballs and the center of the mouth. Some aligned face images were cropped to 32×24 pixels for low-resolution face images, and to 128×96 pixels for high resolution face images. Based on the same training sets, the proposed method was compared to related methods. Each input face image was generated into five different outputs of LR, synthesized LR face images, and HR face images according to the framework. Table 2 demonstrates the list of PSNR values of the hallucinated image with an input face image on the frontal view. The value of K is 100 and $\theta = 0.1$.

Table 2. List of PSNR values of the hallucinated image—frontal view, $K = 100$, and $\theta = 0.1$.

Views	SRGAN	TRNR	LSR	AIHEM
Frontal	31.94	32.49	32.67	32.89
Up	25.32	26.98	26.96	29.94
Down	25.11	26.59	26.71	26.80
Left	30.45	31.13	31.65	32.47
Right	26.67	28.40	29.11	30.19

Table 3 demonstrates the list of PSNR values of the hallucinated image with an input face image on the up view. The value of K was 200 and $\theta = 0.2$. Table 4 demonstrates the list of PSNR values of the hallucinated image with an input face image on the down view. The value of K was 200 and $\theta = 0.2$. Table 5 demonstrates the list of PSNR values of the hallucinated image with an input face image on the left view. The value of K was 200 and $\theta = 0.2$. Table 6 demonstrates the list of PSNR values of the hallucinated image with an input face image on the right view. The value of K was 200 and $\theta = 0.2$. The threshold of similarity was used to find the similarity within a collection of objects. It contained several factors to determine scalability and minimize computational cost. In the testing phase, it generated the low-resolution input images to high-resolution images. The semantic class of the image was essential for solving within the inherent class to achieve the improved result. In-depth features can be implemented to increase the accuracy.

Table 3. List of PSNR values of the hallucinated image—up view, $K = 200$, and $\theta = 0.2$.

Views	SRGAN	TRNR	LSR	AIHEM
Frontal	29.55	30.04	30.84	31.30
Up	31.62	33.64	33.98	34.18
Down	27.44	28.74	28.78	28.88
Left	31.53	32.92	33.56	34.68
Right	30.82	31.36	31.42	31.75

Table 4. List of PSNR values of the hallucinated image—down view, $K = 200$, and $\theta = 0.2$.

Views	SRGAN	TRNR	LSR	AIHEM
Frontal	28.78	29.61	31.52	32.93
Up	27.98	28.34	29.14	30.75
Down	32.15	33.20	33.89	34.17
Left	33.91	34.61	34.65	34.68
Right	31.87	32.70	33.25	34.47

Table 5. List of PSNR values of the hallucinated image—left view, $K = 200$, and $\theta = 0.2$.

Views	SRGAN	TRNR	LSR	AIHEM
Frontal	30.93	31.67	31.98	32.39
Up	29.35	30.55	31.12	31.84
Down	27.24	28.19	29.46	30.82
Left	32.78	33.44	33.78	34.29
Right	31.56	32.28	32.29	32.30

Table 6. List of PSNR values of the hallucinated image—right view, $K = 200$, and $\theta = 0.2$.

Views	SRGAN	TRNR	LSR	AIHEM
Frontal	26.54	26.98	27.43	28.76
Up	30.25	30.54	30.55	30.56
Down	26.12	26.23	26.23	26.24
Left	28.11	28.56	30.16	31.46
Right	32.45	33.12	33.34	33.75

In Figure 23, the accuracy is high because all images are performed by the training phase so that there is no error rate. The precision rate and recall rates are also equally high. Running time for evaluating the efficiency of the proposed methodology was compared to the related methodologies. The proposed method ran faster than the related methods because of its fast optimization techniques. Figure 24 demonstrates the running time. The experimental results showed the higher quality of reconstructed images of the proposed framework over the enhanced methods with interpolation. High-resolution face images of five different views were generated from a single low-resolution face image. According to the experimental results, the reconstructed image was more accurate if the view of the input image was the same as that of the output. Particularly, frontal, up, and down views achieved better estimations than others. The results of the proposed method showed superior reconstruction quality of the HR face image over other related methods in both visualization and PSNR values.

