International Journal of Digital Crime and Forensics
Volume 12 • Issue 3 • July-September 2020

Optimization-Driven Kernel and Deep Convolutional Neural Network for Multi-View Face Video Super Resolution

Amar B. Deshmukh, Vignan University, Guntur, India
N. Usha Rani, Vignan University, Guntur, India

ABSTRACT

One of the major challenges faced by video surveillance is recognition from low-resolution videos or person identification. Image enhancement methods play a significant role in enhancing the resolution of the video. This article introduces a technique for face super resolution based on a deep convolutional neural network (Deep CNN). At first, the video frames are extracted from the input video and the face detection is performed using the Viola-Jones algorithm. The detected face image and the scaling factors are fed into the Fractional-Grey Wolf Optimizer (FGWO)-based kernel weighted regression model and the proposed Deep CNN separately. Finally, the results obtained from both the techniques are integrated using a fuzzy logic system, offering a face image with enhanced resolution. Experimentation is carried out using the UCSD face video dataset, and the effectiveness of the proposed Deep CNN is checked depending on the block size and the upscaling factor values and is evaluated to be the best when compared to other existing techniques with an improved SDME value of 80.888.

KEYWORDS

Deep Convolutional Neural Network, Face Super Resolution, Fractional Grey Wolf Optimizer, Fuzzy Logic, Regression Model

1. INTRODUCTION

Video surveillance is considered as an interesting field of research due to the increasing demand in video surveillance for determining human behaviors and identifying objects. The visual surveillance aims not only to install cameras in the place of human eyes, but also, for achieving complete surveillance attention (Savitha & Ramesh, 2018). Video surveillance is broadly applied in various areas, such as inventory control for retail stores, equipment monitoring for factories, monitoring for intersections, and traffic control security operations for campus, and security surveillance for houses (Porikli et al., 2013; Wang et al., 2018). Tracking and object detection in video surveillance systems mainly depend on background subtraction. Nowadays, video surveillance system uses video compression technology to store images from a total number of cameras to save devices, such as discs, video tapes (Chavda & Dhamecha, 2017). Also, intelligent surveillance systems are utilized

DOI: 10.4018/IJDCF.2020070106

in several areas including protection and the security networks. In various areas, video surveillance face video images are provided for identifying humans. Moreover, the user interest is so far away from the camera, in which the face resolution in the picture is too small for providing information. Due to constrained imaging conditions, it is difficult to obtain high-definition face images. From the generated results, the face images that are captured lose so many meticulous facial features, which are recognized by the users (Qu et al., 2014).

One of the major techniques to improve the resolution of images is super-resolution (SR). SR indicates a class of digital image processing approach, which improve the resolution of an imaging system (Huang et al., 2015). SR combines various low resolution (LR) to generate high resolution (HR) image with best optical resolution. The high frequency content is improved and the degradations produced by the image acquisition are minimized. The LR images are somewhat different, so they consist of dissimilar information about the same scene. SR methods are classified into frequency domain approach, statistical approaches, non-uniform interpolation approach in the spatial domain, and other approaches (Yang et al., 2018). Interpolation approaches based on a single image are sometimes considered as closely related to SR. These techniques lead to a bigger picture size but failed to consider any additional information. In contrast to SR, the high frequency content cannot be recovered. Hence, image interpolation methods are not considered as SR techniques (Dong et al., 2016).

The HR videos are generated from original LR videos and this process is termed as video SR (Huang et al., 2015). Video SR has received considerable attention from both industry and academia. Several HR video devices are developed for storing, producing and transmitting HR videos (Yang et al., 2018). The main aim of SR is to recover HR video or image from its LR finding direct applications that ranges from medical imaging into satellite imaging and also the facilitating tasks like face recognition. Reconstructing HR data from LR input is moreover an extremely ill-posed issue and additional constraint is required for solving (Caballero et al., 2017). The inter-frame temporal relation and the intra-frame spatial relation are the two kinds of relations, which are used for video SR (Liu et al., 2017). In the last two decades, wide variety of SR techniques has been analyzed. The sparse representation (Yang et al., 2010) and self-similar based approaches are used in SR method to reconstruct HR images. Sparse representation techniques are established to evaluate important defect in image processing, often on super resolution and denoising, in which the aim is not for obtaining a compact high-fidelity representation of the observed image, but also for extracting semantic information (Barzigar et al., 2012). One of the representative external example-based SR methods is Sparse-Coding-based (SC) technique (Dong et al., 2016). Depending on cascade of Convolutional Neural Networks (CNNs), Laplacian Pyramid Super-Resolution Network (LapSNR) provides an LR image as input to predict the sub-band residuals in a coarse-to-fine fashion (Lai et al., 2017).

In this paper, a face resolution technique is developed using Deep CNN. Initially, the video frames are extracted from input face video and then, the detection of face is done using Viola-Jones detection algorithm. After that, the detected face image is subjected to the FGWO-based kernel weighted regression model and the Deep CNN, separately. Then, the results obtained from FGWO-based kernel weighted regression model and Deep CNN are combined to achieve the final resolution image using fuzzy logic.

The major contribution of the paper is developing a fusion based technique for multi view super resolution using the FGWO-based kernel weighted regression model and Deep CNN. Finally, the fuzzy logic system combines the results of both the classifiers to generate the SR output with reduced effort, due to the simplicity offered by the fuzzy system.

The rest of the paper is organized as follows: Section 1 depicts a brief introduction to the paper. Section 2 explains literature review and section 3 describes the problem definition. Section 4 discusses the proposed multi-view face resolution technique. Section 5 explains the results and discussions obtained using the proposed method and Section 6 provides the conclusion.

2. LITERATURE SURVEY

This section deals with the existing techniques of face super resolution, explained as follows.

Zeng and Huang (2012) developed face recognition based on non-frontal LR image. The regression method is used for evaluating frontal high resolution (FH) features given by the Non-frontal low resolution (NFL) features. Finally, the feature of FH is utilized for recognizing "one sample per class" gallery. The identities of the input NFL images are obtained by nearest neighboring classifier based on features of FH.

Ren et al. (2012) introduced LR face recognition without using SR pre-processing. Coupled Kernel Embedding (CKE) preserves the locality among neighborhood in the reproducible kernel Hilbert space. CKE analyzed the issue of integrating multimodal data that is critical for conventional methods due to the lack of effective similarity computation. CKE resolves the issue by reducing the inconsistency among the similarity measures produced by their kernel Gram matrices are performed in two spaces.

Izadpanahi and Demirel (2013) developed a video super resolution approach using static and motion areas. Reconstruction method called structure adaptive normalized convolution (SANC) is utilized for generating HR motion and static blocks. Then, discrete wavelet transform (DWT) is utilized for producing HR occluded blocks. The motion blocks and the static blocks are integrated to HR frame. At last, a sharpening process is obtained on HR frame for generating output frame.

Jian and Lam (2015) developed an approach for simultaneous recognition and hallucination of LR faces. In this method, singular values are initially solved to be effective to represent the face images and the values of various resolutions have a linear relation. In this approach, every face image is presented based on SVD. For every LR input face, HR and LR face image pairs are chosen from the face gallery. Depending on these selected LR-HR pairs, the mapping function is performed to interpolate two matrices in SVD for the reconstructing HR face video images.

Ge et al. (2016) developed a spatiotemporal SR approach for improving frame rate and spatial resolution in a hybrid stereo video framework. In this framework, a scene is captured using two cameras to produce two videos, including a high spatial resolution with low-frame rate video and a low spatial resolution by high frame rate video. For low spatial resolution video, the low-resolution frames are super resolved using HR video through stereo matching. Then, the missed frames are interpolated by HR frames by fusing motion compensation and a disparity compensation frame rate.

