Fuzzy Optim. Model. J. 2026; 7(1) : 43-55

Fuzzy Optimization and Modeling Journal (FOMJ)

https://doi.org/10.57647/fomj.2026.0701.03

Original Research

Contour-Based Video Inpainting using Neutrosophic Sets

Amanna Ghanbari Talouki[1,*], Abbas Koochari[2], Seyed Ahmad Edalatpanah[3]

[1]*Department of Technical and Engineering, Ayandegan University, Tonekabon, Iran*
[2]*Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran*
[3]*Department of Applied Mathematics, Ayandegan University, Tonekabon, Iran*

[*]**Corresponding author:** ghanbari@aihe.ac.ir

Article History

Received:
5 January 2026
Revised:
25 February 2026
Accepted:
27 March 2026
Published in Issue:
30 March 2026

Abstract:

Image and video inpainting represent critical challenges within the domain of visual information processing, serving pivotal roles in tasks such as restoration, reconstruction, and occlusion removal. The fundamental objective in these applications is to seamlessly restore missing or corrupted regions without introducing perceptible artifacts or structural inconsistencies. While image inpainting operates within a spatially static framework, video inpainting must additionally contend with temporal continuity, necessitating rigorous preservation of motion coherence and inter-frame consistency. This study presents an advanced inpainting methodology grounded in Neutrosophic logic, a mathematical framework adept at modeling uncertainty and indeterminacy within both spatial and intensity domains. By leveraging Neutrosophic-based segmentation, the proposed approach effectively isolates uncertain regions and guides the subsequent restoration process with enhanced precision. The primary objective of this research is to propose a novel approach for constructing an optimal contour database, which serves as a foundation for achieving superior foreground completion. This method aims to enhance the accuracy and visual coherence of the inpainting process by leveraging precise contour information, ultimately improving the overall quality of video restoration. Comprehensive experimental evaluations validate the efficacy of the proposed method, demonstrating notable improvements in perceptual quality and structural fidelity when benchmarked against state-of-the-art alternatives. Nonetheless, certain limitations persist, particularly in processing scenes characterized by high semantic complexity and rapidly changing backgrounds, which constitute promising directions for future investigation.

Keywords: Video inpainting; Video segmentation; Hole filling; Neutrosophic sets; Indeterminacy

Cite this article: Ghanbari Talouki A, Koochari A, Edalatpanah SA. Contour-Based Video Inpainting using Neutrosophic Sets. Fuzzy Optim. Model. J. 2026;7(1): 43-55. https://doi.org/fomj.2026.0701.03

1. Introduction

With the exponential growth of video content in everyday life and the rapid progress in computer-assisted information retrieval systems, the manual processing and analysis of extensive video datasets have become increasingly impractical. Consequently, image and video processing have emerged as indispensable technologies, underpinning a wide array of applications across domains such as medical diagnostics, remote sensing, security surveillance, and digital entertainment.

Among the critical challenges in this domain, is the task of image and video inpainting (also referred to as visual content restoration) which seeks to reconstruct missing, occluded, or corrupted regions within images or video sequences in a visually coherent manner. Unlike image inpainting, which operates solely within the spatial domain, video inpainting must also account for temporal dynamics, thereby requiring a consistent representation of motion across consecutive frames. This dual-domain complexity renders video inpainting significantly more demanding.

Conventional inpainting approaches often rely on propagating textures or patterns from intact regions into the target areas. While such methods can be effective in static contexts, they frequently fall short when applied to dynamic scenes, where the accurate preservation of object trajectories and structural continuity is paramount.

Major limitations include the inability to handle extensive missing regions, difficulties in edge preservation, and the failure to ensure temporal coherence (especially in videos containing substantial motion or multiple moving entities).

Although recent advancements, particularly those driven by deep learning, have yielded promising results, persistent challenges remain. These include frame-to-frame mismatches, disruption in object continuity, and inaccuracies in motion reconstruction (especially in scenarios involving object removal or the repair of severely degraded content). Therefore, any effective video inpainting framework must prioritize the extraction and integration of semantically relevant and temporally consistent information to reconstruct missing areas while faithfully preserving the overall motion integrity of the scene.