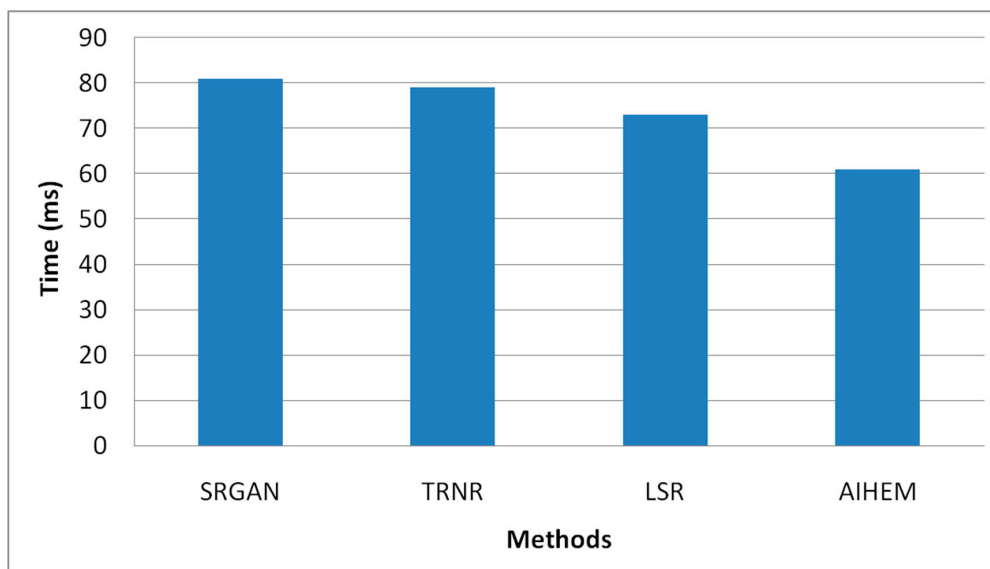


Figure 23. Comparison of Existing and Proposed Method.

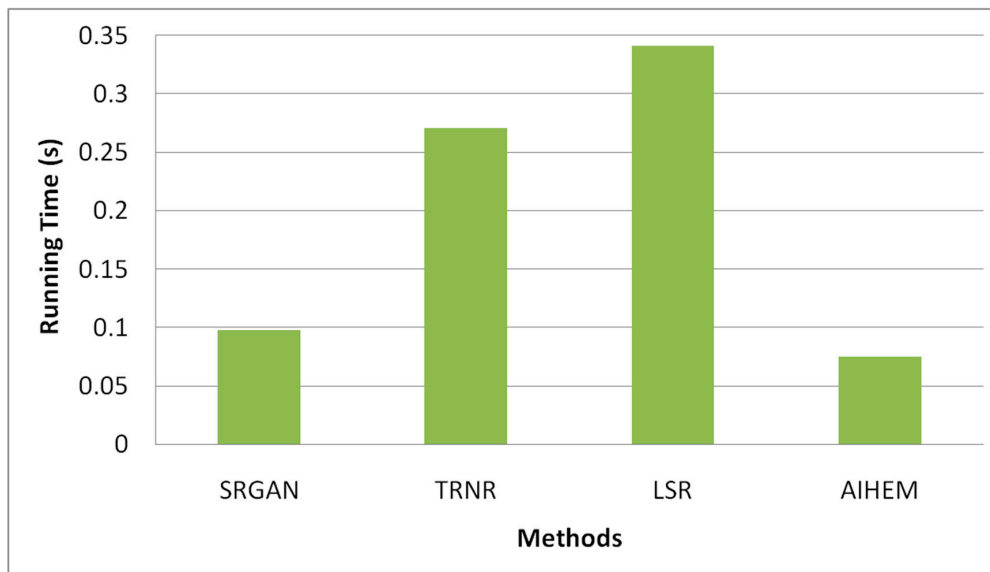


Figure 24. Running time.

The proposed AIHEM methodology was based on the EM algorithm, which is non-deterministic; thus, the performance evaluation was performed multiple times using standard deviation values. The standard deviation value was measured to obtain the PSNR value for the proposed method, as compared to the related methods. The performance results may vary in several iterations of continued evaluation, and the results are shown in Figures 25 and 26. The results showed a standard deviation of 10 and a standard deviation of 100, respectively.