Borsoi et al. (2017) developed a super-resolution reconstruction (SRR) method with enhanced robustness. An instinctive interpretation is developed for representing proximal-point cost function of the regularized least mean squares (RLMS) gradient descent algorithm. The regularization approach was then derived based on statistical information with quicker convergence of the solution in the subspace is linked to innovations while preserving existing evaluated information.

Lee et al. (2018) developed super-resolution technique using adaptive region High Frequency (HF) enhancement approach. At first, the HF signals are reconstructed using self-similar region in a frame, and then, the signals of HF are improved using various enhancement factors, which is depending on curvature region classification. The method is capable of improving perceptual sharpness based on HF signal enhancement, reducing the feasible visual quality degradation by modifying the improvement factors on the regions.

Sajjadi et al. (2018) had developed end-to-end trainable frame recurrent video super-resolution approach, which uses previously inferred HR estimate to solve the subsequent frame. This naturally encourages temporally consistent results by warping one image in every step and minimizes the computational cost. Moreover, the method is capable to accumulate large number of preceding frames without increasing computational demands.

Table 1 lists the brief description of the existing techniques surveyed.

Table 1. Literature review

Authors	Methods	Advantages	Disadvantages
Zeng & Huang (2012)	Radial Basis Function (RBF) based regression method	Highest recognition rate	The method failed to consider more than one gallery image per person
Ren et al. (2012)	Coupled Kernel Embedding (CKE)	Effective improvement in recognition accuracy	The method failed to focus semi supervised learning problems
Izadpanahi and Demirel (2013)	Structure Adaptive Normalized convolution (SANC)	Higher PSNR	The method failed to consider global registration algorithms for video resolution tasks.
Jian and Lam (2015)	Simultaneous face Hallucination and Identification (SHI)	Effective and excellent performance.	The method not have similar holistic structures and patterns to the LR input
Ge et al. (2016)	Spatiotemporal super resolution method	Increase the frame rate	The method failed to consider disparity maps of the LR-HFR video relative to the HR-LFR video between the picture pairs for synchronous frames.
Borsoi et al. (2017)	Video Super-Resolution Reconstruction (SRR) method	Improved robustness	Large number of iterations are needed
Lee et al. (2018)	Self-similarity-based super-resolution method	Improved perceptual sharpness of the video	multiple contexts are needed for implementation
Sajjadi et al. (2018)	End-to-end trainable Frame Recurrent Video Super-Resolution (FRVSR) framework	Reduced computational cost	The method failed to consider additional memory channel

2.1. Challenges

The challenges faced by the existing techniques are listed as follows:

- In learning-based super resolution algorithm (Huang et al., 2015), the learning of image prior information is the major challenge. Moreover, an effective comparison with the state-of-art techniques is required, which the algorithm in (Huang et al., 2015) failed to proceed;
- The challenge of multiframe color SR is critical than that of monochrome imaging and failed to address monochrome methods for several reasons;
- Another challenge faced by most of the SR methods is face shape metrics. Even though these metrics seem to appear as good entity, they are hard to handle (Lai et al., 2017);
- Generating neighborhood pixel in the super-resolution image of the face region is another challenging task in super resolution. It is difficult to predict shape and size of the neighborhood;
- The resolution algorithm poses computational complexity due to LR images as they pose clarification, face log creation (Yang et al., 2010; Zeng & Huang, 2012) etc.

3. PROBLEM DEFINITION

The primary intention in face video SR is to enhance the resolution of the face region of the input video that consists of multiple numbers of frames. Each frame obtained from the input face video is

subjected to face detection algorithm, namely Viola Jones algorithm. The purpose of SR in detected face region is to offer the required information about the face that helps in recognizing the face in application-oriented environment. The resolution output is obtained by increasing the pixel intensity in the face region. Consider FF be the input video sequence with the face video and it comprises of P number of frames and is denoted as $F = \left\{F_r; 1 \le r \le P\right\} F = \left\{F_r; 1 \le r \le P\right\}$. From each frame F_r obtained from the video, the face region is detected by Grey Wolf Viola-Jones algorithm and is expressed as $J = VJ\left(F_r\right)$; VJ indicates the function representing Viola-Jones algorithm for face detection. The detected face region from the frame is expressed as $J^D = \left\{J_{mn}; 1 \le m \le k; 1 \le n \le g\right\}$; such that the size of the frame after the detection is $k \times g$. The face resolution of the image is performed with increasing size of the face region by the interpolation of the optimal kernel weight matrix generation. Upscaling factor is denoted as $r\left(k * \delta \times g * \delta\right)$, which increases the size of the region and the neighborhood pixels in the face region are computed by optimal kernel weight matrix. SR image of the face region is denoted as $J^{SR} = J_{mn}; 1 \le m \le k * \delta; 1 \le n \le g * \delta$.

4. PROPOSED METHOD OF MULTI-VIEW FACE VIDEO SUPER RESOLUTION

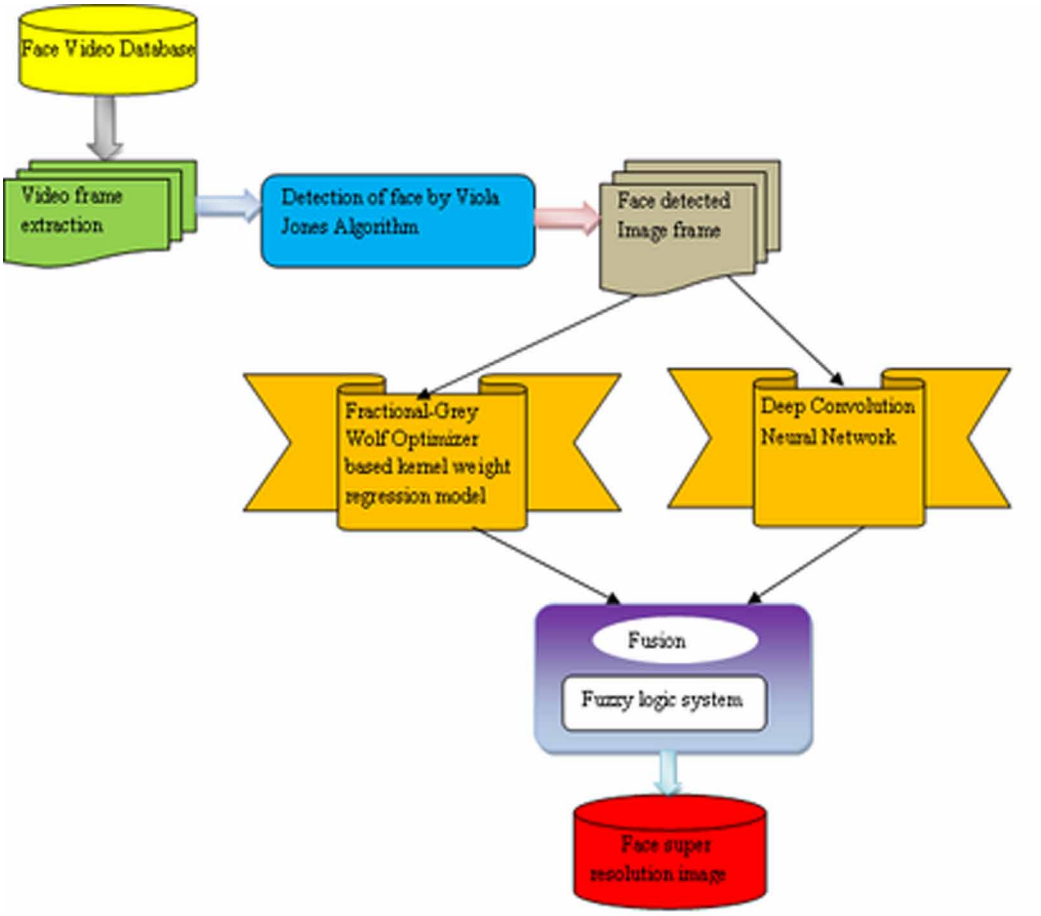
This section presents the proposed method of multi-view face video SR using FGWO based Kernel weight regression model and Deep CNN, with the fuzzy based fusion. SR pixels are generated using optimal weighted kernel regression model and the local patch estimation. Initially, the input face video, which is considered as the input for SR, is read out and the frames from the video are extracted. After extracting the image frames from the input video, the face regions in the frames are detected using Viola-Jones algorithm. The detected face region and the scaling factor are given as the input to the FGWO-based kernel weighted regression model and the Deep CNN (Tu et al., 2017) separately. The results achieved from both the resolution techniques are integrated based on fuzzy logic system to attain the final SR image. Figure 1 depicts the proposed method of multi view face video super resolution.