This study aims to address the aforementioned challenges, with particular emphasis on video sequences characterized by dynamic foreground elements over static backgrounds. The proposed approach seeks to enhance the accuracy of data selection for inpainting tasks, specifically targeting the reconstruction of extensive missing regions, the preservation of object boundaries, and the maintenance of temporal consistency. To this end, the methodology leverages the analytical capacity of Neutrosophic logic as a robust framework for managing uncertainty and guiding the restoration process.

2. Related work

In recent years, video inpainting has evolved from traditional patch-based and optical-flow-guided methods to advanced deep learning frameworks capable of producing highly realistic and temporally coherent results [1]. Early approaches primarily focused on texture synthesis and spatial similarity, yet often struggled to preserve complex structures across frames. The introduction of structure-aware strategies, such as edge- and contour-guided inpainting, addressed this limitation by explicitly modeling geometric information to maintain object boundaries and scene layout [2]. More recently, the application of transformer architectures, diffusion models, and hybrid techniques has further improved the quality of inpainted videos, enabling better handling of dynamic backgrounds, occlusions, and long-term temporal [3]. Within this landscape, the incorporation of neutrosophic set theory offers a novel direction, providing a mathematical foundation for dealing with uncertainty and indeterminacy inherent in real-world video data, and opening new possibilities for more robust and context-aware video restoration [4].

This section presents a comprehensive overview of the existing literature concerning image and video inpainting techniques. Video inpainting methodologies can generally be classified from two primary perspectives: (1) the scale of inpainting performed at each stage, and (2) the underlying mechanism utilized to execute the inpainting process. Each of these perspectives is examined in detail in the subsequent subsections.

2.1 Video inpainting methods based on inpainting scale per stage

Considering inpainting scale, video inpainting methods are broadly categorized into two major classes: patch-based and object-based approaches.

(a) Patch-Based Inpainting Methods These techniques represent an extension of traditional image inpainting methods that operate by transferring patches from known to unknown regions. While effective for still images, their direct application to videos is limited due to the additional requirement of temporal consistency. To overcome this limitation, various adaptations have been proposed to ensure coherence across frames and maintain motion continuity [5, 6].

(b) Object-Based Inpainting Methods Object-based approaches aim to simultaneously maintain both spatial and temporal coherence by explicitly modeling moving objects. These methods generally yield higher perceptual quality, particularly in dynamic scenes, by preserving object trajectories and structural consistency throughout the video sequence [7, 8, 9].

2.2 Video inpainting methods based on inpainting techniques

From a methodological perspective, video inpainting algorithms can be classified into several key categories based on the strategy employed to reconstruct missing content:

(a) Sample-Based Inpainting

These methods typically follow a structured procedure comprising the following steps: (1) manual specification of the target region, (2) computation of a priority function, (3) search and retrieval of a matching patch from alternative frames, and (4) iterative refinement of the inpainted content [10].

(b) PDE-Based Inpainting

This category relies on partial differential equations to iteratively propagate information from the known boundary inward. While effective for small, textured regions, these methods are computationally intensive and tend to generate blurred results when applied to large missing areas [11] and [4].

(c) Texture Synthesis-Based Inpainting

These approaches are capable of replicating complex textures within small regions; however, their performance significantly degrades in natural scenes and they incur substantial computational costs [9].

(d) Semi-Automatic and Fast Inpainting

These methods operate in two stages: boundary specification through user input, followed by patch-based texture synthesis [12]. Some recent studies have considered one of the main challenges of image and video

https://doi.org/10.57647/fomj.2026.0701.03

inpainting, namely the determination of the hole mask, and have attempted to perform the mask-selection stage automatically with the help of neural networks [13].

(e) Hierarchical Super-Resolution Inpainting

This technique integrates sampling-based inpainting with super-resolution frameworks to reconstruct missing areas while simultaneously enhancing spatial detail [14].