The proposed method was a very effective technique and required fewer computational resources, so the processing was easy to produce the solutions for the optimization problems. It is useful to recognize and tune the data with predictable output. The performance improved compared to the related methods, and it was straightforward to train the dataset. The measurement of accuracy is an integral part of image classification when validating the proposed work. The vectors were used to extract the features to achieve the highest amount of accuracy. The iteration model was used to classify the image based on the similarity parameter from a pixel by pixel.

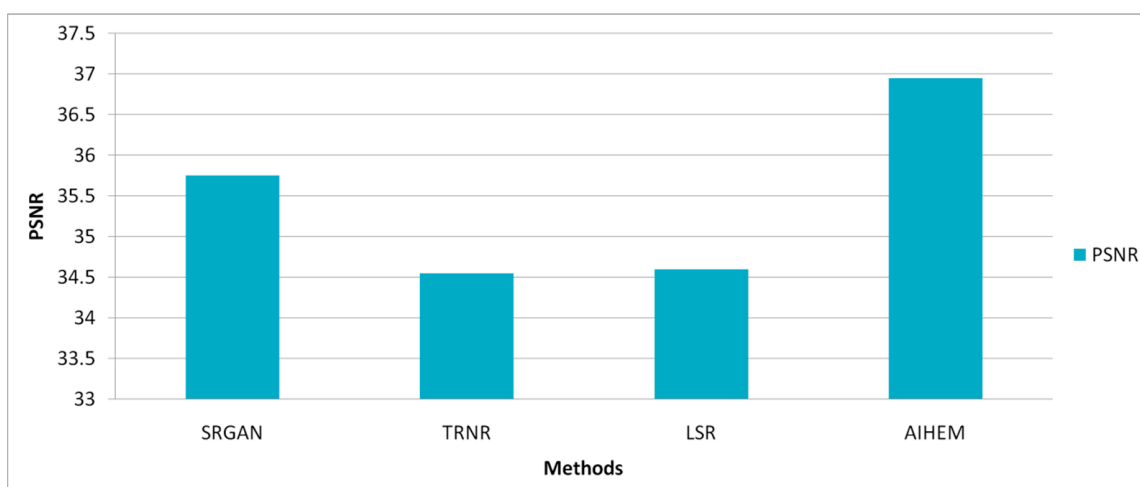


Figure 25. PSNR values for a standard deviation of 10.

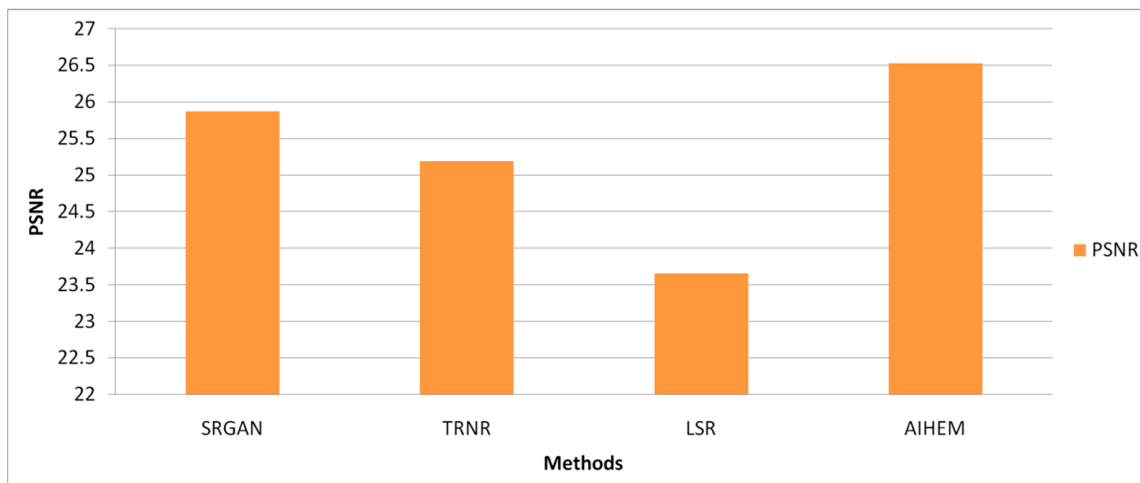


Figure 26. PSNR values for a standard deviation of 100.