The input to the proposed face super resolution is the video that is read out using video reader objects. The video accepted for the experimentation is denoted as F with P total number of frames. The face input video file can be of any format, such as Moving Pictures Expert Group 4 (MP4), Operation Good Guys (OGG), Windows Media Video (WMV), Flash Video Format (FLV), High Definition Video (HDV), Audio Video Interleave (AVI), etc. Thus, the frames extracted from F is expressed as, $F_r = \left\{F_1, F_2, ...F_P\right\}$.

4.1. Detection of Face Using Viola-Jones Algorithm

This section presents the extraction of face region from the extracted video frame. Different types of algorithms are used for face detection in previous works. The traditional approaches used for face detection are binary classification techniques, posing high computational complexity issues. Viola-Jones face detection algorithm is utilized for detecting face region from the extracted frames. Compared to other face detection algorithm, the computational complexity of the Viola-Jones is limited that makes it suitable for various applications, like image databases, user interfaces, teleconferencing and so on. The steps involved in the Viola-Jones face detection algorithm: creation of integral image, feature selection by Haar-like a feature, cascading classifiers, detection of face region and Ada-boosting classifier-based feature selection. After classification and Ada-boost classifier-based feature selection task is performed, the face regions are extracted from each frame is expressed as J^D.

Figure 1. Schematic diagram of proposed multi-view face video super resolution



4.2. Multi-View Face Super Resolution Using the Proposed Fuzzy Based Fusion Model

The proposed fuzzy logic based hybrid system developed for multi view SR by combining the FGWO based Kernel weight regression model and Deep CNN is elaborated in this section. In FGWO based Kernel weight regression model, the face image resolution is performed by interpolation of kernel weight matrix with the sub image, where the kernel weight matrix is generated optimally using FGWO. Meanwhile, Deep CNN generates the SR image from the face detected video frame. Finally, the fuzzy logic system fuses the results attained by both the techniques using a fusion parameter, denoted as α .

4.2.1. FGWO Based Kernel Weight Regression Model

This section presents the FGWO kernel weighted regression model (Deshmukh & Usha Rani1, 2017) developed for multi-view face super resolution. Kernel regression model, also known as non-parametric estimation technique, is adapted for evaluating neighboring pixels in the SR image using local characteristic of the data. In the regression model, the face super resolution methods are used to assign higher weights to the coefficient of the matrix which is constructed using the samples. Here,

the optimal kernel matrix is generated using FGWO to achieve the super resolution of the detected face region. Following are the steps involved in the model.

4.2.1.1. Sub Image Generation

Generation of sub-image is performed using the upscale face detected image J_{mn}.The sub-image is generated from J_{mn} is expressed as $J_{y,x}^{sub}$.Consider the face detected image as $J_{m,n};1\le m\le k;1\le n\le g$ and the generation of sub-image with the upscale factor is expressed as $J_{y,x}^{sub};1\le y\le k*\delta;1\le x\le g*\delta$, where δ indicates the upscaling factor.

4.2.1.2. Interpolation of the Kernel Weight Matrix

Once the sub-image is formed, then the unknown pixel intensities produced by sub-image are filled using the neighboring pixel values and the kernel matrix. The expression of the resolution image is given by:

$$J^{FGW}\left(m,n\right)=\frac{1}{w_1*w_2}\sum_{y=1}^{w_1}\sum_{x=1}^{w_2}J^{sub}\left(y,x\right)*X\left(y,x\right) \tag{1}$$

where, $J^{sub}\left(y,x\right)\in\left(m,n\right)$; $X\left(y,x\right)\in\left(m,n\right)$, where $X\left(y,x\right)$ is generated using FGWO.

4.2.1.3 Generation of Optimal Kernel Weight Matrix Using Fractional Grey Wolf Optimizer

This section elaborates the generation of optical kernel weight matrix using FGWO (Deshmukh & Usha Rani1, 2017). FGWO is developed by integrating Fractional calculus (Bhaladhare & Jinwala, 2014) and Grey Wolf optimizer (GWO) (Mirjalili et al., 2014). GWO is a meta-heuristic approach that consists of four hierarchies, namely alpha, beta, delta, where alpha provides the best solution, beta helps for decision making and delta provides the third best solution. The fractional calculus theory is combined with the GWO algorithm to minimize the convergence issue in GWO.

The steps involved in the Fractional grey wolf optimizer are illustrated below:

1. **Parameter Initialization:** The first step involves the initialization of parameters that are used for position optimization. Population size, maximum number of iterations, $\vec{E}$ and $\vec{L}$ are the parameters used for initialization. The vectors $\vec{E}$ and $\vec{L}$ are computed as:

$$\vec{E}=2\vec{s}.\overrightarrow{h_1}.\vec{s} \tag{2}$$

$$\vec{L}=2.\overrightarrow{h_2} \tag{3}$$

where, h_1 and h_2 indicate the random vectors ranging from [0, 1] and s is a parameter that has the value from 2 to 0. Consider population size of the grey wolfs be m and the maximum iterations be $\max_{itr}$.

2. **Population Initialization:** The initial step in the FGWO is the random initialization of the population. The initial population is expressed as:

$$p=\left\{p_1,p_2,\ldots,p_j,\ldots,p_m\right\} \tag{4}$$

3. **Finding the best solution:** The fitness evaluation selects the optimal position vector as the kernel matrix for the generation and interpolation of the SR image. The fitness evaluation function used in the proposed optimization algorithm is SDME (Panetta et al., 2011). The fitness function of the entire search agent is computed as:

$$Fitness = fit\left(p_i\right) \tag{5}$$

Let us assume, $p = \left\{p_1, p_2, p_3, ...\right\}$ be the position of agents. Initially, the position vector p_1 is engaged as the kernel matrix and the interpolation is performed using the sub-image based on upscaling factor, and hence, SR image is generated. After the generation of resolution image, the value of SDME is evaluated by dividing the image into two blocks.