(f) Contour-Based Inpainting

Ghanbari Talouki and Majdi in [7] proposed a contour-based strategy wherein background regions are restored using content from temporally adjacent frames, while missing parts of moving objects are reconstructed via contour similarity metrics. This approach facilitates the effective recovery of both static and dynamic components [15].

(g) Hybrid inpainting methods

Hybrid approaches combine features from multiple inpainting techniques to improve reconstruction quality in complex scenarios, such as scenes with mixed motion or varying texture patterns [7].

Image segmentation represents one of the most complex and fundamental challenges in the domains of image processing and pattern recognition. It serves a pivotal function in numerous advanced applications, including robotic vision, object recognition, and medical image analysis. Segmentation refers to the process of partitioning an image into distinct regions, such that each region is internally homogeneous while being significantly different from its neighboring regions [16]. The growing demand for high-precision and computationally efficient segmentation methods has intensified research efforts in this area [17].

To address the inherent uncertainty in image data, Zadeh introduced fuzzy set theory in 1965 [18], which models uncertainty by assigning membership values within the [0,1]. However, fuzzy sets are limited in their ability to independently express indeterminacy. To overcome this limitation, Smarandache in [19] introduced the concept of neutrosophic logic and sets (an extension of fuzzy theory) designed to more comprehensively model uncertainty. Neutrosophic sets are characterized by three independent membership functions: truth, indeterminacy, and falsity; each taking values within the range [0,1]. This independence enables a more flexible and nuanced representation of uncertainty, particularly in cases involving ambiguous or incomplete information.

2.3 Fuzzy sets

Fuzzy set theory, originally introduced by Lotfi Zadeh in 1965, provides a mathematical framework for modeling uncertainty and imprecision in set membership. Let U denote the universal set, and A, a subseteq in U, be a fuzzy subset in which each element $u \in U$ is associated with a membership function $F_A : U \rightarrow [0, 1]$. The value $F(u) \in [0,1]$ quantifies the degree to which the element u belongs to the set A. A value close to 1 indicates strong membership, while a value near 0 implies weak or negligible membership. Due to their ability to handle vague or imprecise information, fuzzy sets have been extensively applied in various image processing tasks [20].

2.4 Intuitionistic fuzzy sets

An Intuitionistic Fuzzy Set (IFS) extends the concept of classical fuzzy sets by associating each element u in U with two distinct values: a membership degree $\mu_A(u)$ and a non-membership degree $v_A(u)$. In this formulation, the hesitation margin or indeterminacy associated with u in the set A is defined as $\pi_A(u) = 1-\mu_A(u)-v_A(u)$, representing the degree of uncertainty or hesitation regarding the membership status of the element [21].

2.5 Neutrosophic sets

Neutrosophic sets, formulated by Smarandache in [19], generalize fuzzy set theory by explicitly incorporating the concept of indeterminacy in addition to the degrees of truth and falsity. A neutrosophic set A defined over a universal set U is characterized by three independent membership functions: The truth-membership function T_A, the indeterminacy-membership function I_A, and the falsity-membership function F_A. Each of these functions maps elements from U to the non-standard interval $[-0,1+]$, allowing values that may slightly exceed the classical boundaries of [0,1]. Unlike fuzzy sets or intuitionistic fuzzy sets, neutrosophic sets impose no constraint on the sum $supT(u) + supI_A(u) + supF_A(u)$, thereby providing greater flexibility in modeling incomplete, inconsistent, or uncertain information

