Multimedia Tools and Applications (2022) 81:38887–38919
https://doi.org/10.1007/s11042-022-13160-z

Towards a super-resolution based approach for improved face recognition in low resolution environment

Nalin Singh[1] · Santosh Singh Rathore[2] · Sandeep Kumar[1]

Received: 19 April 2021 / Revised: 23 February 2022 / Accepted: 10 April 2022 /
Published online: 26 April 2022

Abstract

The video surveillance activity generates a vast amount of data, which can be processed to detect miscreants. The task of identifying and recognizing an object in surveillance data is intriguing yet difficult due to the low resolution of captured images or video. The super-resolution approach aims to enhance the resolution of an image to generate a desirable high-resolution one. This paper develops a robust real-time face recognition approach that uses super-resolution to improve images and detect faces in the video. Many previously developed face detection systems are constrained by the severe distortion in the captured images. Further, many systems failed to handle the effect of motion, blur, and noise on the images registered on a camera. The presented approach improves descriptor count of the image based on the super-resolved faces and mitigates the effect of noise. Furthermore, it uses a parallel architecture to implement a super-resolution algorithm and overcomes the efficiency drawback increasing face recognition performance. Experimental analysis on the ORL, Caltech, and Chokepoint datasets has been carried out to evaluate the performance of the presented approach. The PSNR (Peak Signal-to-Noise-Ratio) and face recognition rate are used as the performance measures. The results showed significant improvement in the recognition rates for images where the face didn't contain pose expressions and scale variations. Further, for the complicated cases involving scale, pose, and lighting variations, the presented approach resulted in an improvement of 5%-6% in each case.

Keywords Face recognition · Super resolution · Video surveillance · Face detection

✉ Santosh Singh Rathore
santoshs@iiitm.ac.in

✉ Sandeep Kumar
sandeep.garg@cs.iitr.ac.in

Nalin Singh
nalinsingh1@gmail.com

[1] Department of Computer Science and Engineering, Indian Institute of Technology Roorkee, Roorkee, India

[2] Department of Information Technology, ABV-Indian Institute of Information Technology and Management, Gwalior, India

1 Introduction

Video surveillance applications have grown by leaps and bounds in areas like protecting public monuments, private establishments, and monitoring traffic. This is due to the availability of cheaper hardware and better-performing algorithms. Nowadays, cameras can be found at every corner and intersection. This, in turn, has made them an integral part of maintaining the nearby environment's security. Research in the face recognition field has also developed at a similar pace, and better, sophisticated ways are proposed each day for improving the performance of surveillance systems. Keeping in view the problems faced by face recognition systems, a large amount of time has been devoted to studying unconstrained environment scenarios [1, 32]. These include low lighting, occlusion, low resolution, and noisy scenarios [2, 28, 36].

Viola and Jones [67] has laid down the primitive framework of face recognition systems with an incremental classification of features detected in images. Each classifier is weak and does not denote the whole image accurately. Still, when several of these weak classifiers are taken incrementally, the features can be mapped with a high degree of accuracy, and faces can be identified. Normal digital cameras record videos at a lower resolution than still pictures. Hence, the frames captured from these videos significantly affect face recognition performance. Super-Resolution (SR) is a method that can be adopted to enhance these low-resolution images and video sequences, thereby increasing the face recognition rates [77]. Little work has been done to study the effects of SR on images obtained from a face recognition system [40, 57]. Improving the quality of such images using SR can boost any face recognition technology's performance to a large extent.

This paper proposed a super-resolution enabled system for real-time face recognition in video surveillance. The proposed system aims to detect and collect face images and super resolve them in real-time. Further, the system aims to improve descriptor-count based on the super-resolved faces and explores the effect of noise, scale, and descriptors on face recognition performance. An experimental analysis on the ORL, Caltech, and Chokepoint datasets is performed to evaluate the performance of the presented approach. PSNR and face recognition rate are used as performance evaluation measures. Additionally, a thorough comparison of the proposed system with other state-of-the-art approaches is performed. The results showed an increment in image recognition rates, where the face image didn't contain pose expressions and scale variations. Further, for the complicated cases involving scale, pose, and lighting variations, the presented approach resulted in a 5%-6% performance increment in each case.

1.1 Contributions

Video surveillance involves detecting scene(s) and looking for specific patterns that are indecorous or that may indicate the existence of improper behaviour. The surveillance process includes identifying areas of concern, and by viewing the selected images at appropriate times, it is possible to determine if an improper activity is occurring. However, previously developed face detection systems suffer from low-resolution images or images under severely distorted conditions. This hinders the performance of the video surveillance system. This work focuses on overcoming the problems of low-resolution, blur, and noisy images in face detection systems by employing a super-resolution-based approach. The main contributions of this work include:

1. This paper presents an approach for real-time face recognition and detection in a video surveillance system using the super-resolution.

2. The presented approach addresses the problems of noise, scale, and descriptors on face recognition performance and improves descriptor-count based on the super-resolved faces.
3. An empirical evaluation of the presented approach is performed on three different image datasets, and different performance measures are used to assess the performance of the presented approach.
4. The presented super-resolution-based approach is combined with eigenface and BRISK approaches to overcome the video's low-resolution image constraint.
5. A thorough comparison of the presented approach with original image data and low-resolution image data is carried out to assess the performance improvement of the presented approach.

The rest of the paper is organized as follows. Section 2 discusses the background and related work of face recognition. The presented approach is discussed in Section 3. Section 4 presented experimental objects such as used datasets, performance metrics, and used face detection approaches. The experimental analysis results are discussed in Section 5, followed by conclusions in Section 6.

2 Background and related work

In the face recognition system, image resolution plays a vital role. An image can be divided into two sets, *best resolution* and*minimal resolution*. The best resolution is when the descriptor performs at optimal speeds and provides the best recognition rates. The minimal resolution is the threshold value below which the recognition performance drops sharply. Wang et al. [68] demonstrated the use of facial structure information for face recognition purposes. Some common studies into resolution problem were taken up in [10, 21, 37]. The main takeaway from this was that minimal resolution depends on the system and databases used. Low resolution (LR) images obtained suffered from smaller images and image quality problems. An insufficient number of pixels in the obtained image causes inaccurate descriptions. According to [53], if face size is smaller than 32*24, most conventional methods fail. Depending on focus and illumination, severe blur distortions can degrade image quality and cause misclassifications [54].

Several researchers have developed different approaches for face detection and recognition. Some of them addressed the problems of low-resolution images, while others addressed the issues of blur, noise, and scale of images. In this section, we have discussed different reported works on low-resolution face recognition approaches, super-resolution approaches, and other state-of-the-art approaches for face recognition.

2.1 Super-resolution based approaches for face recognition

Numerous techniques such as MAP-based, example-based, DSR based, S^2R^2 based, and FFD-based have been used for the super-resolution in the face recognition.

MAP based approaches: Given a set of LR-HR pairs, where *I(l)* refers to image in LR and *I(h)* refers to image in HR, this metric calculates *D* (downsampling operator) such that $|I(l) - DI(h)|^2$ is minimized. Capel [16] used MAP-based methods by dividing the face region into six unrelated parts and using PCA to determine important regions for face hallucination. Using Baker's [76] work as a reference, Dedeoglu established spatiotemporal coherence between face images and hallucinated face images to very high magnification

(nearly 16 times). Baker's work also inspired Freeman [12] to integrate a parametric model for hallucination at global face image level and local level. Freeman used PCA linear references to enhance $P(I_l|I_h^g)P(I_g^h)$ thereby getting an optimal face image. The main drawback of the method was that it used explicit down-sampling. Soft and hard constraints were also proposed to beautify faces. Soft constraints made the obtained face image closer to the mean face, whereas the hard constraint was used to faithfully reproduce the discriminating facial details. Noise distributions are assumed to be identical and independently distributed with Gaussian base functions [39].

Example based approaches: This focuses on HR-LR pair similarity and assumes that if a HR image is constructed from linear set of other HR images, then the same could be done in LR domain as well. For this purpose, weights are necessary to balance individual images' contributions properly. Chang et al. [23] used the concept of manifolds for locating similar local geometry patterns in different feature spaces. But authors did not compensate for treating SR as patch-based, thereby losing local details in final images. Park [51] separated face image into texture and PCA-based feature sets and then defined training sets to imitate the model observed. The final LR input image was constructed based on the nearest match in both domains. While dealing with SR using the aforementioned methods. The data domain determines the best possible LR image for reconstruction in the HR domain based on distance or similarity measures. The algorithm domain ensures that the HR image selected belongs to a face. The main limitation arises when it is assumed that changes in HR image are reflected proportionally in the LR domain, as that may not be the case. Hence data constraint optimization fails. Similarly, while recognizing faces using SR algorithms, all the information are not utilized, including class labels in the training set, to enhance accuracy.

DSR based approaches: As the mapping done by humans and machines is different, two constraints are proposed [78] for each scenario. The data constraint is developed to perform the linear mapping between VLR and HR domain, thereby minimizing the error in reconstruction. The constraint provides discriminant analysis for faces as machines use reliable face descriptors for face recognition. Different methods can be used for clustering algorithms with different parameters. In this case, linearity is defined as a measure for clustering pairs in the VLR-HR domain. A set of images is defined as pair P such that all the closest neighbors of this pair have the same linearity. This is ensured by calculating gradients of all possible pairs near location x of the original pair such that gradient difference is minimized. Even if the clusters have been identified for VLR-HR pairs, the correct relationship between pairs is needed to be identified to transform the images to the HR domain. Keeping this in mind a relation factor R is proposed such that when it is applied on $I(l)$ it converts it into $I(h)$. This is unique as earlier methods converted HR to the VLR domain for error determination, resulting in the loss of useful information. The minimization of ε is needed to find a reliable HR image. Face features and discriminants are more important than reliable HR-VLR pairs for machine-based descriptive learning. By using multiple factors for images, error measurements are enhanced, and misclassifications are reduced. The SR algorithm can extract more images with improved data constraints based on linear classifiers. The reconstruction factor tends to overfit the VLR in the HR domain in many cases. This problem is tackled by using linearity as criteria for classification and allowing nonlinear cases to be analyzed as well. The proposed SR algorithm also stands out as the first to use class labels to enhance testing for generic VLR images. Visual quality is detected using RLSR [38] algorithm for which the database CAS-PEAL and YaleB yielded significant improvements. Even in the case of face variations due to external elements, the algorithm performed moderately well.

S^2R^2 **(Simultaneous super-resolution based approaches):** Instead of sequentially reconstructing and then identifying the face in the image, the parameters can be combined

that are then used to classify and enhance the LR image simultaneously. This method is proposed in [25]. Availability of a probe set is assumed, which is also called a test set to which the low-resolution image belongs. The other set is the gallery which contains training data and high-resolution images for the corresponding low-resolution images. The main problem boils down to finding a distance vector such that its magnitude represents the minimum of the distance between the LR and HR image space. For this purpose, authors choose an image x_p from the probe set and calculate its distance from x_g, i.e., an image from the gallery. For the LR scenario, all x_p may not always be available, and hence a need for conversion of LR to HR arises. All LR images are denoted as y_p belonging to the probe set. There are two main ways in which one can perform inter-domain matching. Firstly, an approximate $\bar{x_p}$ can be found corresponding to y_p such that it can be matched in place of actual x_p. For all the pairs (x_g, y_p), the distance metric needs to be minimized, which in turn leads to the determination of parameters mentioned in base cases. The first part of the equation deals with the SR such that the image found out is close to HR space. The second constants denote smoothness for the SR. The third part refers to the features derived from SR and chooses the best among them. This problem can be simplified to determine the weight w such that all the constants are inside the w matrix. The domain is separated into two parts for each set in the gallery and probe. Every part's IDA is calculated and compared for each image pair, and the lowest value score of discriminant is chosen. For calculating α, β, and λ, Powell's method is used. All these values are then combined to form w matrix.

