
An Enhanced Video Inpainting Technique with Grey Wolf Optimization for Object Removal Application

B. Janardhana Rao^{1,*}, Y. Chakrapani² and S. Srinivas Kumar³

¹*ECE, CVR College of Engineering, Hyderabad, Telangana & JNTUK,
Kakinada, AP, India*

²*ECE, ACE Engineering College, Hyderabad, Telangana, India*

³*ECE, UCEK, JNTUK, Kakinada, Andhra Pradesh, India*

E-mail: janardhan.bitra@gmail.com

**Corresponding Author*

Received 19 July 2021; Accepted 10 September 2021;

Publication 11 January 2022

Abstract

Video inpainting is the most trending research topic from the last decade. Video inpainting is the process of restoring the damaged parts of the vintage video or the filling of the regions by removing the unwanted objects with sophisticated techniques. The video inpainting is achieved by dividing the video into frames and the motion of the moving objects in the frames are tracked by applying the motion tracking method. The existing inpainting method proposed by the Criminisi, neglected the local similarities in the images so it suffered from dropping effect in the priority computation. This paper proposed a new priority computation method by introducing gradient operation with the addition of curvature in the data term and local structure measurement function with structure tensor theory as an additional term. Later, the patch matching is achieved with the Sum of Absolute Difference (SAD) distance method. Further, the optimal patch is selected by applying the Grey Wolf Optimization (GWO) algorithm. The efficiency of the proposed video inpainting technique is evaluated with the performance metrics, viz., Peak Signal to Noise Ratio (PSNR), Structural Similarity (SSIM), and Edge

Journal of Mobile Multimedia, Vol. 18.3, 561–582.

doi: 10.13052/jmm1550-4646.1835

© 2022 River Publishers

Similarity (ESIM) executed in MATLAB. The PSNR and SSIM of the proposed method for *Fontaine_chatelet* video is improved by 18.9% and 4.19% than existing method. The proposed method is compared with other existing methods also and it outperformed the existing methods.

Keywords: Inpainting, curvature, structure tensor, grey wolf optimization, sum of absolute difference, PSNR, SSIM.

1 Introduction

Video Inpainting is a method of editing videos for repairing demolished videos due to aging. It is also a process of removing undesired objects in the video frames and filling them with sophisticated algorithms without any artifacts. The frames in the video sequence which contain undesired objects are called target frames; the hole is created after removing the object is called the target region. The frames which are having similar pixel information to fill the target region are called source frames. In order to produce the best video inpainting results, after aligning all the frames to the target frame, it is essential to select the suitable patches from the source region to inpaint the hole in the target frame. The exemplar-based inpainting methods are generally used for inpainting the hole in the target region with patch-by-patch processing.

Video inpainting finds a lot of applications in video processing such as video restoration, video stabilization, and films post-processing, etc. Image inpainting is the technique of restoring the damaged images and removing unwanted information in the image and filling the hole with appropriate data from the remaining region of the image.

The paper is organized as: Section 2 presents available methods of video inpainting as related work. The proposed video inpainting method is described in Section 3. The implementation and experimental results with performance comparison in terms of various metrics are discussed in Section 4. Finally, conclusions are given in Section 5.

2 Video Inpainting: Related Work

Video inpainting techniques are implemented from image inpainting techniques by considering the motion of objects in the videos. Image inpainting stands much attention in research works for decades. This is implemented for restoring damaged images, for text and scratch removal applications with

small target regions using Total Variation (TV), and diffusion-based methods [1–3]. The inpainting of large regions was achieved with enhanced exemplar-based inpainting methods [4]. The exemplar-based inpainting is first proposed by Criminisi et al. [5]. This is further modified as robust exemplar-based inpainting using region segmentation [6]. Yu et al. [7] proposed a contextual attention method in a generative framework to improve the image inpainting. The image inpainting is extended to achieve Video inpainting considering the spatial consistency and temporal coherence [8–11], these methods produced. These methods achieve great for video captured with static cameras with limitation of camera movement. Granados et al. [12], introduced Graph-cuts optimization to implement video inpainting. The homography-based image registration method was used here, to align the input video frames to the target frame. The missing pixels in the video frames are filled with the information taken from registered frames. The cost function minimization is utilized to find the best pixels values to fill in the target region [13, 14].

Newson et al. [15] improved the Granados inpainting by using the patch match algorithm [16] to the spatio-temporal domain with pyramids of frames. Ebdelli et al. [17], achieved reconstructed video by aligning a greater number of frames to the target frame using region-based homography transformation. Huang et al. [18], introduced a video completion technique by including both optical flow and color information of pixels in the target region of the frames in a video. In this paper, temporal consistency is maintained from both pixel-wise flow field and patch-based optimization. Le T.T. et al. [19] is proposed a novel semi-automatic removal of objects from complex background videos using optical flow-based pixel reconstruction inpainting. Okabe et al. [20] proposed an interactive video completion algorithm without any manual interaction. This algorithm estimates the temporally coherent color transition in the space-time holes and flow fields. In order to estimate the color transitions, the energy function was minimized by using L^2 data term. Wang et al. [21] proposed a new structure-guided deep video inpainting algorithm to improve the temporal and structural coherence for best video inpainting results using 3D convolution networks. In the first step of the method, the edges of the removed region were completed with an edge inpainting network with 3D convolution to illustrate scene structures and object shapes. Later the textures in the missing regions were filled by using a coarse to fine synthesis method with Structure Attention Module (SAM) considering the completed edges. The temporal coherence continuity was maintained with the help of motion flows among the adjacent frames during this entire process. The

method requires computationally speed hardware like Graphics Processing Unit (GPU).

Recently, a novel video inpainting algorithm [22] by introducing Recurrent Neural Network (RNN) and Cuckoo Search with Multi-Verse Optimization (CS-MVO). The transformation of one frame to the other frame in a video sequence is achieved by using motion tracking in the first step. The region of the frame is characterized into the structure and smooth with optimized RNN using the texture feature called Gray Level Co-occurrence Matrix (GLCM).