$$0 \leq supT(u) + supI_A(u) + supF_A(u) \leq 3+$$

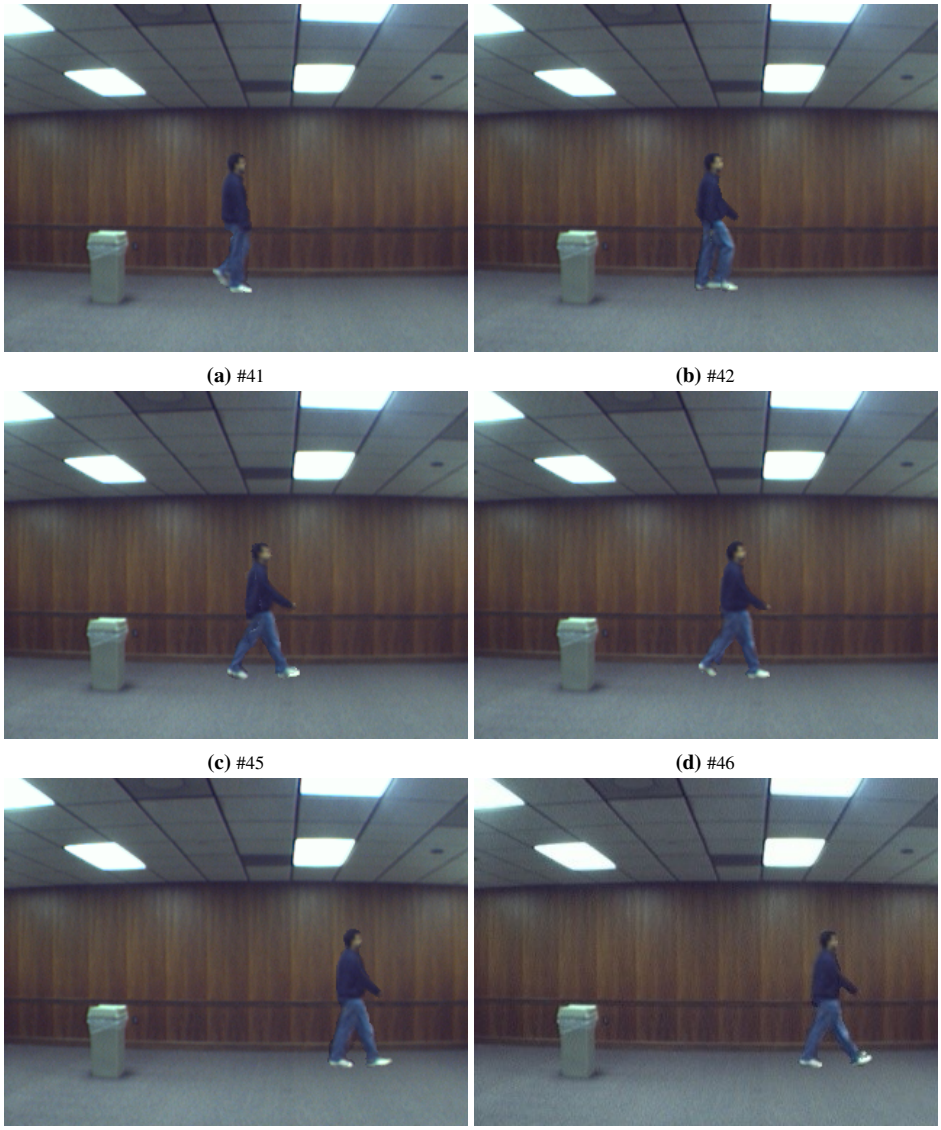
3. Proposed method

Initially, video inpainting techniques were developed as direct extensions of image completion methods, adopting a spatiotemporal perspective. However, as the field progressed, these approaches underwent significant evolution to accommodate the unique characteristics of video data. The principal distinction between image and video inpainting lies in the requirement to maintain temporal continuity of object motion across sequential frames. As a result, the straightforward application of image inpainting algorithms to video often leads to noticeable degradation in visual quality. To address this limitation, it is necessary to adapt and refine such methods to ensure their suitability for video restoration tasks.

Figure 1 presents some frames from a reconstructed video. As illustrated, certain frames exhibit noticeable artifacts; such as misplacement of limbs, where hands and feet are incorrectly reconstructed. In some instances, redundant anatomical features, such as the appearance of three hands, are observed, indicating inconsistencies in the restoration process.