The computational complexity of the proposed system was analyzed using the big oh notation $O(n^2KM)$, for the face hallucination methodology. The pre-processing and alignment methods were crucial to reducing computational complexity. The performance was analyzed using a threshold-based similarity that was 35 times faster than the other methods. The running time for the image in the testing phase was 13.5 s while using the proposed mechanism. The proposed method had a considerable development that attained a steady performance at a size of 20×20 pixel frame. Whenever the size increases, the running time will also increase. To minimize the computational complexity of the proposed algorithm, we maintained the size of the pixel frame as 20×20 pixels for performance evaluation.

4. Conclusions

This paper proposed a new iterative face hallucination using the expectation-maximization based iterative super-resolution algorithm. The M-step is a global face sparse representation model for adaptively selecting proper basis vectors of misalignment variation to input an LR image. The E-step uses an alignment-robust local geometric co-occurrence before compensating for the error generated by the sparse global representation in the first step. The global and local methods are combined iteratively. This iterative method not only takes advantage of both global and local approaches, but it also combines two different strategies to handle the same misalignment problem. Experimental results showed that the proposed method outperformed face super-resolution methods in terms of visual quality. Using the existing system (i.e., sparse representation), the accuracy of the HR image was 80%, but using the proposed method (global face sparse representation), the accuracy of the HR image was 85–90%. In the future, we will study various methods [32–43] and enhance the algorithm on posed face images.