FFD based approaches: Focusing on the face reconstruction problem in non-rigid registration scenarios FFD [31] based SR techniques are proposed. Face distortion and expressions have a major role in accurately registering and reconstructing faces collected from consecutive frames. The FFD technique uses a mesh of control points located on the face image to deform the face and bring it closer to the other view angles for accurate registration. This step is further broken down into local and global registration. As local registration performs precise enhancement, it occurs in the HR grid after B-spline interpolation. The global registration is done in the low-resolution grid with fast and slightly imprecise methods. This multi-level elastic deformation technique performs global deformations to account for expression changes. This precision is further improved using edge information such as SSD. The face edge contour information is between adjacent frames provides accurate registrations. After global registration, a set of sub-image pairs consisting of the global image and the reference image is taken, and correlation coefficients are calculated. If the value of coefficients is very small, price enhancements in HR grids are required. In the end, a POCS algorithm is used for SR reconstruction. The experiments are conducted on a chokepoint video database and record 16% improvement in face recognition accuracy.

Recently, some researchers have used super-resolution (SR) based approaches for different face recognition tasks. Kim et al. [30] proposed an edge and identity preserving network (EIPNet) that uses face SR to reduce distortion by employing a lightweight edge block and identity information. The presented network elaborately restored facial components and generated the high-quality 8 $\times$ scaled SR images. Furthermore, the network successfully reconstructed a 128 $\times$ 128 SR image with 215 fps. The experimental analysis on CelebA and VGGFace2 datasets showed that the presented network outperformed other state-of-the-art methods. Cai et al. [11] have proposed the FCSR-GAN approach based on joint face completion and super-resolution via multi-task learning. The experiments have been performed on CelebA and Helen datasets. Results demonstrated that the proposed approach produced better performance than other state-of-the-art methods for face super-resolution (up to 8 $times$ scale).

Shamsolmoali et al. [59] have used a deep convolution network for surveillance record super-resolution. The presented work aimed to recover the low-resolution objects and points in the surveillance record. The developed model was tested on SCface, and Chokepoint datasets, and PSNR measure was used to evaluate the performance. The results showed that the model produced promising results. Lu et al. [64] presented an approach for very low-resolution (VLR) face recognition and super-resolution based on semi-coupled locality constrained. The presented approach enhances the consistency between VLR and high-resolution local manifold geometries and overcomes the negative effects of one-to-many mapping. The authors have used AR and CMU face recognition datasets to validate the presented approach. The results showed that the proposed method outperformed numerous state-of-the-art SR and recognition algorithms. Some other works such as [55], [14] and [65] also explored the use of super-resolution for the face detection and recognition. Table 1 summarizes the review of the related work.

2.2 Approaches to low resolution face recognition

Low resolution (LR) face recognition methods can be divided into indirect methods and direct methods categories. The indirect method forms high-resolution (HR) images from LR images and then classifies the results using normal HR techniques. The important works that outlined this include Baker et al. [4] and S^2R^2 (Hennings) [25]. Direct methods extract discriminating features from the images independent of resolution. The techniques that follow this are, coupled locality preserving mappings (CLPM) [9] and multidimensional scaling (MDS) [8]. This can be categorized further into resolution robust features and inter-relationship between HR-LR pairs for classifications. Super-resolution methods such as interpolation [69], reconstruction based [5], and learning-based [72] to enhance the images.

Liao et al. [43] presented a JPEG image steganography method based on the dependencies of inter-block coefficients. The inter-block dependencies that describe the interaction among coefficients at the corresponding positions in different discrete cosine transform (DCT) blocks are preserved using the presented method. The experimental analysis has been performed on six brains magnetic resonance imaging (MRI) images. Results showed that the presented method efficiently clustered inter-block embedding changes, improving anti-steganalysis performance. In another work, Liao et al. [42] presented a steganographic embedding function to preserve the correctness and efficiency of the image. The function is utilized to discriminate the image's smoothness. The authors have performed an experimental analysis to validate the presented function by developing and testing special data hiding methods. The results showed that the presented function could perform better than the prior works. Liao et al. [41] presented a separable data hiding method in the encrypted image using compressive sensing and discrete Fourier transform in another similar work. The authors showed that the presented method could generate better image quality when hiding the same embedding capacity through experimental analysis.

Sharma et al. [60] presented the D-FES, a deep face expression recognition system using a recurrent neural network. The presented system could detect six different facial expressions based on the lip structure. The presented system is trained and tested using the JAFFE, MMI, and Cohn-Kanade datasets. Results showed that the presented system had achieved the precision, recall, and f1-score values of 93.8%, 94.5%, and 94.2%, respectively. Kumar et al. [35] presented a superpixel-based color spatial feature approach for salient object detection in another work. The presented approach generates a spatial color feature and combines it with the center-based position before creating the saliency map. The experimental analysis of six datasets showed that the presented approach produced improved

Table 1 A Comparative Analysis of Different Super Resolution Methodologies

Categories	Methods	Summary	Advantages	Drawbacks
Vision- oriented SR	Face Hallucination (Bilinear, Bicubic, Nearest Neighbor) [71, 74]	Using the pixels located in LR image project the intensity values to an HR grid for the nearby pixels.	Fast, Simple	Poor Quality, noise distortions
	MAP (Baker, Freeman, Capel) [3, 70]	Calculate the probability of HR grid containing the same intensity values as in LR by a measure and then use interpolation	Statistical results	Complex two step procedure, computation costs
	Example based [23, 51]	Using pre-trained HR-LR patch pairs replace the close matching LR pairs with HR patch in high dimension	Single step, Better heuristics and quality	Large training set required, May give poor quality for outlier images.
Recognition-oriented SR	S^2R^2 [25]	Use SR and downsampling simultaneously to identify closest matches in an intermediate domain.	Simultaneous SR and recognition	Does not use prior information, poor reconstruction quality
	DSR [38, 78]	Similar to Example based learning yet uses two different parameters for face correction procedures	Multiple learning parameters for human and machine based recognition	Over fitting and complexity
	SR + Registration [3, 75]	Use multiple face images captured at sub pixel shifts to generate HR face image	Edge preservation, good reconstruction quality, fast with parallel implementation	Performs poorly when LR frames are not available

AUC, recall, precision, and F1-score measures. In another similar work, Negi et al. [48] presented a deep neural architecture to detect the face mask amid the Covid-19 pandemic. The presented architecture combines CNN and VGG16 models and is trained on the simulated masked face dataset. The results showed that the best-achieved training, validation, and testing accuracy was 99.47%, 98.59%, and 98.97%, respectively.

A model for predictive analytics for human activities recognition using the residual network has been presented by Negi et al. [47]. The authors have developed a human action recognition system, which uses the ResNet-50 model with transfer learning. The experiments have been performed on the UTKinect Action-3D dataset. The results showed that the presented system yielded better performance than other state-of-the-art methods. Kumar et al. [34] have presented a fast and deep event summarization (F-DES) approach. The presented approach extracts the features, resolves the problem of variations in illumination, removes fine texture details, and detects the objects in a frame. The experimental results showed that the presented F-DES approach has successfully reduced the video content and kept the meaningful information in events. Other similar works for the object-detection and face recognition have been reported by [33, 46, 63].

3 Presented face recognition approach

The main objective of the presented work is to design a system to perform real-time face recognition for a large set of video databases and produces aesthetically better results using the super-resolution system incorporated. Developing an effective face recognition system requires tackling low-resolution and face recognition problems simultaneously. Therefore, it is necessary to use robust methods for each problem separately. Face detection and recognition are the two most important steps in a face recognition system as a whole. Super-resolution (SR) is an important technique for enhancing images to identify faces in the images and videos [15, 62]. Figure 1 shows the overview of the presented face recognition approach. The approach takes a low-resolution video as input and generates detected and recognized the face as the output. The proposed super-resolution-based face recognition system has three main steps; face detection, super-resolution, and face matching and recognition. The description of each component is provided in the upcoming subsections.

3.1 Face Detection

In any face recognition system, the primary step is detecting faces ranging from a wide variety of dimensions. Since in the presented work, very low-resolution face images are considered, therefore, faces with resolutions ranging from (24*24) to (48*48) are captured. The

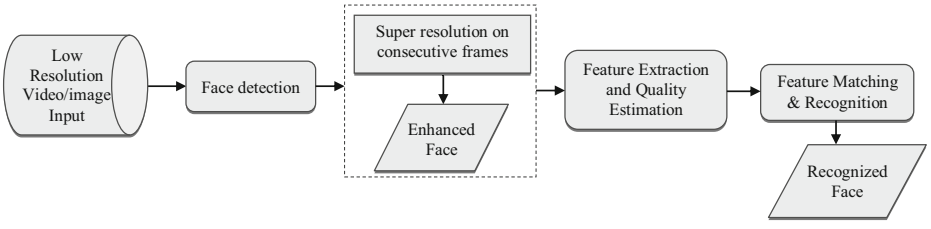

Fig. 1 Overview of the presented face recognition approach

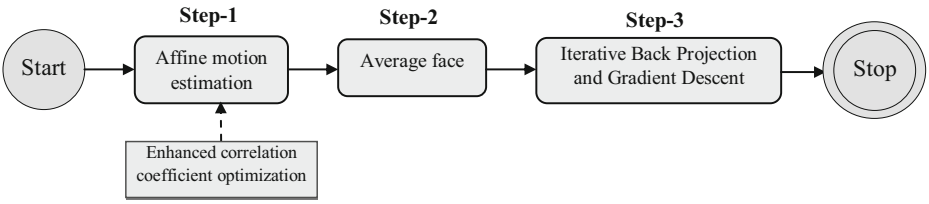

Fig. 2 Super resolution steps

presented face recognition system accepts faces in the range (96*96), this gives us a magnification factor of 2 and 4 to perform super-resolution. First, low-resolution face regions are collected from the input image. Since the input image is of low-resolution (LR), therefore, in the next step, super-resolution is applied to the collected face regions to enhance their quality. A minimum of 8 face images is required to perform super-resolution [52]. If the number is below this limit, it was found that little enhancement in the final image is obtained. Moreover, if many faces are collected, the super-resolution step takes a long time, making it infeasible. The following three steps are applied to collect the face images. (1) Bypass the initial 4 frames as they account for large noise and blur due to motion variations. (2) Collect the frames until the count reaches 8. (3) Pass the collected frame array to the super-resolution module. The pre-processing and collection step takes a negligible amount of time and helps maintain the objective of real-time.