4. **Fractional Solution update:** The location of each search agent is updated using $p_\alpha, p_\beta and p_\delta$, where alpha provides the optimal solution, beta provides the second-best solution and delta provides the third best solution. The solution update is performed using fractional calculus theory. The position of the search agent in GWO is updated as:

$$\vec{p}_{x+1} = \frac{\overrightarrow{p_1} + \overrightarrow{p_2} + \overrightarrow{p_3}}{3} \tag{6}$$

The value of the $\overrightarrow{p_1}, \overrightarrow{p_2}, and \overrightarrow{p_3}$ is expressed as:

$$\begin{aligned}
&\overrightarrow{p_1} = \overrightarrow{p_\alpha} - \overrightarrow{E_1}.N_\alpha; N_\alpha = \overrightarrow{L_1}\cdot\overrightarrow{p_\alpha} - \overrightarrow{p_x} \\
&\overrightarrow{p_2} = \overrightarrow{p_\beta} - \overrightarrow{E_1}.N_\beta; N_\beta = \overrightarrow{L_2}\cdot\overrightarrow{p_\beta} - \overrightarrow{p_x} \\
&\overrightarrow{p_3} = \overrightarrow{p_\delta} - \overrightarrow{E_1}.N_\delta; N_\delta = \overrightarrow{L_3}\cdot\overrightarrow{p_\delta} - \overrightarrow{p_x}
\end{aligned} \tag{7}$$

Grey wolves encircle the prey during the hunt. It is guided by four different parameters beta, delta, gamma and delta. The encircling behavior is depicted as:

$$N = \left|L.p_p\left(t\right) - p\left(t\right)\right| \tag{8}$$

where, $\vec{L}$ represents the vector coefficient and p_p be the location vector of the prey. By the consideration of the fractional calculus (Bhaladhare & Jinwala, 2014), the order of derivative function is changed by the discrete version order χ, which leads to smooth variation. Thus, the updated solution using FGWO is given as:

$$\begin{aligned}
\vec{p}_{x+1} = \frac{1}{3}\Bigg\{ &\chi\,\vec{p}_x - \frac{1}{2}\chi\,\vec{p}_{x-1} + \vec{p}_x\left[\left(\vec{E}_1 + \vec{E}_2 + \vec{E}_3 - 1\right)\right] \\
&+\left[\left(\vec{p}_\alpha + \vec{p}_\beta + \vec{p}_\delta\right) - \left(\vec{E}_1\vec{L}_1\cdot\vec{p}_\alpha + \vec{E}_2\vec{L}_2\cdot\vec{p}_\beta + \vec{E}_3\vec{L}_3\cdot\vec{p}_\delta\right)\right]\Bigg\}
\end{aligned} \tag{9}$$

5. **Iteration:** Once the solution is updated, then the process is repeated until the value of s is reduced;
6. **Termination:** The solution update process for $p_\alpha, p_\beta and p_\delta$ is repeated till the termination condition is satisfied, i.e. $x < \max_{itr}$. The steps from 1 to 5 are repeated for identifying the fittest solution to determine the kernel weight matrix to create the SR face image.

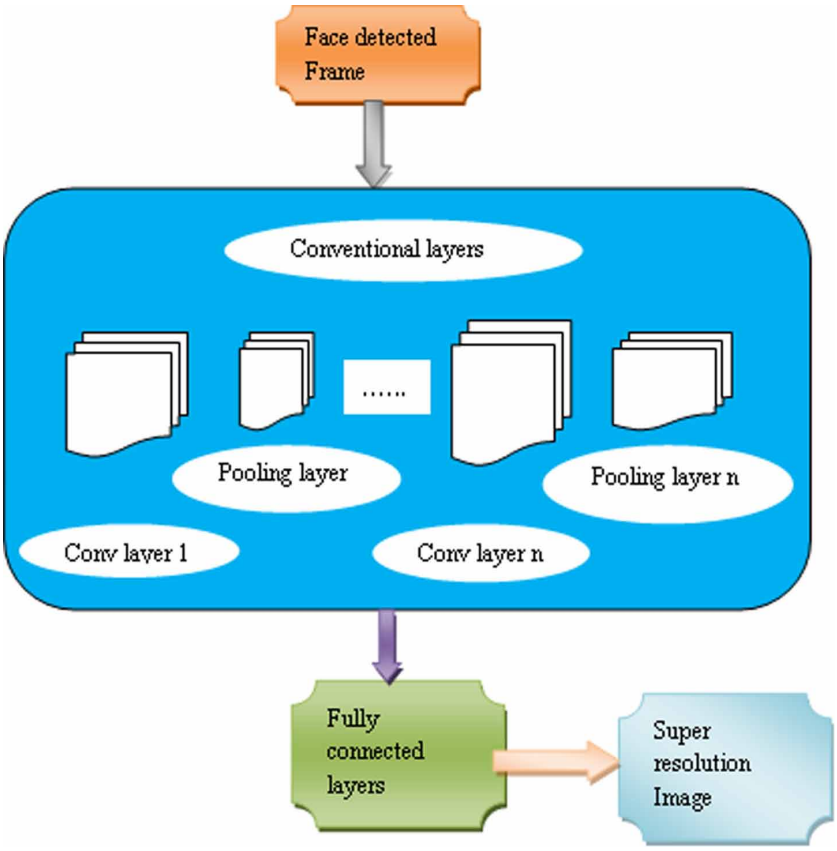
4.2.1.4. Generation of Super-Resolution Image

The super-resolution image is produced with the interpolation of sub image and generated kernel matrix. The unknown pixels present in the sub-image of the kernel matrix are estimated using the weight. The interpolation is performed based on the size of K. Similarly, all the unknown pixel values are computed by interpolation that results the SR image. The SR image is the upscaled face detected image by a factor of 2 with all the neighboring pixel values magnifying the detected face image. In a similar manner, the detected face region inside view is also super-resolved.

4.3. Deep Convolution Neural Network for Multi-View Face Super-Resolution

This section elaborates the deep CNN, which is a type of deep learning, used for multi-view face resolution. Deep CNN (Tu et al., 2017) plays a very significant role in learning. The subsampling layers, like pooling layer, fully connected layer and convolutional layer are the layers present in Deep CNN, as shown in Figure 2. In order to decrease the dimension, subsampling operation is performed,

Figure 2. Block diagram of Deep CNN for face resolution image



where the layer of convolution uses convolution operation for sharing the weight. Without the considerations of manually designed features, Deep CNN can learn the features from the input data for the recognition or detection purpose. Various complex functions can be learned more precisely using Deep CNN compared to shallow networks. Therefore, DCNN can offer more sophisticated deep representations rather than hand-crafted features.

- **Conv layers:** Let us consider the input to the deep CNN is F and hence, the output of the conv layer is expressed as:

$$F^D = \left(A_w^D\right)_{k,g} + \sum_{h=1}^{X_1^{h-1}} \sum_{\tau=-X_1^D}^{X_1^D} \sum_{\upsilon=-X_2^D}^{X_2^D} \left(\varpi_{w,h}^D\right)_{\upsilon,\tau} * \left(F^{D-1}\right)_{k+\tau,g+\upsilon} \tag{10}$$

where, the convolutional operator that paves way to obtain the local patterns using alternative conv layers is denoted as $*$. The output obtained from the preceding $\left(D-1\right)^{th}$ layer forms the input to the D^{th} conv layer. Let us consider the weights to the conv layers be $\varpi_{w,h}^D$, which denotes the D^{th} conv layer weights and the bias of D^{th} conv layer is represented as, A_w^D. Let us consider h, τ, and υ as the notations of feature maps. The neurons in conv layers are arranged in 3-dimensions along the depth, height and width so as to extract the features from all the dimensions of the ReLU layer, which uses an element-wise activation function to simplify the computation using the removal of negative values. The output from the D^{th} layer is the activation function of the preceding $\left(D-1\right)^{th}$ layer, and is expressed as:

$$F^D = Actfn\left(F^{D-1}\right) \tag{11}$$

- **POOL layers:** It is a non-parametric layer with no weights and bias so that it undergoes a fixed operation;
- **Fully connected layers:** The abstract features obtained from the Pooling layers are fed to the fully connected layer. The output obtained from the fully connected layer is expressed as:

$$C_w^D = \delta\left(F^D\right) \text{with } F^D = \sum_{h=1}^{X_1^{h-1}} \sum_{\tau=-X_1^D}^{X_1^D} \sum_{\upsilon=-X_2^D}^{X_2^D} \left(\varpi_{w,h}^D\right)_{\upsilon,\tau} \left(F^{D-1}\right)_{k+\tau,g+\upsilon} \tag{12}$$

The output obtained from the Deep CNN is denoted as J^{DCNN}. The resulting images obtained from FGWO and Deep CNN are subjected to fuzzy logic based fusion model.