In this study, our objective is to eliminate the occluding object (depicted in green in Fig. 2) and subsequently reconstruct the resulting missing region using an image

https://doi.org/10.57647/fomj.2026.0701.03



(a) #41 **(b)** #42

(c) #45 **(d)** #46

(e) #61 **(f)** #42

Figure 1. Errors encountered in video inpainting (corresponding frame numbers indicated beneath each image). (a) and (b) discontinuity in motion, where (b) exhibits the presence of two right feet. (c) and (d) reconstruction artifact which exhibit two right hands. (e) and (f) duplicated right hands are visible in both frames.

completion technique grounded in segmentation within the neutrosophic domain. Furthermore, our novel video inpainting algorithm which applies contour-based approach to generate the foreground database is applied to reconstruct the moving foreground's movement somehow it should not exhibit discontinuity. This method, previously proposed and assessed in [22], serves as the foundation for our inpainting framework.
The proposed video inpainting methodology comprises the following sequential steps:

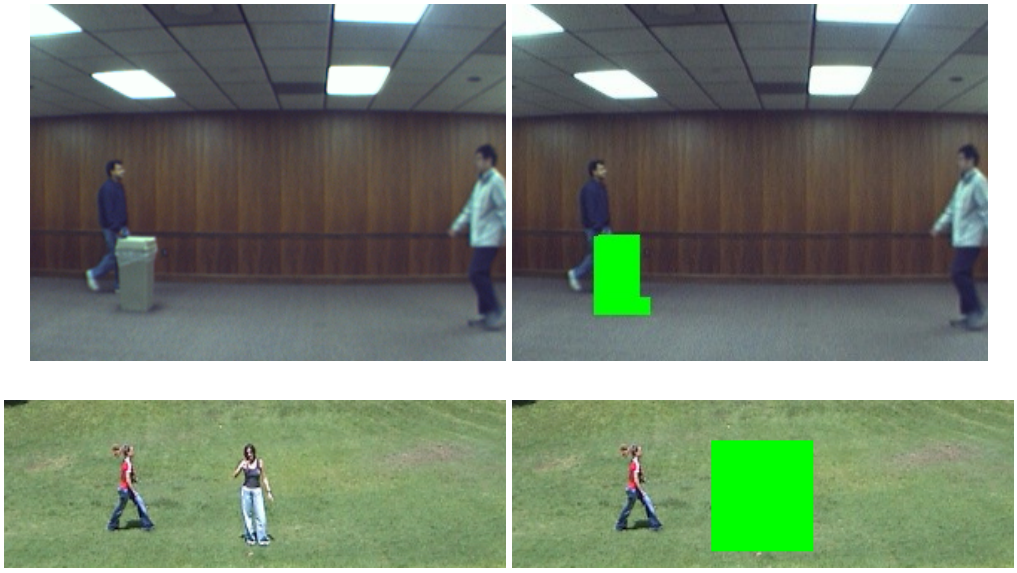
i. The background and the moving foreground objects are separated. The details of this procedure are elaborated in section 3.1.

ii. The occluded background region is reconstructed. In cases where the occluding object remains stationary, no visual information is available from the area behind it; thus, the missing region is filled using a background inpainting technique based on image completion through segmentation within the neutrosophic domain, as described in [22]. Conversely, if the occluding object is in motion, the background information obscured in the current frame may be accessible in other frames. In such instances, the corresponding background content is extracted from adjacent frames and copied into the occluded area to restore visual continuity.

iii. The damaged foreground is completed using the foreground database obtained through the contour-based method in the neutrosophic domain.

3.1 Segmentation of dynamic foreground objects from static background

The extraction of the foreground (moving objects) from the background necessitates the determination of a thresh-



(a) **(b)**

Figure 2. Mask utilized for background reconstruction: (a) An original frame. (b) The occluding object is highlighted in green.

old value, which is applied according to equation (1):

$$FG_t = \begin{cases} 1, & \text{if } |I_t - BG_{t-1}| > thr \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

where I_t denotes the frame captured at time t, BG_{t-1} represents the background model from the previous time step $t-1$, and FG_t corresponds to the extracted foreground at t.

To model the background, a Gaussian Mixture Model (GMM) [23] is employed, wherein each pixel is represented as a combination of k Gaussian distributions. The probability of observing a given pixel value at time t is computed as expressed in equation (2):

$$P(X_t) = \sum_{i=1}^{k} w_{i,t} N\left(X_t, \mu_{i,t}, \sigma_{i,t}\right) \tag{2}$$

where $w_{i,t}$, $\mu_{i,t}$ and $\sigma_{i,t}$ denote the weight, mean, and covariance matrix of the i^{th} Gaussian component at time t.