3.2 Super-Resolution

Super-resolution [29] is a technology used to sharpen out-of-focus images or smooth rough edges in images that have been enlarged using a general up-scaling process (such as a bilinear or bicubic process), thereby delivering an image with high-quality resolution. The proposed super-resolution approach uses the *L1-norm* median face method proposed by Sina et al. [61]. Figure 2 depicts the steps of the proposed super-resolution approach. Several steps are applied to perform super-resolution, as discussed below.

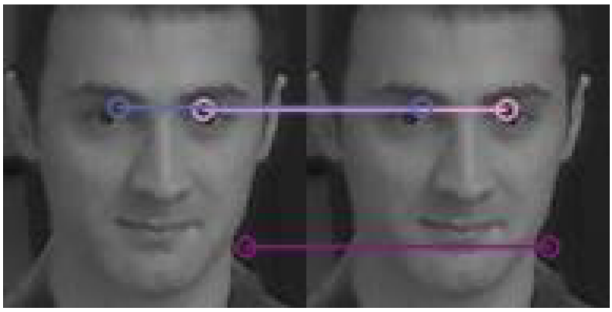

Fig. 3 Results for matching descriptors with minimum distance

3.2.1 Affine motion estimation

To perform super-resolution, it is required that the frames are correctly aligned with each other. For this purpose, the descriptors are used for motion detection that generates a 2*3 matrix for affine motions. This model accounts for rotation, scale, and translation. Since the captured frames are related temporally, the first frame image is assigned as a reference and the motion matrices are calculated correspondingly. An image alignment problem [18] can be expressed as in (1).

$$I_r(x) = I_w(\varphi(x; p)), \forall x \in \tau \tag{1}$$

Where, I_r is the template image, I_w is the warped image, x is the set of coordinates for pixels in template image, and $\varphi(x; p)$ is the set of parametric correspondences for the warped image.

Algorithm-1 depicts the steps of the affine motion estimation process [18]. Figure 3 illustrates an example of the affine motion estimation process. The algorithm performs better than the optical flow optimization techniques proposed by Lucas-Kanade. The efficacy has been proved in the case of 1D translation estimation, and 2D translation estimation in registration [22].

Algorithm 1 Affine Motion Estimation.

Let image1 and image2 be connected by temporal relation with *Detect_perturbations* returning the changes from template image on each image. E*xtract_warp* denotes the warped parameters on the new image. *Matcher* matches and stores the distances between perturbations in two images. Here all frames are registered with respect to first frame i.e. image1 or the template image.

Input: Image1, Image2
Output: Affine transformation matrix
Begin:

Affinemotion (image1, image2)
Keypoint1 ⟵ Detect_perturbations (image1)
Keypoint2 ⟵ Detect_perturbations (image2) ▷ *Key points of both images have been identified after this step*
Descriptor1 ⟵ Extract_warp (Keypoint1, image1)
Descriptor2 ⟵ Extract_warp (Keypoint2, image2) ▷ *Descriptors of both images have been identified after this step*
Matches ⟵ matcher (Descriptor1, Descriptor2)
For each $i \epsilon$ descriptors.rows
Distance[i] ⟵ Matches[i]
End For ▷ *Best matched images have been identified after this step*
Bestmatch ⟵ getleastdistance (Distance[i])
Matrix transform ⟵ getAffineTansform (Bestmatch)
return transform

End

The transformation matrix as proposed in [49] containing the six components is given as follows.

$$\begin{pmatrix} Z_x * cos(a), & -q_1 * sin(a), & d_x \\ q_2 * sin(a) & Z_y * cos(a) & d_y \end{pmatrix}$$

Where, d_x and d_y refer to translations in x and y axis.
Z_x and Z_y are the relative scale operations on x and y axis
'a' is the angle of rotation
'q' is the skew parameter. This causes image to skew on one side

3.2.2 ECC based optimization

When different images of the same scene need to be aligned with respect to geometric distortions, it becomes necessary to consider some objective functions that minimize the overall error rate. Here, an Enhanced Correlation Coefficient (ECC) [18] has been used to solve the affine motion problem. The benefit of using ECC-based optimization is that it is invariant to photometric distortions like brightness and contrast in consecutive frames, and it is fast because of the conversion of the non-linear parametric equation to linear form. The following steps are involved in estimating affine motion by the ECC algorithm.

1. Select a template face image as the reference frame. Here, the first frame is considered the template image.
2. Measure the similarity of the consecutive frames with respect to the first frame and calculate the conversion matrices.
3. Warp the obtained frames with respect to the first frame and store them in the database.
4. Use the SR algorithm to determine the best representation of the face image obtained from all the frames.

3.2.3 Construction of an average image

To perform error minimization, a hypothesis frame is needed to be synthesized. This frame is obtained after aligning the LR frames obtained after affine motion detection and performing suitable transformations. The steps required are mentioned below.

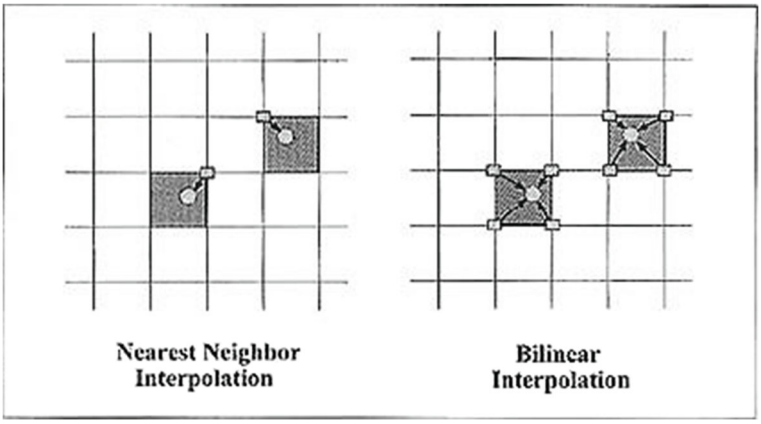

Fig. 4 Interpolation methods for image resizing

1. All the input LR frames are interpolated using the Nearest Neighbour method as shown in Fig. 4. During this, the interpolated pixel is allocated the value of the nearest sample point in the input image. This method is computationally very fast and hence perfect for our goal.
2. After all the LR images are aligned correctly in a high-resolution (HR) grid. The mean of all values in a pixel neighborhood is taken [61]. If the number of images is less than the magnification term (i.e., $N < r^2$), the pixel values are singular, and the mean, median estimator, performs similarly. In other cases, where no singular value is present, no estimate is entered.
3. The above method generates a blurred and noisy high-resolution image.

The result of performing this is a reference frame H that is used as an initial input in the Gradient Descent Iterative Back projection approach.

3.2.4 Iterative back projection

After obtaining affine motion transformations and hypothesis frame H, the main error/cost minimization problem can be reduced to (2) [26].

$$X_{n+1} = X_n - \beta\{\sum_{k=1}^{N} F_k^T H_k^T D_k^T sign(D_k H_k F_k X_n - Y_k) + \lambda \sum_{l=-P}^{P} \sum_{m=0}^{P} \alpha^{|m|+|l|}[I - S_y^{-m} S_x^{-l}] sign(X_n - S_x^m S_y^l X_n)\} \quad (2)$$

Where,
x_n= HR frame as input in n^{th} iteration
x_{n+1} = HR frame obtained after n^{th} iteration
β = Scalar defining step size in direction of gradient
$Y_k = k^{th}$ Low resolution Frame
F_k = Geometric motion operator for the k^{th} LR frame
H_k = Blur operator for the k^{th} LR frame
D_k = Downsampling operator for the k^{th} LR frame

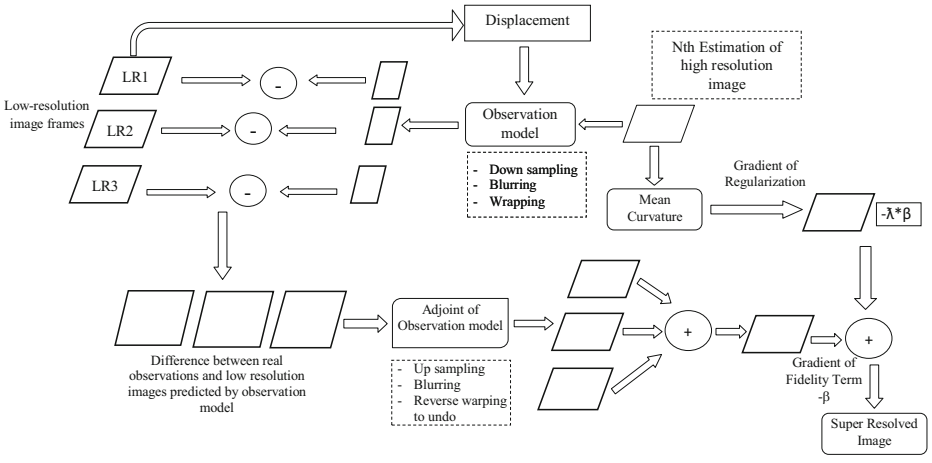

Fig. 5 Super resolution algorithm for misregistration, deblurring and denoising [56]

I = Identity matrix
λ = Regularization factor
α = Smoothing/Decaying addition term
S_x = Shift in x-direction
S_y = Shift in y-direction

3.2.5 Super resolution for misregistration, deblurring, and denoising of image

Figure 5 shows steps of the presented super-resolution approach for misregistration, deblurring, and denoising of image. The hypothesis frame H can be assumed as the first frame X_0 for the algorithm mentioned. The main idea behind the procedure is to minimize the error caused by blur and noise terms using gradient descent. The work reported in [61] proven that when one moves in a direct negative to the gradient, a maximum decrease in cost occurs. The cost function includes errors between predicted LR images and actual LR inputs. Hence, to minimize the cost function (containing the summation of error terms), generation of an error-free HR image is needed iteratively. Algorithm-2 described the working of the super-resolution approach for misregistration, deblurring, and denoising of image [17].

Algorithm 2 Super resolution algorithm for misregistration, deblurring and denoising.

Let *LR_frames* denote the low resolution images and *Shift_matrices* denote the estimated shift for each LR frame. The *Downsampled_matrix* denotes the high resolution to low resolution conversion. *Blur_kernel* denotes the estimation due to camera characteristics. α, β, γ are the parameters determined experimentally for optimal performance. Transpose operation copies value from LR grid to HR grid.

Input: *LR_frames*, H (hypothesis frame), *Shift_matrices*, Downsampling matrix, number of iterations, *Blur_kernel*, α, β, λ.
Output: Super resolved image
Begin:

SuperResolution (No_of_LR_frames, LR_frames, Hypothesis_frame, Shift_matrices, Downsampled_matrices, iterations, Blur_kernel, α, β, γ)
For each $i \epsilon$ iterations ▷ *This step is determining optimal values of super-resolution control parameters*
If ($\lambda > 0$)
Reg_vector = btvregularization (Hypothesis_frame, α)
For each $j \epsilon$ No_of_LR_frames
Sumframe = Sumframe + Transpose * (Blur_kernel * Shift_matrices * Downsampled_matrices) *(Blur_kernel * Shift_matrices * Downsampled_matrices * Hypothesis_frame – LR_framesj)
End For
Hypothesis_frame= Hypothesis_frame – β * Sumframe

Hypotheis_frame= Hypothesis_frame – β * λ * Reg_vector
End For ▷ *A super-resolution image is generated after this step for the input image regions*
Return SuperResolved_image

End

The image representation in the above algorithm is in the form of a vector. Also, the blur downsampling matrices are assumed to be correctly available beforehand in their matrix dimensions. Registration vector contains the information regarding the regularization and helps remove noise while preserving sharp edges. The impact of different parameters in the algorithm is mentioned below.