4.4. Fuzzy Logic Based Fusion Model

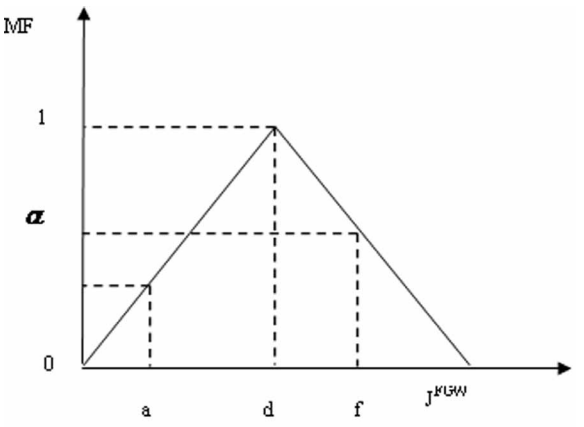
This section explains the fuzzy logic based fusion model (Ravi & Khare, 2016) for SR. The steps involved in fuzzy systems are Fuzzification, Defuzzification and Classification. In the classification step, the input data is subjected to fuzzification to transform the numerical data to form linguistic data and then the linguistic data is matched with the fuzzy rules defined in the rules base. After that, the linguistic data are changed to fuzzy score values by defuzzification. Finally, the fuzzy score is utilized for finding the class. It includes fuzzy membership function

and triangular membership function. Here the membership function is chosen for transferring the input data to the fuzzified value. Triangular membership function comprises of three vertices a, d and f, where a denotes the lower boundary, d be the centre with membership degree as one, and f denotes the upper boundary with the membership degree as zero. The membership function gives the value of fusion parameter that is used to fuse the results of FGWO-based kernel weight regression model and Deep CNN. The formula to calculate the fusion parameter is given as below:

$$\alpha = \begin{cases} 0 & if\ z \le a \\ \dfrac{z-a}{d-a} & if\ a \le z \le d \\ \dfrac{f-z}{f-d} & if\ d \le z \le f \\ 0 & if\ z \ge f \end{cases} \tag{13}$$

Triangular membership functions with defined parameters and their values are depicted in Figure 3. It contains pixel values primarily have membership values close to 1 in one class and member ship

Figure 3. Triangular membership function



values close to 0 in the other classes. The number of membership value depends on the total number of intervals. The maximum and the minimum limit, i.e. the values of f and a are determined by the intermediate value of both and the value of d is calculated accordingly.

Depending on the values assigned to the parameters, α can be obtained and thereby, the fuzzy bound can be computed. For the enhancement of the resolution of the video frame, each pixel location will be computed based on the outputs of both the classifiers with the fuzzy logic as:

$$Y_{jt} = \alpha J_{jt}^{FGW} + (1-\alpha) J_{jt}^{DCNN} \tag{14}$$

where, J_{jt}^{FGW} is the location (j,t) of SR video frame obtained using FGWO-based kernel weight regression model and J_{jt}^{DCNN} is that obtained using Deep CNN. The advantage of fuzzy logic in handling the issues with incomplete information makes the proposed technique more flexible and thereby, generates improved results.

5. RESULTS AND DISCUSSION

The results and discussion of the proposed multi-view face super-resolution technique using hybrid model are demonstrated in this section with an effective comparative analysis to prove the effectiveness of the proposed method.

5.1. Experimental Setup

The experimentation of the proposed technique of Multi-view super resolution is performed in the system with 2 GB RAM, Intel i-3 core processor, Windows 10 Operating System. The proposed method is executed in MATLAB.

5.2. Dataset Description

"UCSD" taken from face video database (The UCSD face video Database taken from http://vision.ucsd.edu/datasets/leekc/disk2/VideoDatabase/testing/, accessed on September 2018) is employed for the experimentation, which is easily available. The face video in the database is recorded in an indoor environment at 15 frames per second with the resolution size 640×480. Each frame in the video sequence exists for 15 seconds. The experimentation is done using four videos from the datasets considering multiple views of the person.

5.3. Evaluation Metric

The performance of the proposed technique is computed using SDME (Panetta et al., 2011) metric as:

$$SDME = \frac{-1}{N_1 * N_2} \sum_{m=1}^{N_1} \sum_{n=1}^{N_2} 20 \ln \left[\frac{J_{\max,m,n}^{SR} - 2 * J_{centre,m,n}^{SR} + J_{\min,m,n}^{SR}}{J_{\min,m,n}^{SR} + 2 * J_{centre,m,n}^{SR} + J_{\min,m,n}^{SR}} \right] \tag{15}$$

where, N_1 and N_2 denotes the blocks of the SR image, $J_{\min,m,n}^{SR}$ and $J_{\max,m,n}^{SR}$ represents the minimum and maximum values of every pixels in block separately, and $J_{Centr,m,n}^{SR}$ be the intensity of the center pixel in every block.

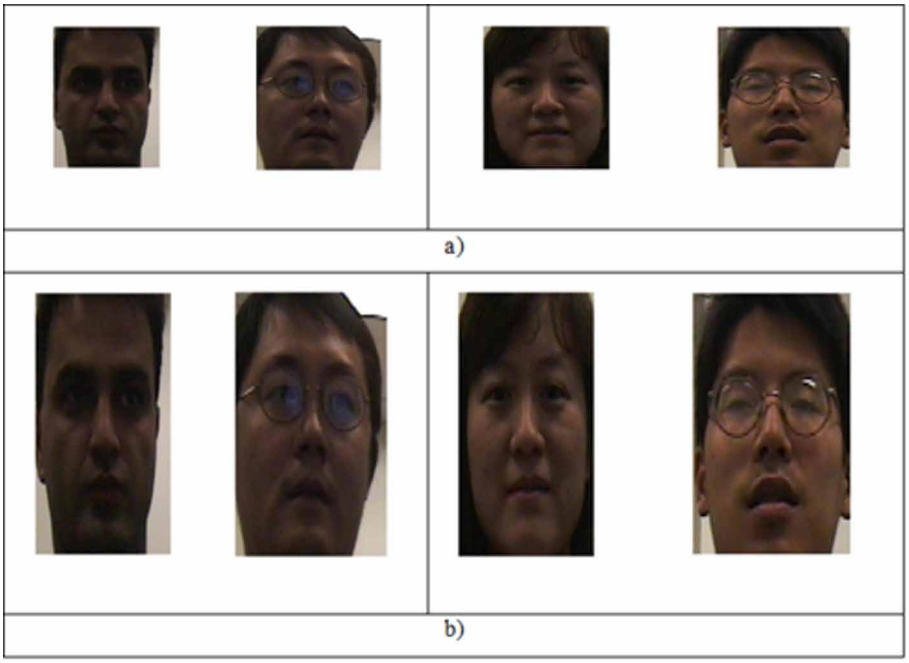
5.4. Experimental Results

The experimental outputs obtained from the proposed face resolution technique are discussed in this section. Figure 4 depicts the experimental outputs obtained from the proposed method. Figure 4a) depicts the sample input video frames and Figure 4b) depicts the SR output attained as a result of the proposed technique that combines FGWO and Deep CNN.