Criminisi et al. [5] proposed a first exemplar-based image inpainting method to restore the images with texture information and full structured along with filling the largely unknown region. This method avoids the local similarities in the images and hence, it consumes more time for inpainting and suffered from dropping effect. The dropping effect means, the rapid decrease of the confidence term value at a smaller number of iterations in the process, due to which, the priority function value drops to a very low value for a smaller number of iterations [23]. Wang et al. [23] proposed a new priority calculation method introducing regularization factor to avoid dropping effect in exemplar-based inpainting method. A similar patch search in the source region was accomplished by combining SSD distance with Normalized Cross-Correlation (NCC). This has disadvantages such as the inability to inpaint images with random textures. Liu et al. [24] presented a local structure measurement function by using structure tensor theory and optimizes the patch priority. This method has disadvantages such as the inability to restore images with different structure information and not able to overcome the dropping effect. Xiang et al. [25] proposed a novel exemplar-based inpainting method by adopting changes in the data term function in the priority calculation process. Patel et al. [26] achieved exemplar-based inpainting with a reduced search region in the source region of the image. This method produced good inpainting results for images having objects with a linear geometry and produced artifacts to inpaint curved geometry. Deng et al. [27] incorporated a new priority calculation method and a little extension on similar patch searching also. This inpainting process well recovers geometric regions and failed to recover curved or cross-shaped structures.

From the literature review, the existing methods of inpainting had limitations in finding the proper inpainting location, i.e., priority computation and similar patch search from the source region to inpaint images with random structures and curved geometry. The appropriate selection of inpainting location and matching patches to inpaint the target region playing a major role

to produce the best video inpainting results. This motivated to implement a new method for finding the proper inpaint location by introducing gradient operation with the addition of curvature in data term and local structure measurement function with structure tensor theory as an additional term. The best patch selection in the source region with Sum of Absolute Difference by applying optimization using the Grey Wolf Optimization (GWO) [28] algorithm.

3 Proposed Framework for Video Inpainting

The video inpainting techniques for various videos are attained by converting the given video into a number of frames. Later the moving objects for static background and also for moving background are estimated by using the process of motion tracking. The inpainting location in the converted frames is identified with the novel priority calculation method by presenting structure tensor theory and gradient operation with the addition of curvature.

The structure tensor theory is used to choose the local structure measurement function. The gradient operation with the addition of curvature is added in the data term while calculating the priority function in order to deal with frames containing the curved geometry. After identifying the inpainting location, the best patch in the source region to inpaint the target region is identified using Sum of Absolute Difference (SAD). Furthermore, the Grey Wolf Optimization technique is employed to identify the optimal patch in the source region.

3.1 Process of Motion Tracking

The location of the moving objects in a video frames is obtained using method called motion tracking. The objective of this is to track the movement of the moving objects. The moving tracking in a video sequence is accomplished in different steps, which are filtering, segmentation, and tracking process.

- Filters are used to remove the noise from marked frames
- Segmentation of background and objects in the frames
- The object tracking is achieved by the variation of intensity values

Later, the frames are marked to track the marked objects in each frame for removing the objects. The neighborhood frames are utilized to get the respective pixels when the background is in motion. The masked object is removed by detecting the pixels of the tracked moving objects.

3.2 Computation of Patch Priority

The frames of the videos which contain the unwanted object to be removed are divided into two regions. The region which contains the hole after removing the unwanted object is called the target region (ψ) and the remaining information in the frame is called the source region (Φ), which are represented in Figure 1.

The patch priority on the boundary of the target region to inpaint is computed by the following equation,

$$F_p = a * C_p + b * D_p + c * G_p \tag{1}$$

where, a , b , and c indicate the constants related to texture features, structure information, and linear structure information in the target region. The corresponding constant is chosen as a high value based on the information available in the source region.

The term C_p is called the confidence term, indicates the amount of reliable information available nearer to the pixel p . It was computed by using the equation as,

$$C_p = \frac{\sum_{t \in (\Psi_p \cap \Phi)} C(t)}{|\Psi_p|} \tag{2}$$

where, t indicates the coordinates of the pixels of Ψ_p and Φ . $|\Psi_p|$ is the number of pixels in the target patch.

The data term $D(p)$, represents the number of isophotes hitting the boundary at pixel p is taken as,

$$D_p = \frac{|\nabla I_p^\perp \cdot n_P| / L_p + d}{255} \tag{3}$$

where, L_p gives the geometric information of the isophote, called the curvature in the direction of the isophote. This curvature is introduced to propagate

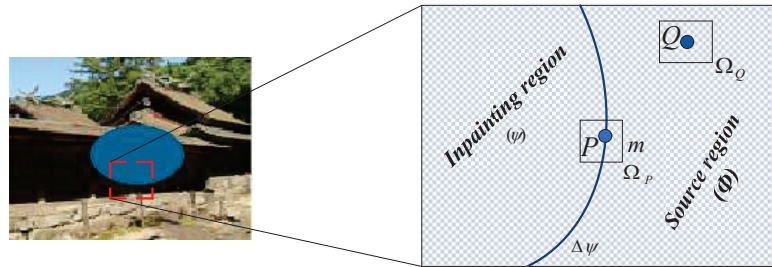


Figure 1 Representation of inpainting location and the source region.

the information in the direction of the isophote. The curvature is taken as,

$$L_p = \nabla \cdot \left(\frac{\nabla I_p}{|\nabla I_p|} \right) \quad (4)$$

The gradient (∇I_p) of the color image at pixel p is computed by using the gradient of the gray component. Here d is the constant value chosen as 3. The data term computed with isophotes did not define the image structures properly. The eigenvalues computed from the structure tensor matrix well describe the local structure information. This motivated to introduce the structure tensor matrix to obtain the inpaint location.

The term, G_p indicates the local structure of the measurement function. Which is computed using structure tensor theory. The function is expressed as

$$G_p = M * g + \exp(-g) \quad (5)$$

where, M is the constant and g is called the local measure function, calculated using eigenvalues (λ_1 and λ_2) obtained from the coefficients taken from the structure tensor matrix.

The structure tensor matrix is taken as

$$g = |\lambda_1 - \lambda_2| \quad (6)$$

The structure tensor matrix from the structure tensor theory is represented as,

$$K_\rho = \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix} = G_\rho * \begin{bmatrix} \frac{\partial^2 I}{\partial^2 x} & \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} \\ \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} & \frac{\partial^2 I}{\partial^2 y} \end{bmatrix} \quad (7)$$

where, G_ρ is the Gaussian function with standard deviation (ρ). The horizontal and vertical components of the gradient vector at each pixel of the image (I) are represented with $\frac{\partial I}{\partial x}$ and $\frac{\partial I}{\partial y}$.