Regularization factor (λ): When the variation in image intensity is weak smoothing should be encouraged equally in all directions and hence this factors value should be sufficiently high to normalize variations and predict intensities using nearby pixels. On the other hand, if a pixel is surrounded by non-similar pixels, BTV considers it a heavily noisy pixel. It uses a large neighborhood to determine whether smoothing is to be performed or not. This ensures that edges are preserved.

Smoothing/Decaying addition term (α): This term gives respective weights to the pixels surrounding the current pixel in a decaying fashion. This ensures that nearby pixel values are given more weight than pixels located far away.

Scalar defining the step size in the direction of gradient (β): This determines the step size in the direction of the gradient. If the step size is large, the error is minimized in fewer iterations, but if it becomes very large, the solution becomes very prone to incorrect estimations. Hence statistical analysis of results is necessary to optimize their values correctly.

This step results in enhanced face images, which can further be processed for face detection and recognition.

3.3 Face matching and recognition

The super-resolution approach increased the aesthetic aspects of the image. However, it is not always guaranteed that the distinguishing face details in the image are enhanced. Therefore, adoption of a robust recognition method is needed to ensure that face recognition is free from variations in illumination, rotation, scale, and pose. Algorithm-3 described the steps of face matching and recognition.

Earlier works used the Eigenface methods for face recognition, but the time-intensive calculations prove them unviable for real-time implementation [45]. Some other works used the SURF method for face recognition, but the matching rates are low compared to other binary descriptors [24]. BRISK (Binary robust invariant scalable keypoints) method also shown a large performance enhancement on a large range of binary descriptors [38]. The capabilities of these methods are used to develop a face recognition system. The presented face recognition system consists of the following sub-modules.

1. **Key point detection and building descriptor:** For finding a good number of key points using BRISK [38], a threshold of 15 was decided. Only the top 50 key points were used for further processing.
2. **Matching and finding good features:** After obtaining descriptors of key points, they are matched against a set of training image descriptors. Since binary descriptors are stored in a string of 1's and 0's, a simple hamming distance match using the XOR operator is sufficient to measure the distance. The nearest neighbor is selected as the best match having the least distance between descriptors. To make matching more robust, a ratio test was also performed. If the distance of the match is less than *0.7 * distance* of the second nearest neighbor, then this is a good match. Image with the maximum number of good matches will define the person's identity.

Algorithm 3 Face matching and recognition.

Let *Training_features* denote the features obtained from training set of images, *Training_labels* denote the id for a given set of *Training_feature*, *SuperResolved_image* is the image obtained after super resolution, *LR_count* is the number of LR images used in super resolution. K- Nearest neighbour is used as matching model.

Input: Super resolved image, Low resolution image count, Training set id, Training set features
Output: Detected identity
Begin:

Recognize (Training_features, Training_labels, SuperResolved_image, LR_count)
If LR_count>16
Return -1
End if
Else
Test_keypoints=BRISK $\longrightarrow$ getKeypoints (SuperResolved_image)
Test_features=BRISK $\longrightarrow$ getDescriptors (SuperResolved_image, Test_keypoints)
Identity $\longleftarrow$-1
Max_matches $\longleftarrow$ 0
For each person ϵ Training_labels
Matches $\longleftarrow$ 0
End For
For each feature ϵ Training_features for person
Matches=Matcher$\longrightarrow$ knnMatch (feature, Test_features)
Good_matches=ratioTest (Matches)
Matches+=good_matches
End For
If Matches > Max_matches
Max_matches=Matches
Identity=person
End If
End Else
If Max_matches < 20
Identity= Recognize (Training_features, Training_labels, SuperResolved_image, LR_count+2)
End If
Return identity

End

3. **Quality Estimation:** If the maximum number of good matches falls below 20, the control is passed back to super resolution module to enhance the LR image set to 10 images. This recursive process enlarges the LR image set by 2 units each time a sufficient match is not found. The process is repeated a maximum of 4 times, i.e., until the LR image set enhances to 16 images. The face image is termed unqualified for recognition, and a new face image is processed. Algorithm 4 shows the precision enhancement process for matching and recognition of the image.

Algorithm 4 Precision enhancement for matching and recognition.

Let matches be the vector of distances stored while matching super resolved image with a training set. Good matches denotes the matches where *first_neighbor_distance* $< 0.7 *$ *second_neighbor_distance*.

Input: Matches vector containing distances
Output: Number of good matches
Begin:

```
RatioTest (Matches)
Good_matches ⟵ NULL
Ratio ⟵ 0.7
For all match in Matches
If match.first_distance < Ratio * match.second_distance
Good_matches ⟶ add (match)
End If
End For
Return Good_matches
```

End

4 Experimental setup and analysis

This section describes the experimental procedure used for the analysis. Three different datasets, ORL, Caltech, and Chokepoints are used to build and evaluate the performance of the presented approach.

4.1 Used image datasets

Three different datasets, the ORL face dataset, Caltech Palestinian dataset, and Chokepoint video surveillance dataset, have been used to evaluate the performance of the presented super-resolution-based approach.

4.1.1 ORL face dataset

The ORL face database consists of 400 images of 40 subjects, including 10 images of each subject [58]. The images were taken at different times for different subjects, varying the lighting, facial expressions, and facial details. The subjects were imaged in an upright, frontal position against a black, uniform background. Each image is 92x112 pixels and has 256 grey levels per pixel. This dataset has been used to evaluate how the presented super-resolution-based approach improves the image quality and helps in accurate face recognition.

4.1.2 Caltech dataset

The Caltech dataset corresponds to 10 hours of 640x480 30Hz video captured from a vehicle travelling in typical traffic in an urban setting [20]. There are around 250,000 annotated frames (in roughly 137 minute-long parts) with a total of 350,000 bounding boxes and

2300 individual pedestrians. This dataset [48] contains varying number of images for each subject. The dataset generation was similar to ORL database. Instead of taking varying Gaussian noise levels, spike noise levels where altered.

4.1.3 Chokepoint dataset

For testing on video surveillance dataset, ChokePoint dataset is used [73]. This dataset can be used for person identification/verification under real-world surveillance. It consists of face images of persons walking through the pedestrian traffic. Three cameras were mounted over two portals (P1 and P2) to capture the video sequences of the subjects entering (E) or leaving (L) the portals in a natural manner. Images are varied in terms of illumination conditions, pose, sharpness, as well as misalignment due to automatic face localization/detection. The dataset includes 25 subjects in portal 1 and 29 subjects in portal 2. The frame rate is set to 30 fps, and the image resolution is 800X600 pixels. A total of 48 video sequences and 64,204 face images are included in the dataset.

4.1.4 Dataset preparation for the system

Figure 6 shows the data preparation process for the system.

4.2 Performance measures

The PSNR and face recognition rate are used as the performance measures.

1. Peak Signal-to-Noise Ratio (PSNR): It is defined as the ratio between the maximum possible power of an image and the power of corrupting noise that affects the quality of its representation. The PSNR value of an image is computed by comparing the image with an ideal clean image with the maximum possible power. It is defined by (3) [19].

$$PSNR = 10log_{10}\frac{(L-1)^2}{MSE} = 20log_{10}\frac{(L-1)}{RMSE} \tag{3}$$

Here, L is the number of maximum possible intensity levels. MSE is defined by the (4) [13].

$$MSE = \frac{1}{m}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}(O(i,j) - D(i,j)^2) \tag{4}$$

Where O is an original image matrix. D is the degraded image matrix. m shows the numbers of rows of pixels, and i shows the index of that row of the image. n shows the number of columns of pixels, and j shows the index of that column of the image.

2. Face Recognition Rate: It is defined as the number of correctly identified faces in the given images. It is calculated as given in (5).

$$Face\ recognition\ rate = \frac{no.\ of\ correctly\ identified\ images}{Total\ no.\ of\ images} * 100 \tag{5}$$

4.3 Used face recognition methods

4.3.1 Eigenface method

The Eigenface approach is based on identifying the most important vectors that can describe faces in the database, termed as face space [66]. Eigenface accomplishes this by capturing

(a) Noise σ = 15, Enhancement factor=4, Spike Noise=5%

(b) Noise σ = 10, Enhancement factor=4, Spike Noise=5%

(c) Noise σ = 15, Enhancement factor=2, Spike Noise=5%

(d) Noise σ = 10, Enhancement factor=2, Spike Noise=5%

Fig. 6 Dataset Preparation for the system

variations in a large set of training images and comparing it with other images without discarding any information in captured pixels. Each face image can be expressed in terms of a linear combination of *M* eigenfaces. For computational efficiency, only the best faces are chosen for forming these *M* eigenfaces.

4.3.2 Fisherface method

Fisherface method uses projections on linear sub-space to determine the classes of faces. Fisherface [7] method improves the Eigenface technique as it tries to maximize the inter-class differences while simultaneously minimizing the intraclass parameters. This helps in classifying images to a higher degree of accuracy.

4.3.3 Scale invariant feature transform (SIFT)

SIFT [44] method is a highly distinctive descriptor that can match objects with high probability over a large collection of similar objects. This approach can be easily used for recognition purposes considering that it provides high accuracy even for highly cluttered and occluded scenarios. SIFT is robust to orientation and scale. The robustness of SIFT method can be described in terms of scale, noise, and orientation.

4.3.4 Speed up robust features (SURF)

SURF [6] was developed as an alternative to SIFT. The SURF method is based on the two components, SURF detector, and SURF descriptor. SURF detector extracts key points in an image by applying LoG masks to an image at varying scales and then calculates the hessian matrix for that scale. The intensity comparison between scales is measured using integral images as only values within a rectangle are compared. SURF descriptor is a rotation and scale-invariant scheme. Rotation invariance is assured by finding the feature's dominant direction and rotating the sampling window to align with that angle. Scale invariance is assured by sampling the descriptor over a window proportional to the detection window size.

4.3.5 BRISK

BRISK solves low lighting problems, pose variation, and scales using Keypoint Detection, Orientation compensation, and Descriptor Construction [38]. BRISK includes a handcrafted sampling pattern consisting of concentric circles of varying radius originating from the center. This causes the generation of 512 sampling pairs taking into account a key point at each center of a circle. The pairs can further be broken down into long and short pairs. If pair distance is below a threshold, it is a short pair; else, it is a long pair. The long pairs determine the orientation, and short pairs provide intensity comparison.