After modeling each pixel, the background is approximated by selecting the largest B Gaussian components from the mixture model, determined according to equation (3):

$$B = \text{argmin}_b \left(\sum_{i=1}^{b} w_k > T \right) \tag{3}$$

where T denotes the minimum information threshold used to classify pixels as background. Following this step, moving objects are identified by subtracting the estimated background from each video frame.

Subsequent to this separation, residual noise may manifest as spurious regions within the frame. To mitigate this, connected components are labeled and their corresponding areas are computed. Only regions exceeding a predefined area threshold are preserved, while smaller noisy regions are eliminated by setting their pixels to zero. This refinement yields a clean and accurate separation of foreground moving objects from the background. Figure 3 illustrates the initial separation with noise, whereas figure 4 depicts the refined separation post noise removal.

Figure 3. Foreground extraction of a moving person amidst background noise interference.

Figure 4. Refined separation of a moving person from the background following noise.

https://doi.org/10.57647/fomj.2026.0701.03

3.2 Background inpainting

The videos considered in this study feature a static background. To reconstruct the occluded background, we adopt the image inpainting method based on segmentation within a neutrosophic domain, as proposed in [22]. As illustrated in figure 5, the occluding object maintains a fixed position throughout the sequence, allowing for the reconstruction of a single reference frame and subsequent replication of the occlusion region across other frames. The outcome of the background filling process, corresponding to the occlusion shown in Fig. 2 (b), is presented in figure 6.

3.3 Foreground inpainting

In the continuation of this study, neutrosophic sets are employed to achieve more precise contour detection. Initially, each frame is treated as an individual input image and is transformed into the neutrosophic domain using equations (4) through (8).

$$t(i,j) = \frac{\overline{g(i,j)} - \overline{g_{\min}}}{\overline{g_{\max}} - \overline{g_{\min}}} \tag{4}$$

$$i(i,j) = \frac{\delta(i,j) - \delta_{\min}}{\delta_{\max} - \delta_{\min}} \tag{5}$$

$$f(i,j) = 3 - t(i,j) - i(i,j) \tag{6}$$

$$\overline{g(i,j)} = \frac{1}{w \times w} \sum_{m=i-w/2}^{i+w/2} \sum_{n=j-w/2}^{j+w/2} g(m,n) \tag{7}$$