The eigenvalues calculated from the coefficients of the matrix K_ρ with the following equation,

$$\lambda_1 = \frac{1}{2} \left(k_{11} + k_{22} + \sqrt{(k_{11} + k_{22})^2 + 4k_{12}^2} \right) \quad (8)$$

and

$$\lambda_2 = \frac{1}{2} \left(k_{11} + k_{22} - \sqrt{(k_{11} + k_{22})^2 + 4k_{12}^2} \right) \quad (9)$$

3.3 Patch Matching

The suitable patch to inpaint the obtained highest priority patch is selected with optimal patch condition taken as

$$\Psi_{q'} = \arg \min_{\Psi_{q^i} \in \Phi} d_{SAD}(\Psi_p, \Psi_{q^i}) \quad (10)$$

where, $d_{SAD}(\Psi_p, \Psi_q)$ is the distance using SAD, which is calculated as,

$$\begin{aligned} d_{SAD}(\Psi_p, \Psi_q) &= d_{SSE}(\Psi_p, \Psi_q) \\ &= \sum |(R_{\Psi_p} - R_{\Psi_q}) + (G_{\Psi_p} - G_{\Psi_q}) + (B_{\Psi_p} - B_{\Psi_q})| \end{aligned} \quad (11)$$

Further, the optimal patch is selected by applying the Grey Wolf Optimization (GWO) technique taking the SAD values between the patches as a parameter.

3.4 Grey Wolf Optimization for Patch Matching

The Grey wolf optimization is initially proposed by Mirjalili and Lewis [28] based on the nature of social order and grey wolves hunting techniques. The best parameter to copy the solicitation is represented by alpha (α), beta (β) indicates the best parameter to copy the solicitation of Wolves, and delta (δ) is 2nd and 3rd best game plans. The omega (ψ) wolves that are left of the competitor game plans. The steps for GWO [29] are given as follows,

1. Initialize algorithm parameters:
Intensity threshold and orientation thresholds lower bound (LB) & upper bound vectors (UB), number of variables $Nva = 2$, Maxi iteration ($MaxItr$).
2. Initialize α, β, δ size $1 \times Nva$.
3. Initialize the positions Ps_{ij}
4. Find the fitness of the initial position
5. Divide the solution with respect to fitness.
Take the fitness attribute into consideration and note the first three best fitnesses as α, β , and δ .
6. Define 'v' vector that decreases with iteration number $= 2 - (Itr / MaxItr)$
7. Perform Update Position
Update α position

Define random numbers between 0 to 1 for $r1$ and $r2$.
 Compute self-assertive parameters for hyper circles $A1, C1$ as
 $A1 = 2 * a * r1 - a$ and $C1 = 2 * r2$; with $r1, r2$ irregular vectors
 in $[0, 1]$.

$$D_{\alpha} = C1 * \alpha_{Ps}(j) - Ps_{ij}$$

$$X1_{\alpha} = \alpha_{Ps}(j) - A1 * D_{\alpha}$$

Update β position

Define random numbers between 0 to 1 for $r1$ and $r2$.
 Compute $A2, C2$ as $A2 = 2 * a * r1 - a$ and $C2 = 2 * r2$.

$$D_{\beta} = C2 * \beta_{Ps}(j) - Ps_{ij}$$

$$X2_{\beta} = \beta_{Ps}(j) - A2 * D_{\beta}$$

Update δ position

Define random numbers between 0 to 1 for $r1$ and $r2$.
 Compute $A3, C3$ as $A3 = 2 * a * r1 - a$ and $C3 = 2 * r2$

$$D_{\delta} = abs(C3 * \delta_{Ps}(j) - Ps_{ij})$$

$$X3_{\delta} = \delta_{Ps}(j) - A3 * D_{\delta}$$

$$Pos_{ij} = (X1_{\alpha} + X2_{\beta} + X3_{\delta})/3$$

8. Check the boundaries of the variables.

$$\text{If } (Ps_{ij} > UB), Pos_{ij} = UB$$

or

$$\text{If } (Ps_{ij} < LB), Ps_{ij} = LB$$

9. Repeat steps 4 to 7 for population size and the number of iterations.

The alpha (α) is superior in planning the hunt, beta (β), and delta (δ) update the position of the prey to represent the pursuit of the dark wolves numerically. Record the three best acts and required omegas (ψ) to recompute positions with its boss's best positions, respectively, for the chase as indicated in step 6.

4 Implementation and Experimental Results

In this section, the proposed video inpainting algorithm is implemented on videos, two videos taken from the YouTube-VOS dataset [30] and two videos from the DAVIS dataset [31]. There are four test videos considered in this work, test video 1-Fontaine_chatelet, test video 2-Les_loulous, test video 3-tennis, and test video 4-horse jump high. In the proposed implementation, the optimized patch selection carries out a significant role in the inpainting process.

The performance metrics viz., PSNR, SSIM [32], and Edge Similarity index are utilized for analyzing the performance of the proposed system. The proposed Grey Wolf Optimization (GWO) and Enhanced Patch Priority (EPF) based inpainting is compared with the inpainting method through various optimization algorithms viz., Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Artificial Bee Colony (ABC) Optimization in terms of performance metrics. In addition, compared with existing video inpainting techniques in the literature.

4.1 Performance Comparison in Terms of PSNR

The PSNR values for the video inpainting method using the optimization algorithms viz., GA, PSO, ABC, and the proposed method (GWO) are tabulated in Table 1.

The comparative graphical representation of PSNR for the video inpainting method with various optimization algorithms is shown in Figure 2.