4.4 Implementation details

All the experiments were performed on an Intel core *i5* 2.4 GHz machine with 4 GB RAM. The platform used was OpenCV/C++ with Ubuntu 10.04 as Operating System. Face detection is performed on videos having low resolutions ranging from 400*300 to 200*150 for each frame. The frame rate is around 25 fps giving Haar-based face detection an average time of 100ms per frame [50]. Since super-resolution using the given frames takes an estimated 1s for 7-8 frames, the detection, super-resolution, and recognition rate are selected as 6 fps. An important thing to note here is that if the video resolution is very small, the super-resolution time decreases drastically, ranging from 500-600ms for 8 frames, and increases the overall processing rate to 10 fps. This entails the possibility of using super-resolution for face recognition in real-time surveillance videos [27]. In the face recognition system, the face image is affected by blur, motion, and noise while being captured by a camera. These parameters are not known accurately, and to perform super-resolution, the motion characteristics have been estimated using descriptors while presuming that the blur kernel is known for the given camera [61].

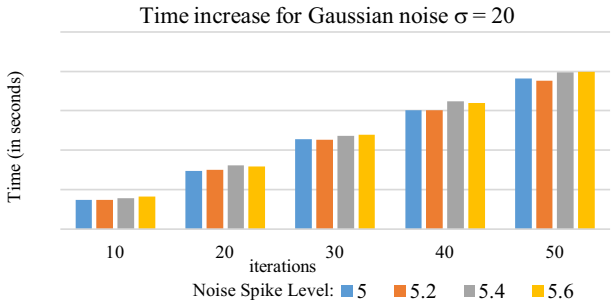

Fig. 7 Time vs iteration scenario for proposed super resolution based approach at noise=20 (Different series are corresponding to different noise spike levels, 5, 5.2, 5.4 and 5.6)

5 Results and comparative analysis

his section discusses the presented face recognition approach results on the ORL, Caltech, and Chokepoint datasets. The proposed super-resolution approach has been applied to datasets with different noise levels to evaluate their robustness and further the face recognition rate. The results for different performance measures are discussed in the upcoming subsections.

5.1 Robustness of the presented approach for the different noise and spike levels

Figure 7 shows the real-time performance benchmark for the presented super-resolution-based approach with successive iterations. Different noise spike levels, 5, 5.2, 5.4, and 5.6 are used. As the iterations have been increased, a linear increase in time is observed. It is true for all the spike levels. When the noise content has been changed from 20 to 10 as shown in Fig. 8, it is observed that a change in noise content does not affect the real-time performance much significantly. The approach works similarly for different noise levels. This proves the robustness of the presented super-resolution approach to the noise.

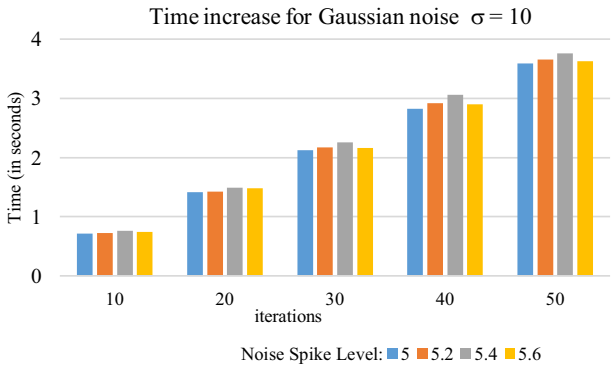

Fig. 8 Time vs iteration scenario for proposed super resolution based approach at noise=10 (Different series are corresponding to different noise spike levels, 5, 5.2, 5.4 and 5.6)

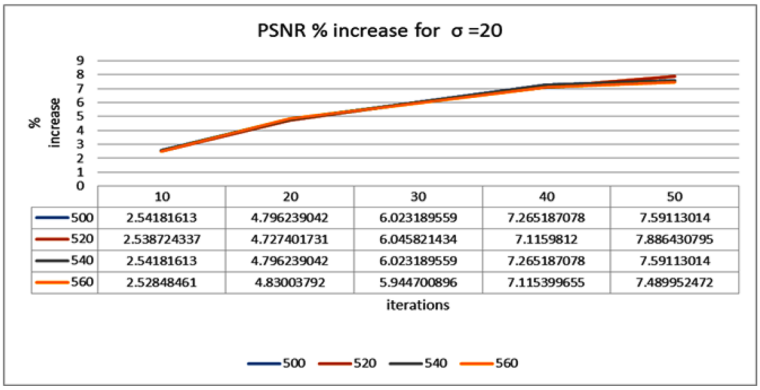

Fig. 9 PSNR *Vs.* iteration graph for proposed super resolution based approach at various spike noise levels with noise=20 (Different series are corresponding to different noise spike levels, 500, 520, 540 and 560)

5.2 PSNR analysis of the presented approach

A PSNR ratio analysis has been performed of the presented approach for the original image to measure the quality of face reconstruction. The better the face reconstruction higher is the PSNR gain. Figures 9 and 10 show the PSNR ratio values of the super-resolution approach with noise level of 20 and 10, respectively. Different noise spike levels have been considered, 500, 520, 540, and 560 for the PSNR analysis. From the figures, it can be observed that the PSNR ratio values are increased with successive iterations. It is true for all the considered noise spike levels. The highest achieved PSNR value is 0.8 approximately for the noise spike level 520. These results showed that the presented super-resolution approach performs better for face reconstruction. The PSNR values increase as the number of iterations increases. The super-resolution approach produced an improved performance for both the considered noise levels.

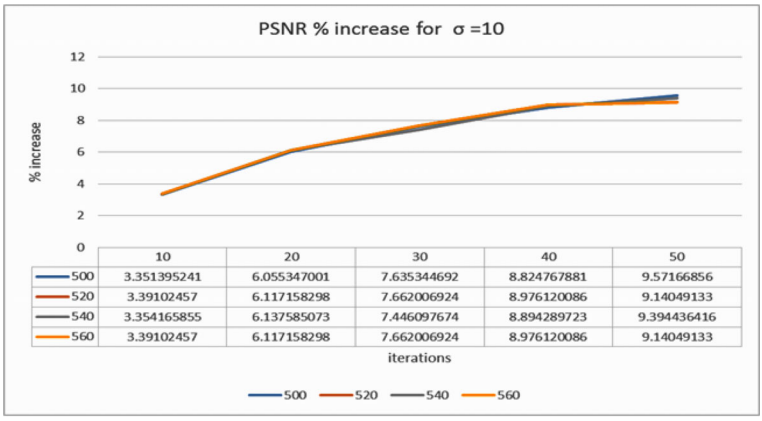

Fig. 10 PSNR *Vs.* iteration graph for proposed super resolution based approach at various spike noise levels with noise=10 (Different series are corresponding to different noise spike levels, 500, 520, 540 and 560)

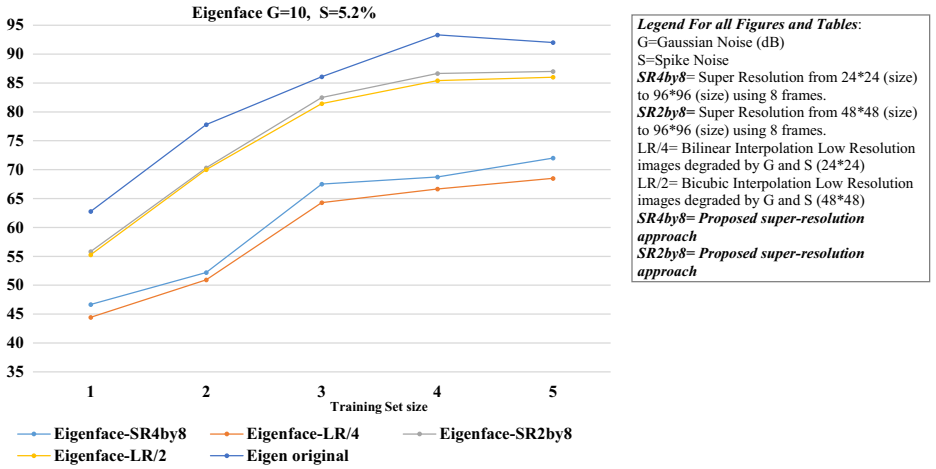

Fig. 11 Face recognition rates on ORL Dataset using proposed super resolution based approach

5.3 Face recognition rate of the presented approach

The super-resolution approach's performance has been evaluated using the face recognition rate measure for the ORL and Caltech datasets. The super-resolution and other face recognition approaches (eigenface and BRISK) have been combined, and the results have been evaluated. Initially, the eigenface and BRISK approaches have been to the original images extracted from the datasets. Then, original images have been degraded by applying noise and spike levels. They are called LR/2 (48*48 size) and LR/4 (24*24 size). These are low-resolution images. The eigenface and BRISK approaches have been applied again on the LR images, and performance has been recorded. Finally, the presented super-resolution approach has been applied to the generated LR/2 and LR/4 images to enhance their quality. They are called SR4by8 (super-resolution from 24*24 to 96*96) and SR2by8 (super-resolution from 48*48 to 96*96). The eigenface and BRISK approaches with the presented super-resolution approach have been applied, and the performance improvement achieved by the super-resolution approach has been recorded. Different training set sizes have been used to determine the worst-case performance of the presented face recognition system. The training set size (K) 1 to 5 is formed by randomly selecting images for training and testing.

Table 2 Face recognition rate on Caltech face dataset using proposed super resolution based approach and other approaches with different training set size, K

	Training Set size (K)	1	2	3	4	5
Original image data	Eigen original	62.7778	77.8125	86.0714	93.3333	92.0
Low-resolution image data	Eigenface-LR/2	55.2778	70.0	81.4286	85.4167	86.0
	Eigenface-LR/4	44.4444	50.9375	64.2857	66.6667	68.5
Data after applying presented super-resolution approach	Proposed Eigenface-SR2by8	55.8333	70.3125	82.5	86.6667	87.0
	Proposed Eigenface-SR4by8	46.6667	52.1875	67.5	68.75	72.0

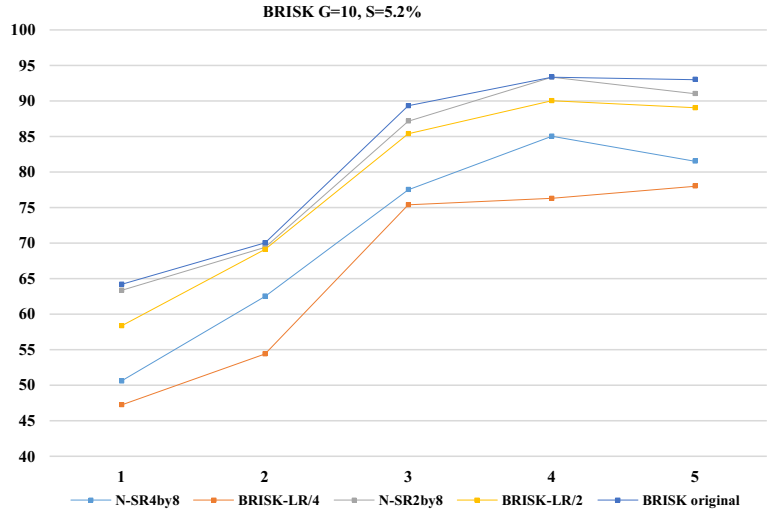

Fig. 12 Face recognition rates on ORL Dataset using proposed super resolution based approach and BRISK