Author Contributions: K.L. is responsible for data collection and preprocessing; R.S.K. and E.G.J. are responsible for first draft of methodology and testing; Y.H.R. and R.K. are responsible for first draft writing; L.H.S. and D.T.B. are responsible for methodology and second draft revision; T.X.H. is responsible for image processing algorithms and processing; P.S. and P.T.T.N. are responsible for validation by experiments and discussion section. All authors have read and agree to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: This research was funded by Vietnam Academy of Science and Technology (VAST), Vietnam under grant number VAST01.09/17-18: “Research and development of techniques to support Museum exhibits based on Virtual reality technology”.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liu, C.; Shum, H.-Y.; Freeman, W.T. Face Hallucination: Theory and Practice. *Int. J. Comput. Vis.* **2007**, *75*, 115–134. [[CrossRef](#)]
2. Liu, L.; Chen, L.; Chen, C.L.P.; Tang, Y.Y.; Pun, C.M. Weighted joint sparse representation for removing mixed noise in image. *IEEE Trans. Cybern.* **2017**, *47*, 600–611. [[CrossRef](#)] [[PubMed](#)]
3. Bhuvaneshwari, N.R.; Sivakumar, V.G. A comprehensive review on sparse representation for image classification in remote sensing. In Proceedings of the International Conference on Communication and Electronics Systems (ICCES), Funchal, Portugal, 30 March 2017.
4. Lakshminarayanan, K.; Ramesh, G.P. Image Compression using Frequency Band Suppression in VLSI Design based Discrete Wavelet Transform. *Int. J. Control. Theory Appl.* **2017**, *10*, 133–139.
5. Liu, L.; Chen, C.L.P.; You, X.; Tang, Y.Y.; Zhang, Y.; Li, S. Mixed noise removal via robust constrained sparse representation. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *18*, 2177–2189. [[CrossRef](#)]
6. Liu, L.; Chen, C.L.P.; Li, S.; Tang, Y.Y.; Chen, L. Robust face hallucination via locality-constrained bi-layer representation. *IEEE Trans. Cybern.* **2018**, *48*, 1189–1201. [[CrossRef](#)] [[PubMed](#)]
7. Jiang, J.; Ma, J.; Chen, C.; Jiang, X.; Wang, Z. Noise robust face image super-resolution through smooth sparse representation. *IEEE Trans. Cybern.* **2017**, *47*, 3991–4002. [[CrossRef](#)]
8. Wang, N.; Tao, D.; Gao, X.; Li, X.; Li, J. A Comprehensive Survey to Face Hallucination. *Int. J. Comput. Vis.* **2013**, *106*, 9–30. [[CrossRef](#)]
9. Nie, H.; Lu, Y.; Ikram, J. Face hallucination via convolution neural network. In Proceedings of the 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI), San Jose, CA, USA, 6–8 November 2016; pp. 485–489.
10. Li, Y. Sparse Representation for Machine Learning. In Proceedings of the Canadian Conference on Artificial Intelligence, Canadian AI 2013: Advances in Artificial Intelligence, Regina, SK, Canada, 28–31 May 2013; pp. 352–357.
11. Chen, L.; Pan, J.; Hu, R.; Han, Z.; Liang, C.; Wu, Y. Modeling and Optimizing of the Multi-Layer Nearest Neighbor Network for Face Image Super-Resolution. *IEEE Trans. Circuits Syst. Video Technol.* **2019**. [[CrossRef](#)]
12. Shi, J.; Liu, X.; Zong, Y.; Qi, C.; Zhao, G. Hallucinating Face Image by Regularization Models in High-Resolution Feature Space. *IEEE Trans. Image Process.* **2018**, *27*, 2980–2995. [[CrossRef](#)]
13. Zhang, G.; Liang, G.; Li, W.; Fang, J.; Wang, J.; Geng, Y.; Wang, J. Learning Convolutional Ranking-Score Function by Query Preference Regularization. In Proceedings of the International Conference on Intelligent Data Engineering and Automated Learning, Guilin, China, 30 October–1 November 2017; pp. 1–8. [[CrossRef](#)]
14. Arandjelovic, O. Reimagining the central challenge of face recognition: Turning a problem into an advantage. *Pattern Recognit.* **2018**, *83*, 388–400. [[CrossRef](#)]
15. Moon, T.K. The expectation-maximization algorithm. *IEEE Signal. Process. Mag.* **1996**, *13*, 47–60. [[CrossRef](#)]
16. Jian, M.; Cui, C.; Nie, X.; Zhang, H.; Nie, L.; Yin, Y. Multi-view face hallucination using SVD and a mapping model. *Inf. Sci.* **2019**, *488*, 181–189. [[CrossRef](#)]
17. Singh, A.; Sidhu, J.S. Super Resolution Applications in Modern Digital Image Processing. *Int. J. Comput. Appl.* **2016**, *150*, 6–8. [[CrossRef](#)]
18. Zhang, H.; Zhang, Y.; Huang, T.S. Efficient sparse representation based image super resolution via dual dictionary learning. In Proceedings of the IEEE International Conference on Multimedia and Expo, Barcelona, Spain, 11–15 July 2011.
19. Chen, Y.; Tai, Y.; Liu, X.; Shen, C.; Yang, J. FSRNet: End-to-end learning face super-resolution with facial priors. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 1874–1883.
20. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. GANs trained by a two time-scale update rule converge to a local nash equilibrium. *Adv. Neural Inf. Process. Syst. Nips* **2017**, *30*, 6626–6637.
21. Lakshminarayanan, K.; Ramesh, G.P. Vlsi Architecture for 2D-Discrete Wavelet Transform (DWT) Based Lifting Method. *Indian J. Public Health Res. Dev.* **2017**, *8*, 284–289.
22. Lee, C.-H.; Zhang, K.; Lee, H.-C.; Cheng, C.-W.; Hsu, W. Attribute augmented convolutional neural network for face hallucination. In Proceedings of the IEEE Computer Vision and Pattern Recognition Workshop—New Trends in Image Restoration and Enhancement, Salt Lake City, UT, USA, 18–22 June 2018.

23. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326. [[CrossRef](#)]
24. Xie, C.; Liu, Y.; Zeng, W.; Lu, X. An improved method for single image super-resolution based on deep learning. *Signal. Image Video Process.* **2019**, *13*, 557–565. [[CrossRef](#)]
25. Yang, J.; Wright, J.; Huang, T.S.; Ma, Y. Image Super-Resolution via Sparse Representation. *IEEE Trans. Image Process.* **2010**, *19*, 2861–2873. [[CrossRef](#)]
26. Jiang, J.; Yu, Y.; Tang, S.; Ma, J.; Aizawa, A.; Aizawa, K. Context-patch face hallucination based on thresholding locality-constrained representation and reproducing learning. *IEEE Trans. Cybern.* **2018**, *50*, 324–337. [[CrossRef](#)]
27. Jiang, J.; Chen, C.; Huang, K.; Cai, Z.; Hu, R. Noise robust position-patch based face super-resolution via Tikhonov regularized neighbor representation. *Inf. Sci.* **2016**, *367–368*, 354–372. [[CrossRef](#)]
28. Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 105–114.
29. Zhao, Y.; Wang, R.; Jia, W.; Yang, J.; Wang, W.; Gao, W. Local patch encoding-based method for single image super-resolution. *Inf. Sci.* **2018**, *433–434*, 292–305. [[CrossRef](#)]
30. Chang, H.; Yeung, D.-Y.; Xiong, Y. Super-resolution through neighbor embedding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Washington, DC, USA, 27 June–2 July 2004; pp. 1275–1282.
31. Ma, X.; Zhang, J.; Qi, C. Hallucinating face by position-patch. *Pattern Recognit.* **2010**, *43*, 2224–2236. [[CrossRef](#)]
32. Jha, S.; Son, L.H.; Kumar, R.; Priyadarshini, I.; Smarandache, F.; Long, H.V. Neutrosophic Image Segmentation with Dice Coefficients. *Measurement* **2019**, *134*, 762–772. [[CrossRef](#)]
33. Kaur, S.; Bansal, R.K.; Mittal, M.; Goyal, L.M.; Kaur, I.; Verma, A.; Son, L.H. Mixed pixel decomposition based on extended fuzzy clustering for single spectral value remote sensing images. *J. Indian Soc. Remote Sens.* **2019**, *47*, 427–437. [[CrossRef](#)]
34. Son, L.H.; Tuan, T.M.; Fujita, H.; Dey, A.; Ashour, A.S.; Ngoc, V.T.N.; Anh, L.Q.; Chu, D.T. Dental Diagnosis from X-Ray Images: An Expert System based on Fuzzy Computing. *Biomed. Signal. Process. Control.* **2018**, *39C*, 64–73. [[CrossRef](#)]
35. Ali, M.; Son, L.H.; Khan, M.; Tung, N.Y. Segmentation of Dental X-ray Images in Medical Imaging using Neutrosophic Orthogonal Matrices. *Expert Syst. Appl.* **2018**, *91*, 434–441. [[CrossRef](#)]
36. Hemanth, D.J.; Anitha, J.; Son, L.H.; Mittal, M. Diabetic Retinopathy Diagnosis from Retinal Images using Modified Hopfield Neural Network. *J. Med. Syst.* **2018**, *42*, 247–253. [[CrossRef](#)]
37. Hemanth, J.; Anitha, J.; Naaji, A.; Geman, O.; Popescu, D.; Son, L.H. A Modified Deep Convolutional Neural Network for Abnormal Brain Image Classification. *IEEE Access* **2018**, *7*, 4275–4283. [[CrossRef](#)]
38. Son, L.H.; Thong, P.H. Some Novel Hybrid Forecast Methods Based On Picture Fuzzy Clustering for Weather Nowcasting from Satellite Image Sequences. *Appl. Intell.* **2017**, *46*, 1–15. [[CrossRef](#)]
39. Son, L.H.; Tuan, T.M. Dental segmentation from X-ray images using semi-supervised fuzzy clustering with spatial constraints. *Eng. Appl. Artif. Intell.* **2017**, *59*, 186–195. [[CrossRef](#)]
40. Son, L.H.; Tuan, T.M. A cooperative semi-supervised fuzzy clustering framework for dental X-ray image segmentation. *Expert Syst. Appl.* **2016**, *46*, 380–393. [[CrossRef](#)]
41. Tuan, T.M.; Ngan, T.T.; Son, L.H. A Novel Semi-Supervised Fuzzy Clustering Method based on Interactive Fuzzy Satisficing for Dental X-Ray Image Segmentation. *Appl. Intell.* **2016**, *45*, 402–428. [[CrossRef](#)]
42. Son, L.H. Generalized Picture Distance Measure and Applications to Picture Fuzzy Clustering. *Appl. Soft Comput.* **2016**, *46*, 284–295. [[CrossRef](#)]
43. Ngan, T.T.; Tuan, T.M.; Son, L.H.; Minh, N.H.; Dey, N. Decision making based on fuzzy aggregation operators for medical diagnosis from dental X-ray images. *J. Med. Syst.* **2016**, *40*. [[CrossRef](#)]