The PSNR values of the proposed video inpainting technique for the Fontaine_chatelet video is 18.5%, 13.5%, and 9.5% efficient compared to GA, PSO, and ABC. For Les_loulous video is 14.4%, 4.8%, and 3.16% superior

Table 1 PSNR values for video inpainting method with various optimization algorithms

| Test Videos | GA | PSO | ABC | Proposed (GWO) |
|-------------------|--------|--------|--------|----------------|
| Fontaine_chatelet | 17.876 | 18.971 | 19.845 | 21.942 |
| Les_loulous | 15.134 | 16.826 | 17.12 | 17.679 |
| Tennis | 21.453 | 21.654 | 21.712 | 22.756 |
| Horse jump high | 19.214 | 19.453 | 19.725 | 20.647 |

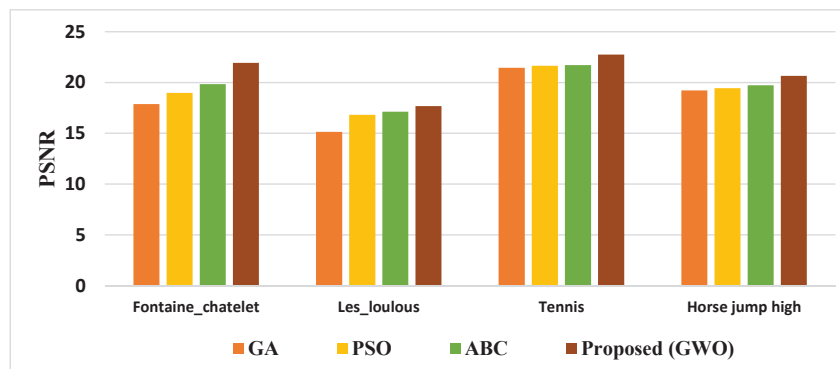


Figure 2 Comparison with various Optimization algorithms in terms of PSNR.

Table 2 PSNR values of Existing Video inpainting methods in the literature

| Test Videos | Newson et al. [15] | Huang et al. [18] | Le et al. [19] | Proposed Method |
|-------------------|--------------------|-------------------|----------------|-----------------|
| Fontaine_chatelet | 17.79 | 21.64 | 21.56 | 21.942 |
| Les_loulous | 15.96 | 16.12 | 16.27 | 17.679 |
| Tennis | 21.5946 | 21.652 | 20.3179 | 22.756 |
| Horse jump high | 19.9035 | 19.8662 | 19.6732 | 20.647 |

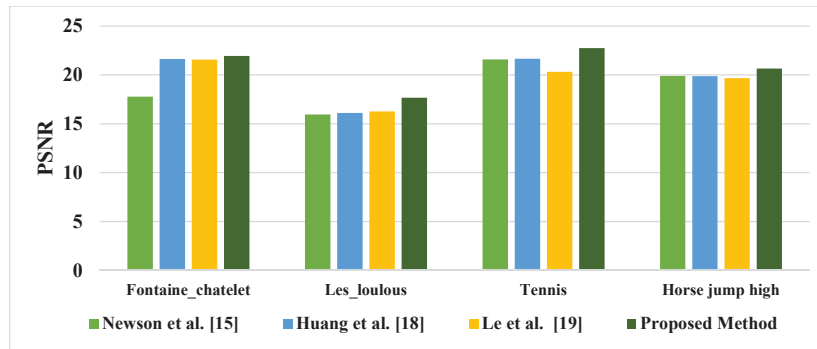


Figure 3 Comparison with existing methods of video inpainting in terms of PSNR.

compared to GA, PSO, and ABC. For tennis video is 5.7%, 4.8%, and 4.6% efficient compared to GA, PSO, and ABC. For horse jump high video 6.94%, 5.78%, and 4.46% superior compared to GA, PSO, and ABC.

From the graphical representation, it is confirmed that the proposed method with GWO optimization has the highest PSNR values compared to the implementation with various optimization algorithms viz., GA, PSO, and ABC.

The PSNR values of the existing methods of video inpainting in the literature viz., Newson et al. [15], Huang et al. [18], and Le et al. [19] along with the proposed method are listed in Table 2.

The PSNR values of the proposed method for Fontaine_chatelet video is increased by 18.9%, 1.3%, and 1.7% than Newson et al. [15], Huang et al. [18], and Le et al. [19]. For Les_loulous 9.72%, 8.81%, and 7.96% than Newson et al. [15], Huang et al. [18], and Le et al. [19]. For tennis video 5.1%, 4.85%, and 10.71% than Newson et al. [15], Huang et al. [18], and Le et al. [19]. For horse jump high 3.6%, 3.78%, and 4.72% than Newson et al. [15], Huang et al. [18] and Le et al. [19].

The comparative graphical representation of PSNR for the existing video inpainting method in the literature with the proposed method using GWO is shown in Figure 3.

From the comparative graphical representation, the proposed video inpainting technique using GWO is outperformed in terms of PSNR values compared to available methods of video inpainting in the literature viz., Newson et al. [15], Huang et al. [18], and Le et al. [19].

4.2 Performance Comparison in Terms of SSIM

The SSIM values for the video inpainting method using the optimization algorithms viz., GA, PSO, ABC, and the proposed method (GWO) are tabulated in Table 3.

The SSIM values of the proposed method for the Fontaine_chatelet test video are 2.14% efficient than GA, 1.41% than PSO, and 1.16% than ABC. For Les_loulous test video, 2.14% superior than GA, 0.96% than PSO, and 0.31% than ABC. For tennis test video, 2.66% more than GA, 2.38% than PSO, and 1.94% than ABC. For horse jump high test video, 11.5% superior than GA, 11.09% than PSO, 3.94% than ABC.

The comparative graphical representation of SSIM for the video inpainting method with various optimization algorithms is shown in Figure 4. From the graphical representation, it is confirmed that the proposed method with GWO optimization has superior SSIM values compared to the implementation with various optimization algorithms viz., GA, PSO, and ABC.

Table 3 SSIM values for video inpainting method with various optimization algorithms

| Test Videos | GA | PSO | ABC | Proposed (GWO) |
|-------------------|--------|--------|--------|----------------|
| Fontaine_chatelet | 0.9352 | 0.9413 | 0.9481 | 0.9598 |
| Les_loulous | 0.9225 | 0.9383 | 0.9416 | 0.9489 |
| Tennis | 0.9385 | 0.9412 | 0.9454 | 0.9642 |
| Horse jump high | 0.8421 | 0.8465 | 0.9145 | 0.9521 |