5.5. Competing Methods

The methods like nearest interpolation (Dong et al., 2011), bicubic interpolation and bilinear interpolation (Yang et al., 2008), multikernel regression (Marquina & Osher, 2008), GWO (Mirjalili et al., 2014), FGWO (Deshmukh & Usha Rani1, 2017) are used for the comparison with the proposed FGWO+ Deep CNN for the analysis.

Figure 4. Experimental outputs of the proposed face SR technique: (a) Input video frame; (b) SR output



5.6. Comparative Analysis

The comparative analysis of the proposed FWGO classifier by evaluating the performance of other comparative techniques is presented in this section. The comparative analysis is performed by changing the upscaling factors and the block sizes, and the results are evaluated based on metrics, such as, Second Derivative like Measure of Enhancement (SDME).

5.6.1. Based on Upscaling Factor

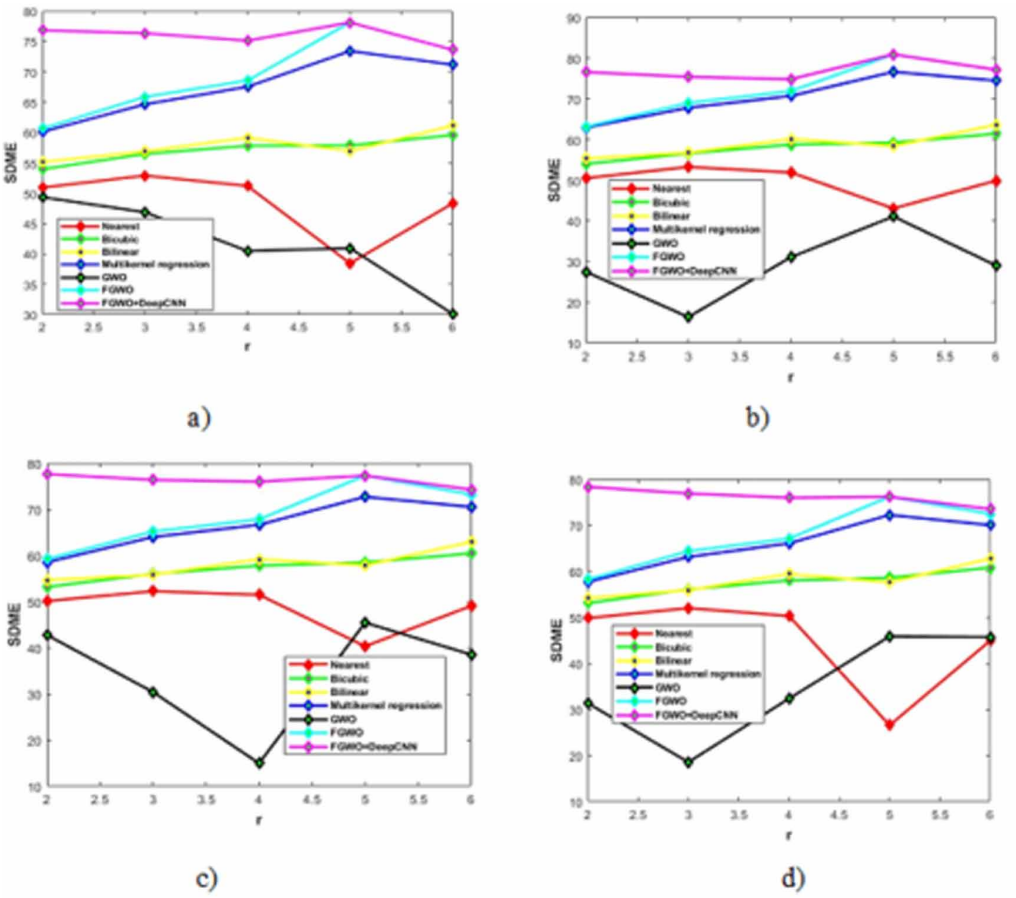
The comparative analysis of the methods is discussed based on SDME using Figure 5 for varying upscaling factors, denoted as r. The analysis based on SDME for video 1 is shown in Figure 5a. When the value of r is fixed to 2, the values of SDME for existing techniques, like nearest interpolation, bicubic interpolation, bilinear interpolation, multikernel regression, GWO, FGWO and proposed FGWO+Deep CNN are 50.902, 53.928, 55.119, 60.104, 49.318, 60.630, and 76.801. Similarly, for the value of r is fixed to 5, the SDME values measured by nearest interpolation, bicubic interpolation, bilinear interpolation, multikernel regression, GWO, FGWO and proposed FGWO+Deep CNN are 51.179, 57.787, 59.101, 67.512, 40.467, 68.532, and 75.099. It is clear that the proposed method acquired a maximal value of SDME when compared with the existing works.

The analysis based on SDME with varying the values of r for video 2 is shown in Figure 5b. When the value of r is fixed to 5, the values of SDME for nearest interpolation, bicubic interpolation, bilinear interpolation, multikernel regression, GWO, FGWO and proposed FGWO+Deep CNN are 43.039, 59.184, 58.379, 76.589, 41.059, 80.887, and 80.887. Similarly, for the value of r is fixed to 6, the SDME values measured by nearest interpolation, bicubic interpolation, bilinear interpolation, multikernel regression, GWO, FGWO and proposed FGWO+Deep CNN are 49.844, 61.388, 63.571, 74.424, 29.034, 77.088 and 77.088.

The analysis based on SDME metric for video 3 is depicted in Figure 5c. When the value of r is fixed to 5, the corresponding SDME values obtained by nearest interpolation, bicubic interpolation,

bilinear interpolation, multikernel regression, GWO, FGWO and proposed FGWO+Deep CNN are 40.312, 58.482, 57.862, 72.716, 45.415, 77.317, and 77.317. Similarly, when the value of r is fixed to 3, the SDME values calculated by nearest interpolation, bicubic interpolation, bilinear interpolation, multikernel regression, GWO, FGWO and proposed FGWO+Deep CNN is 52.251, 56.005, 55.808, 64, 30.387, 65.173, and 76.374. The analysis based on SDME metric for video 4 is depicted in Figure 5d.

Figure 5. SDME analysis based on upscaling factor for: (a) Video 1; (b) Video 2; (c) Video 3; and (d) Video 4



When the value of r is fixed to 2, the corresponding SDME values obtained by nearest interpolation, bicubic interpolation, bilinear interpolation, multikernel regression, GWO, FGWO and proposed FGWO+Deep CNN are 49.817, 53.062, 54.178, 57.677, 31.323, 58.1949, and 78.323. For the value of r is fixed to 4, the SDME values calculated by nearest interpolation, bicubic interpolation, bilinear interpolation, multikernel regression, GWO, FGWO and proposed FGWO+Deep CNN are 50.277, 57.984, 59.413, 66.051, 67.079 and 75.960. It is clear that the proposed method acquired a maximal value of SDME for all the considered videos when compared to previous works.

5.6.2. Based on Block Size

The comparative analysis of the methods is analyzed based on SDME for varying block sizes using Figure 6. The analysis based on SDME for video 1 is shown in Figure 6a. When the block size value is

fixed to 2, the values of SDME for nearest interpolation, bicubic interpolation, bilinear interpolation, multikernel regression, GWO, FGWO and proposed FGWO+Deep CNN are 51.179, 57.787, 59.1, 67.512, 36.022, 68.532, and 73.217. Similarly, for the block size value is 5, the SDME values measured by nearest interpolation, bicubic interpolation, bilinear interpolation, multikernel regression, GWO, FGWO and proposed FGWO+Deep CNN are 51.433, 52.514, 53.450, 57.197, 21.205, 57.150, and 72.304, respectively. It is clear that the proposed method acquired a maximal value of SDME when compared with the existing techniques.