Since each person has 10 images with varying pose and lighting, the performance variations with BRISK and eigenfaces approaches can be determined while demonstrating that super-resolution increases each scenario's performance. Figure 11 shows the face recognition rate on the ORL dataset using eigenface with the presented super-resolution approach. Table 2 shows the face recognition rate on Caltech face dataset with different training set sizes, K using eigenface, and the presented super-resolution. The figure and table show that eigenface produced the best performance on the original images. The performance of the eigenface decreased significantly on the LR images (LR/2 and LR/4). However, the performance has increased significantly when eigenface is applied with the presented super-resolution approach. Similarly, Fig. 12 shows the face recognition rate on the ORL dataset using super-resolution and BRISK approaches. Table 3 shows the Face recognition rate on Caltech face dataset with different training set sizes, K using super-resolution and BRISK approaches. By using BRISK (Table 3 and BRISK (Fig. 12), again, it is observed that BRISK produced the best performance on the original images, and performance decreased for the LR images. However, again, when super-resolution and BRISK are both applied on the LR images, the performance of BRSIK has increased significantly. Further, it is found that BRISK is more resistant to changes in noise and gives better recognition rates under normal conditions. Table 3 shows the face recognition rate on Caltech face dataset

Table 3 Face recognition rate on Caltech face dataset with different training set size, K

	Training Set Size (K)	1	2	3	4	5
Original image data	BRISK original	64.1667	70	89.2857	93.3333	93.0
Low-resolution image data	BRISK-LR/2	58.3333	69.0625	85.3571	90.0	89.0
	BRISK-LR/4	47.2222	54.375	75.3571	76.25	78.0
Data after applying presented super-resolution approach	Proposed BRISK-SR2by8	63.3333	69.375	87.1429	93.3333	91.0
	Proposed BRISK -SR4by8	50.5556	62.5	77.5	85.0	81.5

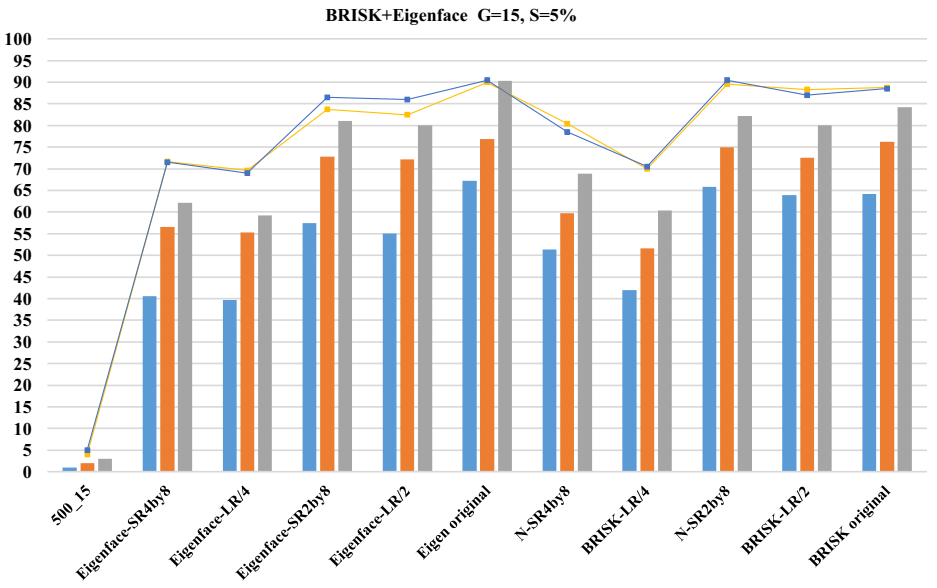

Fig. 13 Face recognition rate on ORL face dataset with different spike noise levels (Different series are corresponding to different noise spike levels)

with different training set sizes, *K* using BRISK. It is observed from the table that for the *K=4*, BRISK and proposed-SR2by8 produced the best performance. These results showed that the presented super-resolution approach improves the performance of both eigenface and BRISK approaches for the low-resolution images. The results produced by combining the presented SR approach with eigenface and BRISK approaches are comparable to the original images' results.

A comparison of BRISK and Eigenfaces method is done in Fig. 13 and Table 4 with the presence of Gaussian noise =15 and varying spike noise levels to demonstrate the effectiveness of the BRISK method in combination with super-resolution. Even LR images with

Table 4 Face recognition rate on Caltech face dataset with different Gaussian noise and spike noise levels

	Training Set Size (K)	1	2	3	4	5
Original image data	BRISK original	64.1667	76.25	84.2857	88.75	88.5
	Eigen original	67.2222	76.875	90.3571	90.0	90.5
Low-resolution image data	BRISK-LR/2	63.8889	72.5	80.0	88.3333	87.0
	BRISK-LR/4	41.9444	51.5625	60.3571	70.0	70.5
	Eigenface-LR/2	55.0	72.1875	80.0	82.5	86.0
	Eigenface-LR/4	39.7222	55.3125	59.2857	69.5833	69.0
Data after applying presented super-resolution approach	Proposed BRISK-SR2by8	65.8333	75.0	82.1429	89.5833	90.5
	Proposed BRISK-SR4by8	51.3889	59.6875	68.9286	80.4167	78.5
	Proposed Eigenface-SR2by8	57.5	72.8125	81.0714	83.75	86.5
	Proposed Eigenface-SR4by8	40.5556	56.5625	62.1429	71.6667	71.5

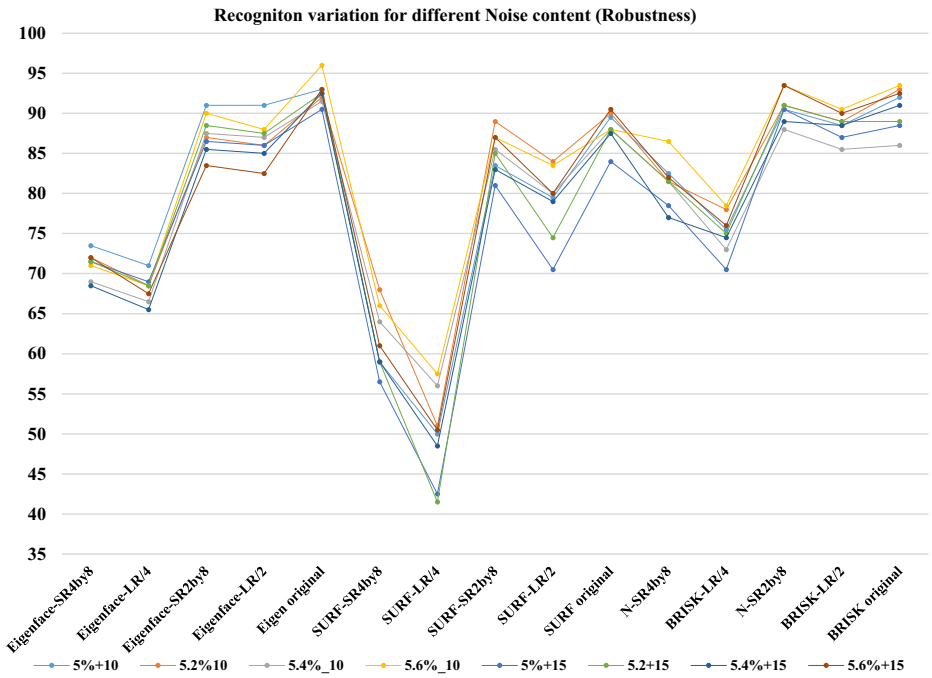

Fig. 14 Face recognition rate on ORL face dataset with different Gaussian noise levels

high noise content can be identified correctly under severe lighting variations. Figure 14 and Table 5 demonstrate the same concept with varying noise levels proving that BRISK is not affected much by the presence of Gaussian and spike noise due to the effects of Multi-frame super-resolution.

5.4 Result analysis on chokepoint dataset

For testing on the video surveillance dataset, the ChokePoint dataset has been used [73]. The ChokePoint dataset has been used by various authors previously, and it is available in the public repository for access. Therefore, this dataset is selected. For comparative analysis, the same dataset is used. Dataset consists of images captured in a real-world unconstrained environment and hence is quite challenging as images contain variation in lighting, pose, expression, etc., with low-quality images. Three cameras were mounted over two portals (P1 and P2) to capture the video sequences of the subjects entering (E) or leaving (L) the portals in a natural manner. Considering that no noise content has been added in from outside in these videos, the performance gain calculation obtained by our technique is required. For training 10 images have been selected randomly from the S1_C1 sequence for each subject. The PSNR values are calculated and reported in Fig. 15. Figure 15(a), (b), and (c) demonstrate that the SR algorithm performs correctly for real-world images where the noise content is inherently present. It is to be noted that only simple translation motions are considered while capturing images. This means that for each expression change or poses variation, translated versions of images have been generated for the scene. Each result shows interpolation to 96*96 level using 15 iterations.

Table 5 Face recognition rate on Caltech face dataset with different Gaussian noise levels

Gaussian noise + Spike noise ->		5%+10	5.2%10	5.4%_10	5.6%_10	5%+15	5.2+15	5.4%+15	5.6%+15
Original image data	Eigen original	93.0	92.0	91.5	96.0	90.5	92.5	92.5	93.0
	BRISK original	92.0	93.0	86.0	93.5	88.5	89.0	91.0	92.5
	SURF original	89.5	90.0	88.0	88.0	84.0	88.0	87.5	90.5
Low-resolution image data	Eigenface-LR/2	91.0	86.0	87.0	88.0	86.0	87.5	85.0	82.5
	Eigenface-LR/4	71.0	68.5	66.5	68.5	69.0	68.5	65.5	67.5
	BRISK-LR/2	88.5	89.0	85.5	90.5	87.0	89.0	88.5	90.0
	BRISK-LR/4	75.5	78.0	73.0	78.5	70.5	75.0	74.5	76.0
	SURF-LR/2	79.5	84.0	80.0	83.5	70.5	74.5	79.0	80.0
	SURF-LR/4	50.0	51.0	56.0	57.5	42.5	41.5	48.5	50.5
Data after applying presented super-resolution approach	Proposed Eigenface-SR2by8	91.0	87.0	87.5	90.0	86.5	88.5	85.5	83.5
	Proposed Eigenface-SR4by8	73.5	72.0	69.0	71.0	71.5	71.5	68.5	72.0
	Proposed BRISK SR2by8	90.5	91.0	88.0	93.5	90.5	91.0	89.0	93.5
	Proposed BRISK SR4by8	82.5	81.5	81.5	86.5	78.5	81.5	77.0	82.0
	Proposed SURF-SR2by8	83.5	89.0	85.5	87.0	81.0	85.0	83.0	87.0
	Proposed SURF-SR4by8	59.0	68.0	64.0	66.0	56.5	59.0	59.0	61.0