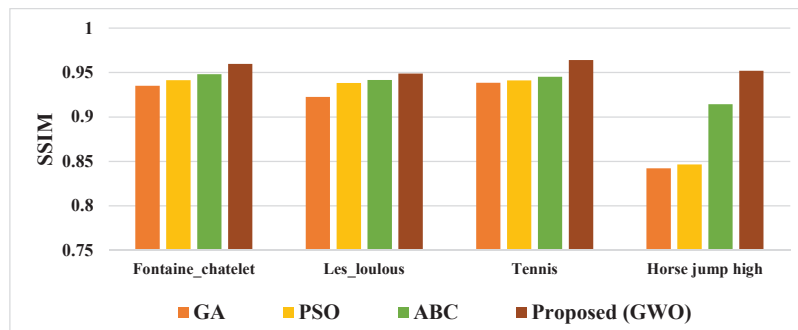
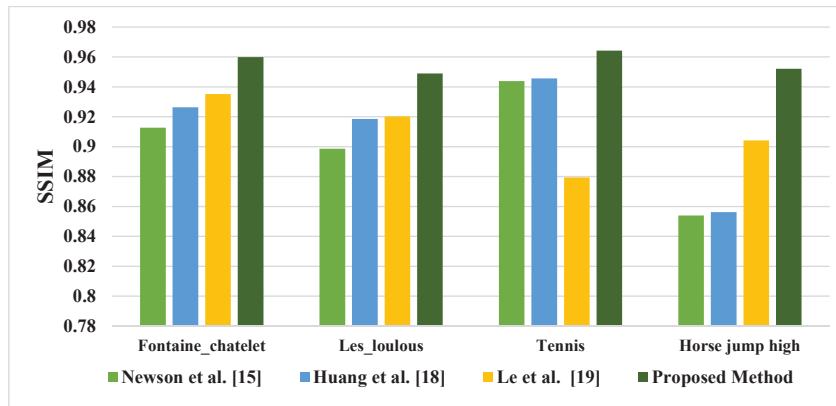


Figure 4 Comparison with various Optimization algorithms in terms of SSIM.

Table 4 SSIM values of Existing Video inpainting methods in the literature

| Test Videos | Newson et al. [15] | Huang et al. [18] | Le et al. [19] | Proposed Method |
|-------------------|--------------------|-------------------|----------------|-----------------|
| Fontaine_chatelet | 0.9127 | 0.9263 | 0.93514 | 0.9598 |
| Les_loulous | 0.8986 | 0.9185 | 0.92023 | 0.9489 |
| Tennis | 0.94385 | 0.94571 | 0.8795 | 0.9642 |
| Horse jump high | 0.8539 | 0.8562 | 0.90408 | 0.9521 |

**Figure 5** Comparison with existing methods of video inpainting in terms of SSIM.

The SSIM values of the existing methods of video inpainting in the literature viz., Newson et al. [15], Huang et al. [18], and Le et al. [19] along with the proposed method are listed in Table 4.

The SSIM values of the proposed method for the Fontaine_chatelet test video is 4.19% efficient than Newson et al. [15], 2.77% than Huang et al. [18], and 1.84% than Le et al. [19]. For Les_loulous test video, 4.56% superior than Newson et al. [15], 2.45% than Huang et al. [18], and 2.26% than Le et al. [19]. For tennis test video, 2.11% more than Newson et al. [15], 1.91% than Huang et al. [18], and 8.78% than Le et al. [19]. For horse jump high test video, 1.03% superior than Newson et al. [15], 1% than Huang et al. [18], 5.04% than Le et al. [19].

The comparative graphical representation of SSIM for the existing video inpainting method in the literature with the proposed method using GWO is shown in Figure 5. From the comparative graphical representation, the proposed video inpainting technique using GWO has produced the best SSIM values compared to available methods of video inpainting in the literature viz., Newson et al. [58], Huang et al. [54] and Le et al. [19].

4.3 Performance Comparison in Terms of Edge Similarity

The Edge Similarity values for the video inpainting method using the optimization algorithms viz., GA, PSO, ABC, and the proposed method (GWO) are tabulated in Table 5.

The Edge Similarity index values of the proposed method for the Fontaine_chatelet test video is 2.67% efficient than GA, 1.12% than PSO, and 0.66% than ABC. For Les_loulous test video, 2.61% superior than GA, 1.81% than PSO, and 0.03% than ABC. For tennis test video, 1.52% more than GA, 0.88% than PSO, and 0.031% than ABC. For horse jump high test video, 3.62% superior than GA, 2.57% than PSO, 2.33% than ABC.

The comparative graphical representation of Edge Similarity index values for the video inpainting method with various optimization algorithms is shown in Figure 6.

From the graphical representation, it is confirmed that the proposed method with GWO optimization has the highest Edge Similarity index values compared to the implementation with various optimization algorithms viz., GA, PSO, and ABC.

Table 5 Edge similarity values for video inpainting method with various optimization algorithms

| Test Videos | GA | PSO | ABC | Proposed (GWO) |
|-------------------|--------|--------|--------|----------------|
| Fontaine_chatelet | 102.17 | 103.8 | 104.28 | 104.98 |
| Les_loulous | 131.69 | 132.78 | 134.82 | 135.23 |
| Tennis | 135.25 | 136.12 | 136.91 | 137.34 |
| Horse jump high | 101.31 | 102.41 | 102.67 | 105.12 |

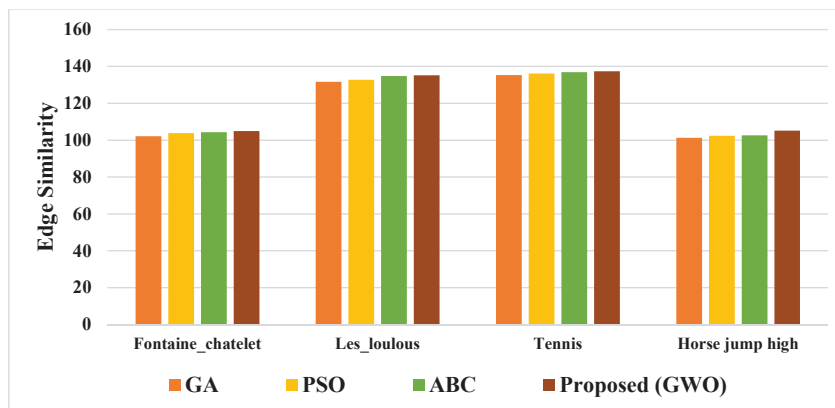


Figure 6 Comparison with various optimization algorithms in terms of edge similarity.