The analysis based on SDME with varying block size for video 2 is shown in Figure 6b. When the block size value is fixed to 2, the values of SDME for the techniques, such as nearest interpolation, bicubic interpolation, bilinear interpolation, multikernel regression, GWO, FGWO and proposed FGWO+Deep CNN are 51.868, 58.727, 60.098, 70.660, 33.164, 71.828 and 72.729, respectively. Similarly, for the block size value is 6, the SDME values measured by nearest interpolation, bicubic interpolation, bilinear interpolation, multikernel regression, GWO, FGWO and proposed FGWO+Deep CNN are 49.524, 50.936, 51.767, 26.543, 57.919, and 71.277. The analysis based on SDME metric for video 3 is depicted in Figure 6c. When the block size value is 3, the corresponding SDME values obtained by nearest interpolation, bicubic interpolation, bilinear interpolation, multikernel regression, GWO, FGWO and proposed FGWO+Deep CNN are 56.213, 54.872, 55.602, 62.055, 25.687, 63.091 and 73.073, respectively. Similarly, for the block size value is 5, the SDME values calculated by nearest interpolation, bicubic interpolation, bilinear interpolation, multikernel regression, GWO, FGWO and proposed FGWO+Deep CNN are 50.421, 51.445, 52.472, 56.332, 10.894, 56.221 and 72.577, respectively. It is clearly shown that the proposed method acquired a maximal value of SDME for video 3 when compared with the existing methods.

The analysis based on SDME metric for video 4 is depicted in Figure 6d. When the block size value is 2, the corresponding SDME values obtained by nearest interpolation, bicubic interpolation, bilinear interpolation, multikernel regression, GWO, FGWO and proposed FGWO+Deep CNN are 50.277, 57.985, 59.413, 66.052, 23.389, 67.079 and 73.497, respectively. Similarly, for the block size value is 4, the SDME values calculated by nearest interpolation, bicubic interpolation, bilinear interpolation, multikernel regression, GWO, FGWO and proposed FGWO+Deep CNN are 49.551, 52.340, 52.996, 57.474, 26.063, 57.978, and 71.043, respectively. From the above data, it is clear that the proposed method acquired a maximal value of SDME when compared to existing methods.

5.7. Discussion

Table 2 describes the discussion regarding the maximum values attained by the existing techniques with the proposed technique. The values computed by nearest interpolation method while varying r and block size provides SDME value as 53.232 and 56.295, respectively. Then the values computed by bicubic interpolation method by varying r and block size provides SDME values as 61.388 and 58.727. The values computed by bilinear interpolation while varying r and block size provides SDME values as 62.947 and 60.098, respectively. Similarly, the SDME values measured by multikernel regression method by varying r and block size are 76.589 and 70.660. Likewise, the SDME values measured by GWO algorithm while varying r and blocksize are 49.318 and 41.484 and the values computed by FGWO algorithm by varying r and block size provides SDME values as 78.099 and 71.828. From the comparative discussion, it is concluded that the SDME values measured by the proposed FGWO+Deep CNN algorithm by varying r and block size are 80.888 and 73.497. Thus, the proposed method shows the superior performance than the existing methods.

6. CONCLUSION

This paper presents the proposed technique of multi-view face resolution based on deep learning and FGWO–based kernel weighted regression model. Initially, the video frames are extricated from the input face video and face regions in the frames are detected using Viola-Jones algorithm. The

Figure 6. SDME analysis based on block size for: (a) Video 1; (b) Video 2; (c) Video 3; and (d) Video 4

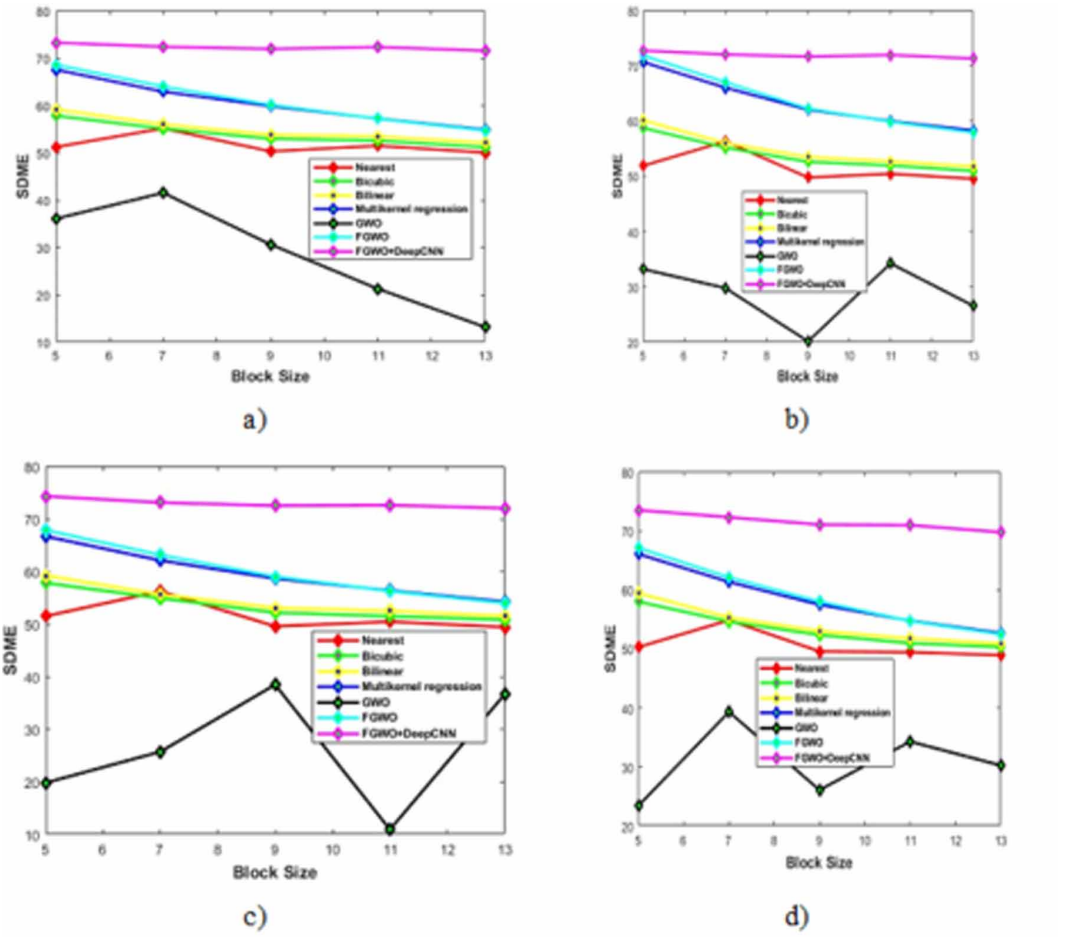


Table 2. Comparative discussion

Methods	SDME	
	Upscaling Factor	Block Size
Nearest interpolation	53.232	56.295
Bicubic interpolation	61.388	58.727
Bilinear interpolation	62.947	60.098
Multikernel regression	76.589	70.660
GWO	49.318	41.484
FGWO	78.099	71.828
FGWO+Deep CNN	**80.888**	**73.497**

detected face image and the scaling factors are fed as the input to the FGWO-based kernel weighted regression model and the deep CNN separately. Then, the results obtained from both the techniques are integrated using fuzzy logic system for final super-resolution image and thereby, the proposed technique has improved visual clarity. SDME is the metric considered for computing the performance of the proposed super resolution method. The proposed technique is experimented using UCSD face video dataset. From the analysis, it is noted that the developed technique enhanced the performance of super-resolution by achieving the maximum SDME value while varying r and block size as 80.888 and 73.497, respectively. Hence the results obtained from the proposed method shows best behaviour than the existing methods