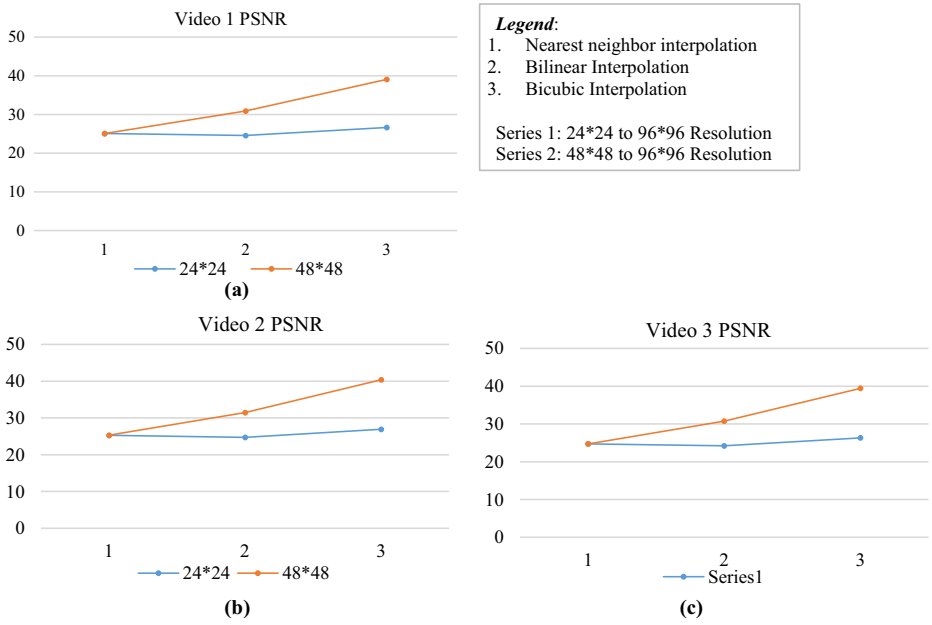

Fig. 15 PSNR results on P1L sequences of ChokePoint dataset (Series1 = 24*24, Series2 = 48*48)

Table 6 shows the recognition rates for the images obtained after super-resolution. It is to be noted that the dataset was trained using random images obtained from each video and then calculating the face recognition rates for each case individually. The training set ranged from 5 images to 9 images per video, showing that face recognition rates increase when more training data is available. The scale change considered was from 24*24 ⇒96*96 and from 48*48⇒96*96.

In unconstrained motion scenarios, the person walking towards the camera may exhibit expression changes. These changes cause distortion in the face image. After applying for ECC-based registration on 8 frames, the face image collected after SR was processed and stored. The training image set consisted of the 7 LR frames obtained initially, with a comparison performed between the initial LR image and obtained SR image. BRISK was used as a descriptor measure for identification. In Fig. 16(a), the video sequence contained a person entering with the camera facing the front. This ensured maximum face coverage. In Fig. 16(b), the subject entered with the camera facing at his side. This also gave good results proving that BRISK was able to extract unique features from the image even if the face was partially visible. A steady increase in descriptors was obtained in all the cases, even for highly complicated motions and pose variations. This shows the effectiveness of the system proposed even for videos (Fig. 16(c)).

5.5 Discussion

Unconstrained face recognition conditions and low-resolution images are serious constraint to the accuracy of automated video surveillance. Images from surveillance camera are typically low contrast and feature a large blur and noise in real-life surveillance circumstances.

Table 6 Face Recognition results on P1L sequences of ChokePoint dataset

	5	6	7	8	9
SR4	73.0769	80.76	87.5	93.26	97.11
LR/4	59.61	71.15	80.76	88.46	96.15
SR2	76.92	86.34	96.15	97.11	99.03
LR/2	75	85.57	90.38	94.23	98.32
Original	81.73	87.5	96.15	99.03	100
SR4	57.29	63.54	72.91	79.167	82.29
LR/4	43.75	59.37	66.66	69.79	72.91
SR2	60.41	71.87	84.37	87.5	91.67
LR/2	53.125	64.58	73.95	83.33	84.37
Original	64.58	72.91	83.33	86.45	90.625
SR4	70.19	78.84	87.5	89.42	93.26
LR/4	62.5	71.15	75.96	81.73	84.61
SR2	83.65	90.38	94.23	97.11	98.2
LR/2	76.9	87.5	91.45	94.21	97.12
Original	82.69	89.42	92.3	95.19	98.07

The existing methods developed for high-resolution images do not generalize well for low-resolution images, and therefore, the face recognition task becomes challenging. This work presented a super-resolution-based approach to enhance the quality of the low-resolution images and improve the accuracy of the face recognition system. The presented super-resolution approach is combined with eigenface and BRISK approaches to overcome the video's low-resolution image constraint. A performance evaluation for three different image and video datasets has been performed. The following observations have been drawn from the experimental analysis.

- The results found that a significant performance improvement for face recognition could be achieved by combining BRISK descriptors with the presented multi-frame super-resolution approach. The presented approach with BRISK (BRISK-SR2by8 and BRISK-SR 4by8) achieved around 5% improvement in the face recognition rate compared to low-resolution images (Table 3 reports these results).
- The noise and spike level analysis showed that the presented approach is robust to the noise. The performance of the approach remains the same for different noise levels. Figures 7 and 8 showed that when the noise spike level increased from 5.0 to 5.6, the performance of the presented approach decreased marginally only.
- The presented approach achieved a high PSNR value for different noise levels. It shows that the presented approach can accurately reconstruct the face from low-resolution images. Figures 9 and 10 reports these results. The highest achieved PSNR value is 0.8 approximately for the noise spike level 520.
- The evaluation using Chokepoint video surveillance dataset demonstrated that the presented approach has successfully extracted the unique features from the blur or noisy images. It confirms the effectiveness of the presented approach.
- Finally, the comprehensive experimental analysis has demonstrated that the presented super-resolution approach increases the efficacy of the system in face recognition and could be used in severe noisy and blurred conditions.

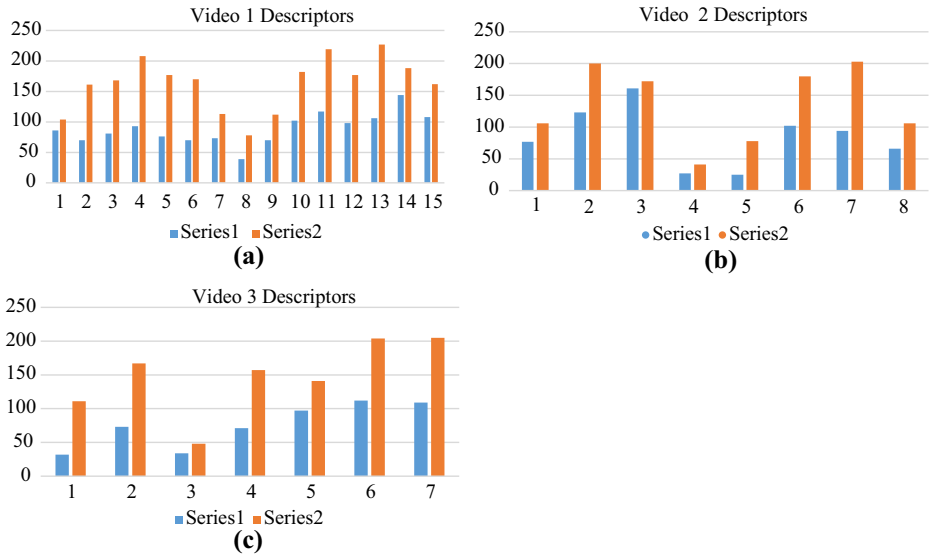

Fig. 16 BRISK descriptor count on P1L sequences of ChokePoint dataset for video 3 descriptors (Series1 = 24*24, Series2= 48*48)

6 Conclusions and future work

For video-based surveillance techniques, recognizing face images from a long distance is crucial yet challenging task due to the low image quality. To address this problem, the low-resolution (LR) images need to be enhanced to make them viable for recognition. The presented work aimed to demonstrate and verify the effect of a multi-frame super-resolution technique on the latest binary descriptor-based face recognition techniques. This work developed a system that could generate a super-resolved image from multiple frames and verify the face recognition performance. The efficacy of the system was determined by training the system on a set of few HR images and then testing them on 24*24 and 48*48 LR images. The experimental analysis was performed on three video surveillance and image datasets, ORL, Caltech, and checkpoints. The results showed an increase in image recognition rates where the face image didn't contain pose expressions and scale variations. Similarly, an increase in BRISK descriptor count for complicated cases involving scale, pose, and lighting variations have been observed. For the LR images, it has been observed that after applying SR and interpolating them to 96*96 resolution, a performance increment of 5%-6% was observed in each case.

In the future, the use of the proposed system for unconstrained face recognition conditions will be explored. Further, a better registration mechanism would be devised to correctly align subsequent frames for each other and switch the super-resolution framework with an example-based super-resolution to enhance speed and visual accuracy.

Declarations Funding and/or Conflicts of interests/Competing interests:

Disclosure of potential conflicts of interest The authors declare that they have no conflict of interest.

Research involving Human Participants and/or Animals This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent This article does not contain any studies with human participants.

References

1. Ahsan MM (2018) Real time face recognition in unconstrained environment. Lamar University-Beaumont
2. Ali W, Tian W, Din SU, Iradukunda D, Khan AA (2021) Classical and modern face recognition approaches: a complete review. Multimed Tools Appl 80(3):4825–4880
3. Ataer-Cansizoglu E, Jones M (2018) Super-resolution of very low-resolution faces from videos. In: British machine vision conference
4. Baker S, Kanade T (2000) Hallucinating faces. In: Proceedings fourth IEEE international conference on automatic face and gesture recognition (Cat. No. PR00580), pp 83–88 IEEE
5. Baker S, Kanade T (2002) Limits on super-resolution and how to break them. IEEE Trans Pattern Anal Mach Intell 24(9):1167–1183
6. Bay H, Tuytelaars T, Van Gool L (2006) Surf: speeded up robust features. In: European conference on computer vision, pp 404–417. Springer
7. Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs fisherfaces: recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell 19(7):711–720
8. Biswas S, Bowyer KW, Flynn PJ (2011) Multidimensional scaling for matching low-resolution face images. IEEE Trans Pattern Anal Mach intell 34(10):2019–2030
9. Bo L, Chang H, Shan S, Chen X (2009) Low-resolution face recognition via coupled locality preserving mappings. IEEE Signal Process Lett 17(1):20–23
10. Boom BJ, Beumer GM, Spreeuwers LJ, Veldhuis RNJ (2006) The effect of image resolution on the performance of a face recognition system. In: 2006 9Th international conference on control, automation, robotics and vision, pp 1–6. IEEE
11. Cai J, Han Hu, Shan S, Chen X (2019) Fcsr-gan: joint face completion and super-resolution via multi-task learning. IEEE Trans Biometr Behav Ident Sci 2(2):109–121
12. Capel D, Zisserman A (2001) Super-resolution from multiple views using learnt image models. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, vol 2, pp II–II. IEEE
13. Chai T, Draxler RR (2014) Root mean square error (rmse) or mean absolute error (mae). Geosci Model Develop Discuss 7(1):1525–1534
14. Chen C, Gong D, Wang H, Li Z, Wong K-YK (2020) Learning spatial attention for face super-resolution. IEEE Trans Image Process 30:1219–1231
15. Chen Y, Phonevilay V, Tao J, Xi C, Xia R, Zhang Q, Yang K, Xiong J, Xie J (2020) The face image super-resolution algorithm based on combined representation learning. Multimed Tools Appl:1–23
16. Dedeoglu G, Kanade T, August J (2004) High-zoom video hallucination by exploiting spatio-temporal regularities. In: Proceedings of the 2004 IEEE computer society conference on computer vision and pattern recognition, 2004. CVPR 2004., vol 2, pP II–II. IEEE
17. Emami S, Suciu VP (2012) Facial recognition using opencv. J of Mobile Embedded Distrib Syst 4(1):38–43
18. Evangelidis GD, Psarakis EZ (2009) An ecc-based iterative algorithm for photometric invariant projective registration. Int J Artif Intell Tool 18(01):121–139
19. Faragallah O, El-Hoseny H, El-Shafai W, El-Rahman WA, El-Sayed HS, El-Sayed E-R, El-Samie FA, Geweid GGN (2020) A comprehensive survey analysis for present solutions of medical image fusion and future directions. IEEE Access
20. Fink M, Perona P (2003) The full images for natural knowledge caltech office db
21. Fookes C, Lin F, Chandran V, Sridharan S (2012) Evaluation of image resolution and super-resolution on face recognition performance. J Vis Commun Image Represent 23(1):75–93
22. Fortun D, Storath M, Rickert D, Weinmann A, Unser M (2018) Fast piecewise-affine motion estimation without segmentation. IEEE Trans Image Process 27(11):5612–5624
23. Gao X, Zhang K, Tao D, Li X (2012) Image super-resolution with sparse neighbor embedding. IEEE Trans Image Process 21(7):3194–3205