Table 6 Edge Similarity values of Existing Video inpainting methods in the literature

| Test Videos | Newson et al. [15] | Huang et al. [18] | Le et al. [19] | Proposed Method |
|-------------------|--------------------|-------------------|----------------|-----------------|
| Fontaine_chatelet | 102.15 | 103.43 | 104.23 | 104.98 |
| Les_loulous | 125.46 | 128.54 | 131.27 | 135.23 |
| Tennis | 133.14 | 134.37 | 135.86 | 137.34 |
| Horse jump high | 100.42 | 101.85 | 103.53 | 105.12 |

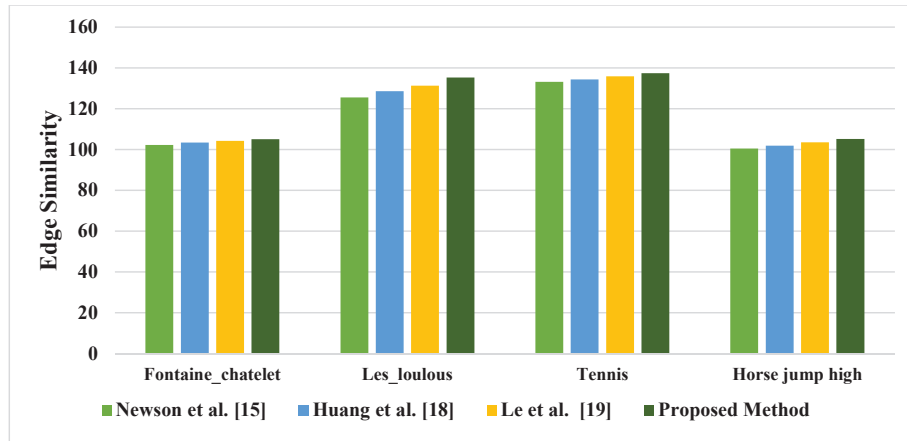


Figure 7 Comparison with existing methods of video inpainting in terms of edge similarity.

The Edge Similarity index values of the existing methods of video inpainting in the literature viz., Newson et al. [15], Huang et al. [18], and Le et al. [19]. along with the proposed method are listed in Table 6.

The Edge Similarity index values of the proposed method for the Fontaine_chatelet test video is 2.69% efficient than Newson et al. [15], 1.47% than Huang et al. [18], and 0.71% than Le et al. [19]. For Les_loulous test video, 7.22% superior than Newson et al. [15], 4.94% than Huang et al. [18], and 2.92% than Le et al. [19]. For tennis test video, 3.05% more than Newson et al. [15], 2.16% than Huang et al. [18], and 1.07% than Le et al. [19]. For horse jump high test video, 4.47% superior than Newson et al. [15], 3.11% than Huang et al. [18], 1.51% than Le et al. [19].

The comparative graphical representation of PSNR for the existing video inpainting method in the literature with the proposed method using GWO is shown in Figure 7.

From the comparative graphical representation, the proposed video inpainting technique using GWO is outperformed in terms of PSNR values

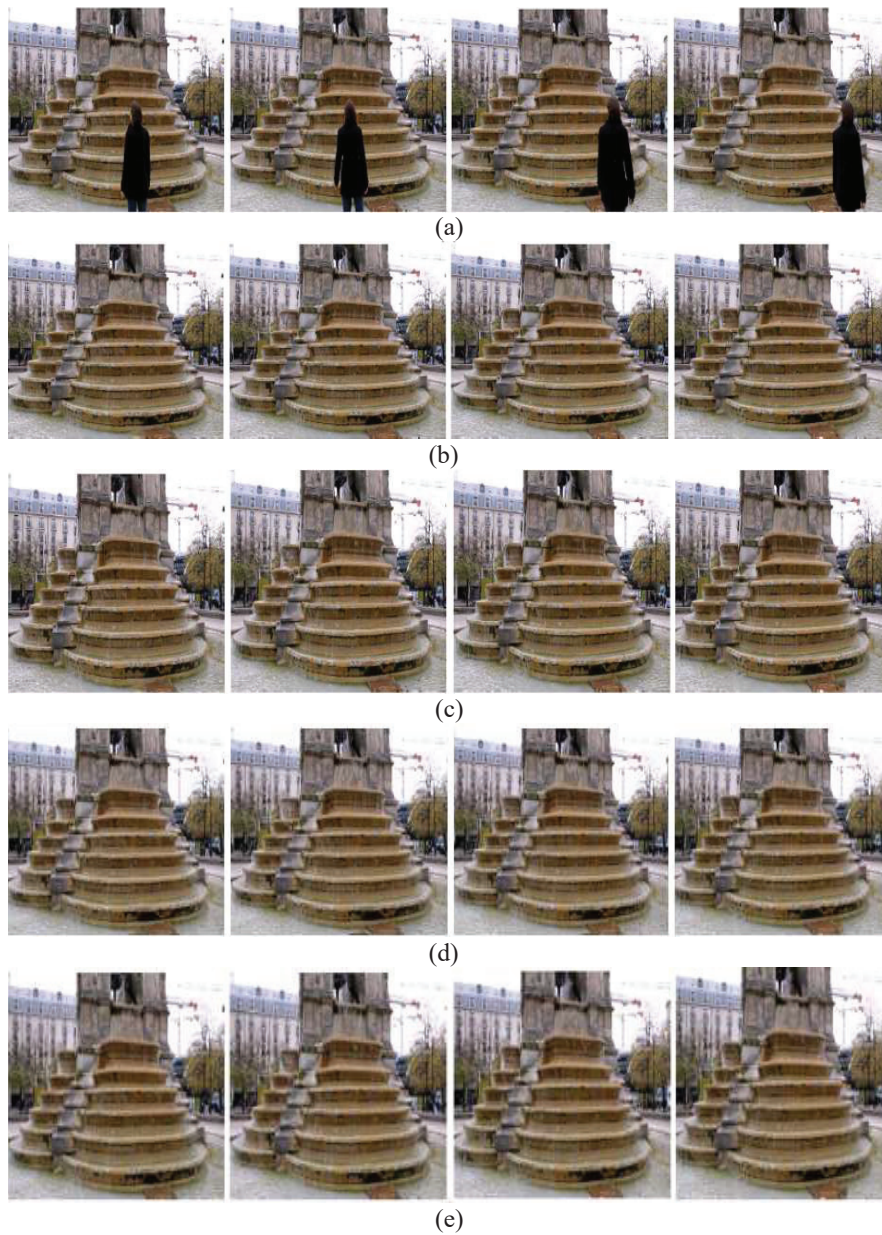


Figure 8 Comparison of results with [15, 18] and [19] (a) Input video frames with object to be removed (b) Results from [15] (c) Results from [18] (d) Results from [19] (e) Results from proposed method.