REFERENCES

Barzigar, N., Roozgard, A., Cheng, S., & Verma, P. (2012). A robust super resolution method for video. *Proceedings of Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)* (pp. 1679-1683). Academic Press. doi:10.1109/ACSSC.2012.6489318

Bhaladhare, P. R., & Jinwala, D. C. (2014). A clustering approach for the-diversity model in privacy preserving data mining using fractional calculus-bacterial foraging optimization algorithm. Advances in Computer Engineering. doi:10.1155/2014/396529

Borsoi, R. A., Costa, G. H., & Bermudez, J. C. M. (2017). A new adaptive video SRR algorithm with improved robustness to innovations. *Proceedings of 25th European Signal Processing Conference (EUSIPCO)*. Academic Press. doi:10.23919/EUSIPCO.2017.8081460

Caballero, J., Ledig, C., Aitken, A., Acosta, A., Totz, J., Wang, Z., & Shi, W. (2017). *Real-Time Video Super-Resolution with Spatio-Temporal Networks and Motion Compensation*. *Proceedings of CVPR*, *1*(2), 7.

Chavda, H. K., & Dhamecha, M. (2017). Moving object tracking using PTZ camera in video surveillance system. *Proceedings of International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)* (pp. 263-266). Academic Press. doi:10.1109/ICECDS.2017.8389917

Deshmukh, A.B. & Usha Rani, N. (2017). Fractional-Grey Wolf optimizer-based kernel weighted regression model for multi-view face video super resolution. *International Journal of Machine Learning and Cybernetics*, 1–19.

Dong, W., Zhang, L., Shi, G., & Wu, X. (2011). Image Deblurring and Super-Resolution by Adaptive Sparse Domain Selection and Adaptive Regularization. *IEEE Transactions on Image Processing*, *20*(7), 1838–1857. doi:10.1109/TIP.2011.2108306 PMID:21278019

Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(2), 295–307. doi:10.1109/TPAMI.2015.2439281 PMID:26761735

Ge, J., Liu, J., Yuan, H., Ge, C., & Zhang, B. (2016). A spatiotemporal super-resolution algorithm for a hybrid stereo video system. *Signal, Image and Video Processing*, *10*(3), 559–566. doi:10.1007/s11760-015-0774-4

Huang, Y., Wang, W., & Wang, L. (2015). Bidirectional recurrent convolutional networks for multi-frame super-resolution. In Advances in Neural Information Processing Systems (pp. 235-243). Academic Press.

Izadpanahi, S., & Demirel, H. (2013). Motion block based video super resolution. *Digital Signal Processing*, *23*(5), 1451–1462. doi:10.1016/j.dsp.2013.04.002

Jian, M., & Lam, K. (2015). Simultaneous Hallucination and Recognition of Low-Resolution Faces Based on Singular Value Decomposition. *IEEE Transactions on Circuits and Systems for Video Technology*, *25*(11), 1761–1772. doi:10.1109/TCSVT.2015.2400772

Lai, W., Huang, J., Ahuja, N., & Yang, M. (2017). Deep laplacian pyramid networks for fast and accurate super resolution. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Press. doi:10.1109/CVPR.2017.618

Lee, D. Y., Lee, J., Choi, J., Kim, J., Kim, H. Y., & Choi, J. S. (2018). GPU-based real-time super-resolution system for high-quality UHD video up-conversion. *The Journal of Supercomputing*, *74*(1), 456–484. doi:10.1007/s11227-017-2136-1

Liu, D., Wang, Z., Fan, Y., Liu, X., Wang, Z., Chang, S., & Huang, T. (2017). Robust video super-resolution with learned temporal dynamics. *Proceedings of IEEE International Conference on computer vision* (pp. 2526-2534). IEEE Press.

Marquina, A., & Osher, S. J. (2008). Image super-resolution by TV-regularization and Bregman iteration. *Journal of Scientific Computing*, *37*(3), 367–382. doi:10.1007/s10915-008-9214-8

Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey Wolf Optimizer. *Advances in Engineering Software*, *69*, 46–61. doi:10.1016/j.advengsoft.2013.12.007

Panetta, Y., Zhou, S., & Jia, H. (2011). Nonlinear Unsharp Masking for Mammogram Enhancement. *IEEE Transactions on Information Technology in Biomedicine*, *15*(6), 918–928. doi:10.1109/TITB.2011.2164259 PMID:21843996

Porikli, F., Brémond, F., Dockstader, S. L., Ferryman, J., Hoogs, A., Lovell, B. C., & Venetianer, P. L. et al. (2013). Video surveillance: Past, present, and now the future. *IEEE Signal Processing Magazine*, *30*(3), 190–198. doi:10.1109/MSP.2013.2241312

Qu, S., Hu, R., Chen, S., Chen, L., & Zhang, M. (2014). Robust face super-resolution via position-patch neighborhood preserving. *Proceedings of IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. IEEE Press. doi:10.1109/ICMEW.2014.6890650

Ravi, C., & Khare, N. (2016). BGFS: Design and Development of Brain Genetic Fuzzy System for Data Classification. *Journal of Intelligent System*, *27*(2), 231–247. doi:10.1515/jisys-2016-0034

Ren, C., Dai, D., & Yan, H. (2012). Coupled Kernel Embedding for Low-Resolution Face Image Recognition. *IEEE Transactions on Image Processing*, *21*(8). PMID:22481822

Sajjadi, M. S. M., Vemulapalli, R., & Brown, M. (2018). Frame-Recurrent Video Super-Resolution. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 6626-6634). IEEE Press.

Savitha, C., & Ramesh, D. (2018). Motion detection in video surveillance: A systematic survey. *Proceedings of 2nd International Conference on Inventive Systems and Control (ICISC)* (pp. 51-54). Academic Press. doi:10.1109/ICISC.2018.8398880

Tu, F., Yin, S., Ouyang, P., Tang, S., Liu, L., & Wei, S. (2017). Deep Convolutional Neural Network Architecture with Reconfigurable Computation Patterns. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, *25*(8), 2220–2233. doi:10.1109/TVLSI.2017.2688340

Wang, M., Cheng, B., & Yuen, C. (2018). Joint Coding-Transmission Optimization for a Video Surveillance System with Multiple Cameras. *IEEE Transactions on Multimedia*, *20*(3), 620–633. doi:10.1109/TMM.2017.2748459

Yang, J., Wright, J., Huang, T., & Ma, Y. (2008, June). Image super-resolution as sparse representation of raw image patches. *Proceedings of the 2008 IEEE conference on computer vision and pattern recognition* (pp. 1-8). IEEE.

Yang, J., Wright, J., Huang, T. S., & Ma, Y. (2010). Image Super-Resolution Via Sparse Representation. *IEEE Transactions on Image Processing*, *19*(11). PMID:20483687

Yang, W., Feng, J., Xie, G., Liu, J., Guo, Z., & Yan, S. (2018). Video super-resolution based on spatial-temporal recurrent residual. *Computer Vision and Image Understanding*, *168*, 79–92. doi:10.1016/j.cviu.2017.09.002

Zeng, X., & Huang, H. (2012). Super-resolution method for multiview face recognition from a single image per person using nonlinear mappings on coherent features. *IEEE Signal Processing Letters*, *19*(4).

The UCSD face video database. (n.d.). taken from http://vision.ucsd.edu/datasets/leekc/disk2/VideoDatabase/testing/