24. Hermosilla G, Solar JR, Verschae R, Correa M (2012) A comparative study of thermal face recognition methods in unconstrained environments. Pattern Recogn 45(7):2445–2459
25. Hennings-Yeomans PH, Baker S, Kumar BVKV (2008) Simultaneous super-resolution and feature extraction for recognition of low-resolution faces. In: 2008 IEEE Conference on computer vision and pattern recognition, pp 1–8. IEEE
26. Irani M, Peleg S (1991) Improving resolution by image registration. CVGIP: Graphical Models and Image Processing 53(3):231–239
27. Jebadurai J, Peter JD (2018) Super-resolution of retinal images using multi-kernel svr for iot healthcare applications. Futur Gener Comput Syst 83:338–346
28. Jie X (2021) A deep learning approach to building an intelligent video surveillance system. Multimed Tools Appl 80(4):5495–5515
29. Keren D, Peleg S, Brada R (1988) Image sequence enhancement for super-resolution image sequence enhancement. In: Proceedings of the IEEE Conference on computer vision and pattern recognition. pp 742–746
30. Kim J, Li G, Yun I, Jung C, Kim J (2021) Edge and identity preserving network for face super-resolution. Neurocomputing 446:11–22
31. Kong Y, Zhang S, Cheng P (2013) Super-resolution reconstruction face recognition based on multi-level ffd registration. Optik 124(24):6926–6931
32. Kumar D, Garain J, Kisku DR, Sing JK, Gupta P (2020) Unconstrained and constrained face recognition using dense local descriptor with ensemble framework. Neurocomputing 408:273–284
33. Kumar K, Kumar A, Ayush B (2017) D-cad: deep and crowded anomaly detection. In: Proceedings of the 7th international conference on computer and communication technology, pp 100–105
34. Kumar K, Shrimankar DD (2017) F-des: fast and deep event summarization. IEEE Trans Multimed 20(2):323–334
35. Kumar A, Singh N, Kumar P, Vijayvergia A, Kumar K (2017) A novel superpixel based color spatial feature for salient object detection. In: 2017 Conference on Information and Communication Technology (CICT), pp 1–5. IEEE
36. Kushwaha A, Khare A, Srivastava P (2021) On integration of multiple features for human activity recognition in video sequences. Multimed Tools Appl 80(21):32511–32538
37. Lemieux A, Parizeau M (2002) Experiments on eigenfaces robustness. In: object recognition supported by user interaction for service robots, vol 1, pp 421–424. IEEE
38. Leutenegger S, Chli M, Siegwart RY (2011) Brisk: binary robust invariant scalable keypoints. In: 2011 International conference on computer vision, pp 2548–2555. IEEE
39. Li Y, Lin X (2004) An improved two-step approach to hallucinating faces. In: Third International Conference on Image and Graphics (ICIG'04), pp 298–301. IEEE
40. Li P, Prieto L, Mery D, Flynn PJ (2019) On low-resolution face recognition in the wild: comparisons and new techniques. IEEE Trans Inform Forens Secur 14(8):2000–2012
41. Liao X, Li K, Yin J (2017) Separable data hiding in encrypted image based on compressive sensing and discrete fourier transform. Multimed Tools Appl 76(20):20739–20753
42. Liao X, Qin Z, Ding L (2017) Data embedding in digital images using critical functions. Signal Process Image Commun 58:146–156
43. Liao X, Yin J, Guo S, Li X, Sangaiah AK (2018) Medical jpeg image steganography based on preserving inter-block dependencies. Comput Electr Eng 67:320–329
44. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110
45. Mulyono IUW, Susanto A, Rachmawanto EH, Fahmi A et al (2019) Performance analysis of face recognition using eigenface approach. In: 2019 International Seminar on Application for Technology of Information and Communication (ISEMANTIC), pp 1–5. IEEE
46. Negi A, Chauhan P, Kumar K, Rajput RS (2020) Face mask detection classifier and model pruning with keras-surgeon. In: 2020 5th IEEE international conference on recent advances and innovations in engineering (ICRAIE), pp 1–6. IEEE
47. Negi A, Kumar K, Chaudhari NS, Singh N, Chauhan P (2021) Predictive analytics for recognizing human activities using residual network and fine-tuning. In: International Conference on big data analytics, pp 296–310. Springer
48. Negi A, Kumar K, Chauhan P, Rajput RS (2021) Deep neural architecture for face mask detection on simulated masked face dataset against covid-19 pandemic. In: 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), pp 595–600. IEEE
49. Nguyen N, Milanfar P, Golub G (2001) Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement. IEEE Trans Image Process 10(9):1299–1308

50. Owusu E, Abdulai J-D, Zhan Y (2019) Face detection based on multilayer feed-forward neural network and haar features. Softw: Pract Exp 49(1):120–129
51. Park J-S, Lee S-W (2008) An example-based face hallucination method for single-frame, low-resolution facial images. IEEE Trans Image Process 17(10):1806–1816
52. Park SC, Park MK, Kang MG (2003) Super-resolution image reconstruction: a technical overview. IEEE signal Process Magaz 20(3):21–36
53. Phillips PJ, Grother P, Micheals R, Blackburn DM, Tabassi E, Bone M (2003) Face recognition vendor test 2002. In: 2003 IEEE International SOI Conference. Proceedings (Cat. No. 03CH37443), p 44. IEEE
54. Phillips PJ, Flynn JPJ, Beveridge WR, Scruggs T, O'toole AJ, Bolme D, Bowyer KW, Draper BA, Givens GH, Lui YM et al (2009) Overview of the multiple biometrics grand challenge. In: International conference on biometrics. pp 705–714
55. Qin Z, He W, Deng F, Li M, Liu Y (2019) Srprid: pedestrian re-identification based on super-resolution images. IEEE Access 7:152891–152899
56. Quevedo E, Marrero G, Tobajas F (2016) Approach to super-resolution through the concept of multicamera imaging. In: Book: Radhakrishnan S, ed. Recent advances in image and video coding. pp 101–123
57. Rajput SS, Arya KV (2020) A robust face super-resolution algorithm and its application in low-resolution face recognition system. Multimed Tools Appl 79(33):23909–23934
58. Samaria FS, Harter AC (1994) Parameterisation of a stochastic model for human face identification. In: Proceedings of 1994 IEEE workshop on applications of computer vision, pp 138–142. IEEE
59. Shamsolmoali P, Zareapoor M, Jain DK, Jain VK, Yang J (2019) Deep convolution network for surveillance records super-resolution. Multimed Tools Appl 78(17):23815–23829
60. Sharma S, Kumar K, Singh N (2017) D-fes: deep facial expression recognition system. In: 2017 Conference on Information and Communication Technology (CICT), pp 1–6. IEEE
61. Sina Farsiu M, Robinson D, Elad M, Milanfar P (2004) Fast and robust multiframe super resolution. IEEE Trans Image Process 13(10):1327–1344
62. Singh R, Kushwaha AKS, Srivastava R (2019) Multi-view recognition system for human activity based on multiple features for video surveillance system. Multimed Tools Appl 78(12):17165–17196
63. Solanki A, Bamrara R, Kumar K, Singh N (2020) Vedl: a novel video event searching technique using deep learning. In: Soft computing: theories and applications, pp 905–914. Springer
64. Tao L, Chen X, Zhang Y, Chen C, Xiong Z (2018) Slr: semi-coupled locality constrained representation for very low resolution face recognition and super resolution. IEEE Access 6:56269–56281
65. Tian C, Yong X, Zuo W, Zhang B, Fei L, Lin C-W (2020) Coarse-to-fine cnn for image super-resolution. IEEE Trans Multimed 23:1489–1502
66. Turk MA, Pentland AP (1991) Face recognition using eigenfaces. In: Proceedings of the 1991 IEEE computer society conference on computer vision and pattern recognition, pp 586–587. IEEE Computer Society
67. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, volume 1, pp I–I. IEEE
68. Wang J, Zhang C, Shum H-Y (2004) Face image resolution versus face recognition performance based on two global methods. In: Proceedings of Asia Conference on Computer Vision, vol 47, pp 48–49. Citeseer
69. Wang X, Tang X (2005) Hallucinating face by eigentransformation. IEEE Trans Syst Man Cybern Part C (Applications and Reviews) 35(3):425–434
70. Wang Z, Miao Z, Jonathan WQM, Wan Y, Tang Z (2014) Low-resolution face recognition: a review. Vis Comput 30(4):359–386
71. Wang N, Tao D, Gao X, Li X, Li J (2014) A comprehensive survey to face hallucination. Int J Comput Vis 106(1):9–30
72. Wei W, Liu Z, He X (2011) Learning-based super resolution using kernel partial least squares. Image Vis Comput 29(6):394–406
73. Wong Y, Chen S, Mau S, Sanderson C, Lovell BC (2011) Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition. In: CVPR Workshops, pp 74–81. IEEE, p 2011
74. Yang C-Y, Liu S, Yang M-H (2013) Structured face hallucination. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 1099–1106
75. Yue L, Shen H, Li J, Yuan Q, Zhang H, Zhang L (2016) Image super-resolution: the techniques, applications, and future. Signal Process 128:389–408
76. Zhang W, Cham W-K (2008) Learning-based face hallucination in dct domain. In: IEEE Conference on computer vision and pattern recognition, pp 1–8, p 2008

77. Zhou L, Wang Z, Luo Y, Xiong Z (2019) Separability and compactness network for image recognition and superresolution. IEEE Trans Neural Netw Learn Syst 30(11):3275–3286
78. Zou WWW, Yuen PC (2011) Very low resolution face recognition problem. IEEE Trans on Image Process 21(1):327–340