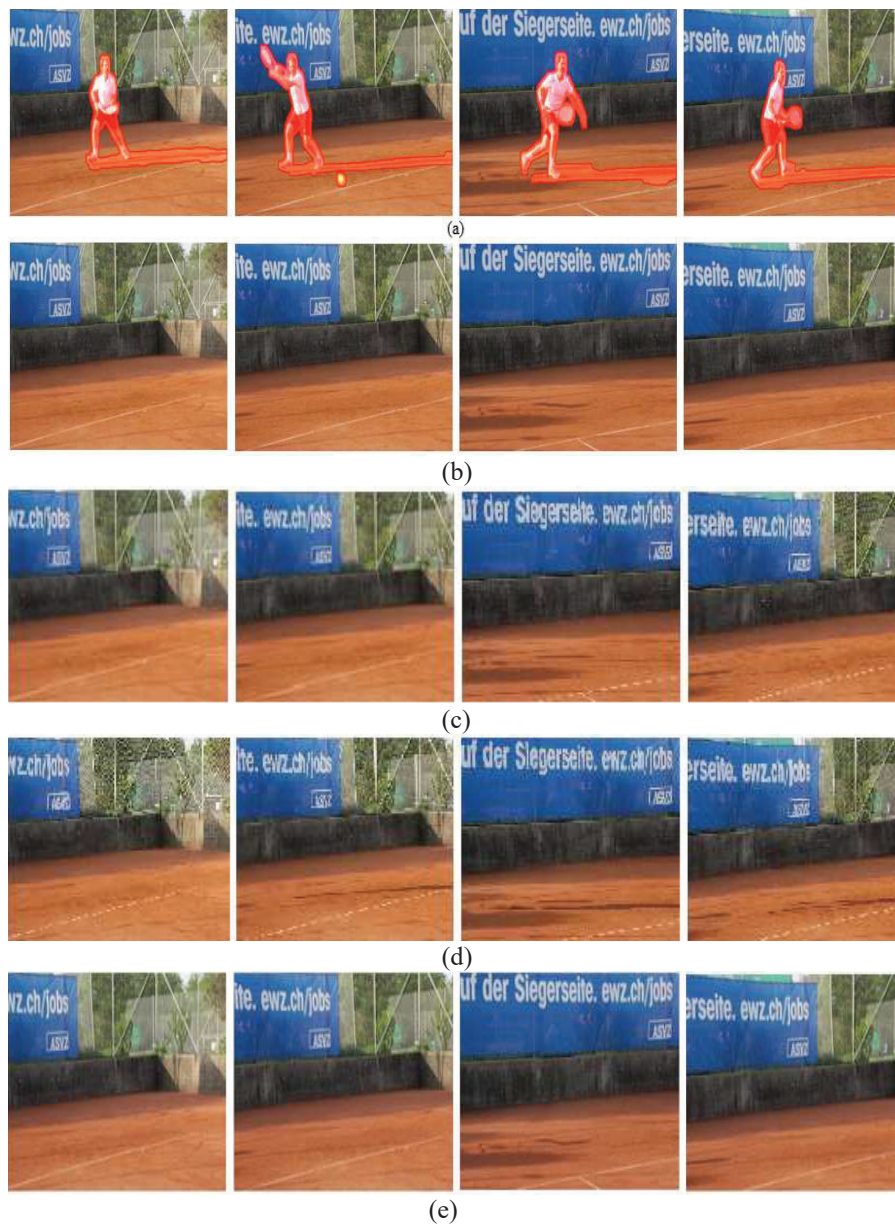


Figure 9 Comparison of results with [15, 18] and [19] (a) Input video frames with object to be removed (b) Results from [15] (c) Results from [18] (d) Results from [19] (e) Results from proposed method.

to the available methods of video inpainting in the literature viz., Newson et al. [15], Huang et al. [18] and Le et al. [19].

The frames of the video sequences with an object to be removed and inpainted results of Newson et al. [15], Huang et al. [18] and Le et al. [19], and the proposed GWO method are presented here. The results of the test video1 are shown in Figure 8, the results of the test video2 are shown in Figure 9. From these, it is noted that results obtained from the proposed method inpainted the target region more efficiently in comparison with existing methods.

5 Conclusions

Video inpainting is the process of restoring the damaged parts of the vintage video or the filling of the regions by removing the unwanted objects with sophisticated techniques. In this work, proposed a new priority calculation method to detect the inpaint location. The gradient operation with an addition of curvature in the data term is considered. The additional term, local structure measurement function with structure tensor theory is incorporated in the priority computation. The Grey Wolf Optimization is applied for the selection of similarity patches in the source region with SAD as a distance metric. The efficacy of the proposed method is confirmed in terms of metrics, viz., PSNR, SSIM, and ESIM. The values of the performance metrics proved that the proposed method is superior compared to the existing methods in the literature.

References

- [1] G. Sridevi, S.S. Kumar, “Image inpainting and enhancement using fractional order variational model”, *Defence Sci. J.* 67(3), 308–315 (2017).
- [2] G. Sridevi, S.S. Kumar, “Image inpainting based on fractional-order nonlinear diffusion for image reconstruction”, *Circuits, Systems, and Signal Processing* (2019): 1–16.
- [3] G. Sridevi, and S. Srinivas Kumar. “p-Laplace Variational Image Inpainting Model Using Riesz Fractional Differential Filter.” *International Journal of Electrical and Computer Engineering* 7.2 (2017): 850.

- [4] Janardhana Rao, B., Y. Chakrapani, and S. Srinivas Kumar. "Image Inpainting Method with Improved Patch Priority and Patch Selection." *IETE Journal of Education* 59.1 (2018): 26–34.
- [5] A. Criminisi, P. Perez, K. Toyama, "Region filling and object removal by exemplar-based image inpainting", *IEEE Trans. Image Process.* 13(9), 1200–1212 (2004).
- [6] Lee, Jino, Dong-Kyu Lee, and Rae-Hong Park. "Robust exemplar-based inpainting algorithm using region segmentation." *IEEE Transactions on Consumer Electronics* 58, no. 2 (2012): 553–561.
- [7] Yu, Jiahui, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S. Huang. "Generative image inpainting with contextual attention." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5505–5514. 2018.
- [8] Y. Matsushita, E. Ofek, W. Ge, X. Tang, and H.-Y. Shum, "Full-frame video stabilization with motion inpainting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 7, pp. 1150–1163, Jul. 2006.
- [9] K. A. Patwardhan, G. Sapiro, and M. Bertalmio, "Video inpainting under constrained camera motion," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 545–553, Feb. 2007.
- [10] T. K. Shih, N. C. Tang, and J.-N. Hwang, "Exemplar-based video inpainting without ghost shadow artifacts by maintaining temporal continuity," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 3, pp. 347–360, Mar. 2009.
- [11] T. K. Shih, N. C. Tan, J. C. Tsai, and H.-Y. Zhong, "Video falsifying by motion interpolation and inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
- [12] M. Granados, J. Tompkin, K. I. Kim, J. Kautz, and C. Theobalt, "Background inpainting for videos with dynamic objects and a free moving camera," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 682–695.
- [13] O. Whyte, J. Sivic, and A. Zisserman, "Get out of my picture! Internet based inpainting," in *Proc. Brit. Mach. Vis. Conf.*, 2009, pp. 1–11.
- [14] X. Chen, Y. Shen, and Y. H. Yang, "Background estimation using graph cuts and inpainting," in *Proc. Graph. Inter.*, 2010, pp. 97–103.
- [15] A. Newson, A. Almansa, M. Fradet, Y. Gousseau, and P. Pérez, "Video inpainting of complex scenes," *SIAM J. Imag. Sci.*, vol. 7, no. 4, pp. 1993–2019, 2014.
- [16] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "Patch Match: A randomized correspondence algorithm for structural image editing," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 24:1–24:11, Jul. 2009.

- [17] M. Ebdelli, O. Le Meur, and C. Guillemot, "Video inpainting with short term windows: application to object removal and error concealment," *IEEE Trans. Image Processing*, vol. 24, no. 10, pp. 3034–3047, May 2015.
- [18] Huang, J.B., Kang, S.B., Ahuja, N. and Kopf, J., 2016. Temporally coherent completion of dynamic video. *ACM Transactions on Graphics (TOG)*, 35(6), pp. 1–11.
- [19] Le, Thuc Trinh, et al. "Object removal from complex videos using a few annotations", *Computational Visual Media* 5.3 (2019): 267–291.
- [20] M. Okabe, K. Noda, Y. Dobashi and K. Anjyo, "Interactive video completion," *IEEE computer graphics and applications*, 40(1), pp. 127–139, 2019.
- [21] C. Wang, X. Chen, S. Min, Z. J. Zha, and J. Wang, "Structure-Guided Deep Video Inpainting," *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.
- [22] B Janardhana Rao, Y Chakrapani, S Srinivas Kumar, Hybridized Cuckoo Search with Multi-Verse Optimization-Based Patch Matching and Deep Learning Concept for Enhancing Video Inpainting, *The Computer Journal*, 2021; bxab067, <https://doi.org/10.1093/comjnl/bxab067>.
- [23] J. Wang, K. Lu, D. Pan, N He, and B. Bao, "Robust object removal with an exemplar-based image inpainting approach," *Neurocomputing*, pp. 150–155, 2014.
- [24] Y. Liu, C. J. Liu, H. L. Zou, S.S Zhou, Q. Shen, and T.T. Chen, "A novel exemplar-based image inpainting algorithm," In *2015 International Conference on Intelligent Networking and Collaborative Systems*, *IEEE* September 2015, pp. 86–90.
- [25] Xiang, C., Cao, Y., Duan, P., & Shi, L. (2014, August). An improved exemplar-based image inpainting algorithm. In *2014 9th International Conference on Computer Science & Education* (pp. 770–775). *IEEE*.
- [26] J. Patel and T. K. Sarode, "Exemplar based image inpainting with reduced search region," *International Journal of Computer Applications*, 92(12), 2014.
- [27] L. J. Deng, T. Z. Huang and X. L. Zhao, "Exemplar-based image inpainting using a modified priority definition," *PloS one*, 10(10), e0141199, 2015.
- [28] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer," *Adv Eng Softw*, 69, pp. 46–61, 2014.

- [29] Khwaja Muinuddin Chisti Mohammed, Srinivas Kumar S., Prasad Gandikota, “Segmenting the Defective Leather Texture Image Using Modified Region Growing With GWO Technique,” *Journal of Adv Research in Dynamical & Control Systems*, Vol. 10, no. 04, Oct. 4, 2018.
- [30] N. Xu, L. Yang, Y. Fan, J. Yang, D. Yue, Y. Liang, and T. Huang, “Youtube-vos: Sequence-to-sequence video object segmentation,” In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 585–601.
- [31] Perazzi, Federico, Jordi Pont-Tuset, Brian McWilliams, Luc Van Gool, Markus Gross, and Alexander Sorkine-Hornung. “A benchmark dataset and evaluation methodology for video object segmentation.” In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 724–732. 2016.
- [32] Odugu VK, C VN, K SP. An efficient VLSI architecture of 2-D finite impulse response filter using enhanced approximate compressor circuits. *Int J Circ Theor Appl*. 2021;1–16. doi:10.1002/cta.3114

Biographies



B. Janardhana Rao received his B.Tech from Nagarjuna University and M.Tech from Nagarjuna University. Presently he is pursuing his Ph.D at JNT University, Kakinada. Currently working as an Associate Professor in the Department of ECE, CVR College of Engineering, Hyderabad, Telangana. His areas of interest include Image Inpainting, Video Inpainting, Image restoration, and enhancement.



Y. Chakrapani received his B.Tech in ECE from JNT University Anantapur. He completed his Master's Degree from NIT Warangal. He completed his Ph.D in Image Processing from JNTU Anantapur. He has 30 years of teaching experience. He worked as a professor and HOD of ECE in G. Pulla Reddy Engineering College, Kurnool; presently he is working as a professor of ECE in ACE Engineering College, Hyderabad. His research interest includes Image Processing and Video Signal Processing.



S. Srinivas Kumar is working as a Professor in ECE Department, UCEK, JNTUK, Kakinada, Andhra Pradesh, India. He received his M.Tech from JNTU, Hyderabad, India. He received his Ph.D from E&ECE Department, IIT, Kharagpur. He has 35 years of experience in teaching and research. He has published more than 125 research papers in National and International Journals, and also in proceedings of reputed conferences. His research interests are Digital Image Processing, Computer Vision, and the application of Artificial Neural Networks and Fuzzy logic to engineering problems.