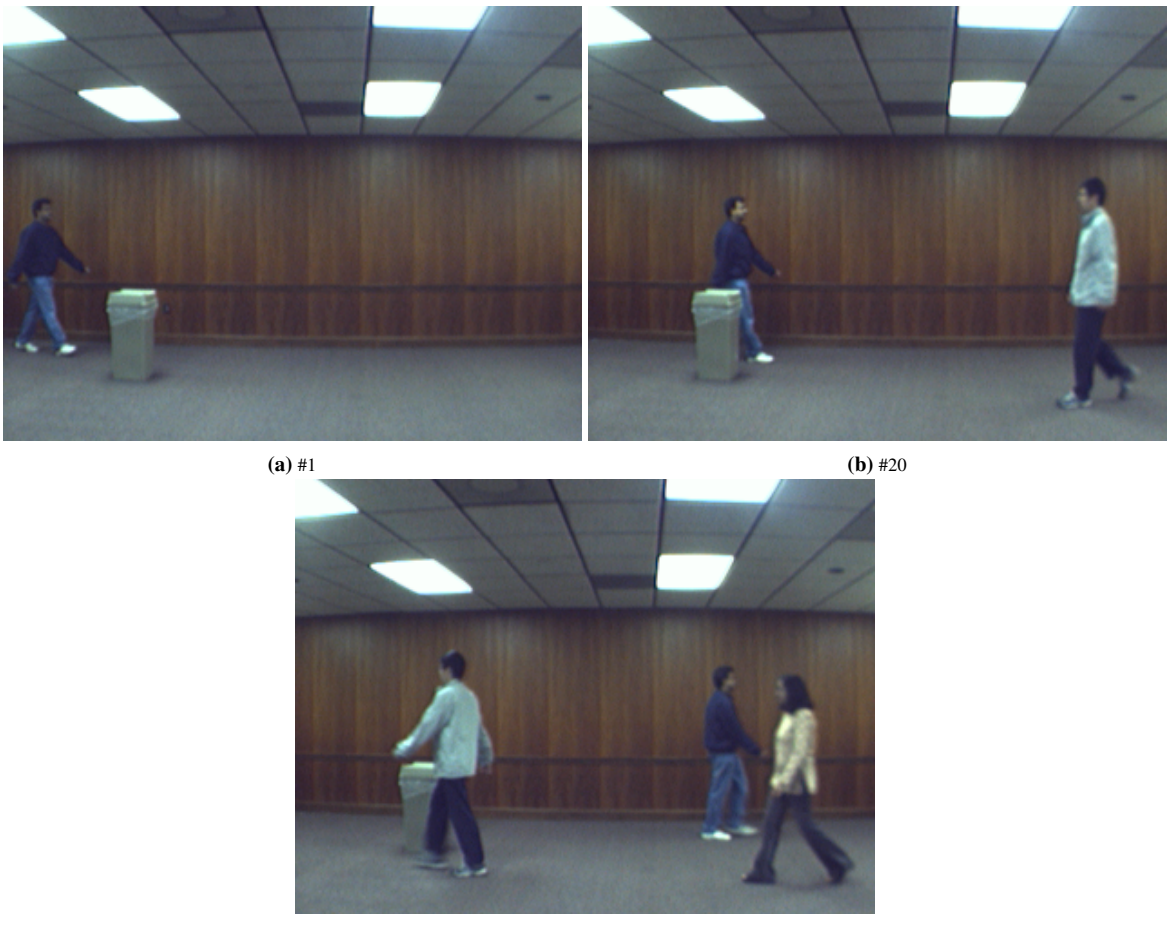
$$\delta(i,j) = abs\left(g(i,j) - \overline{g(i,j)}\right) \tag{8}$$

Figure 7 illustrates the orientation of pixels. As shown in the figure, three principal orientations can be evaluated for each pixel: omnidirectional, horizontal, and vertical. To determine the orientation of a given pixel, if the magnitude of the horizontal gradient exceeds that of the vertical gradient, the pixel is classified as horizontal. Conversely, if the vertical gradient magnitude is greater, the pixel is classified as vertical. In cases where both gradient magnitudes are comparable, the pixel orientation is considered omnidirectional.

The three masks M_1, M_2 and M_3 correspond to the filtering masks for omnidirectional, horizontal, and vertical orientations, respectively.

$$M_1 = \frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

$$M_2 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$



(a) #1 **(b)** #20

(c) #60

Figure 5. A static occlude (trash can). The frame indices are labeled beneath the image. (a) frame number 1. (b) frame number 20. (c)) frame number 60.

https://doi.org/10.57647/fomj.2026.0701.03

Figure 6. Outcome of background inpainting.

$$M_3 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Equations (9) through (11) represent the results of applying the three filtering masks M_1, M_2 and M_3 to a neutrosophic image.

$$R_1 = conv(M_1, Im_{Ns}) \tag{9}$$

$$R_2 = conv(M_2, Im_{Ns}) \tag{10}$$

$$R_3 = conv(M_3, Im_{Ns}) \tag{11}$$

where $conv$ denotes the convolution function. The mean directional function is defined according to equation (12).

$$F\left(Im(i,j)\right) = \begin{cases} R_1 & |G_h(i,j) - G_v(i,j)| \le \theta \\ R_2 & G_h(i,j) - G_v(i,j) > \theta \\ R_3 & G_v(i,j) - G_h(i,j) < \theta \end{cases} \tag{12}$$

where $G_h(i,j)$ and $G_v(i,j)$ represent the horizontal and vertical gradient magnitudes at point (i,j) of the neutrosophic image Im_{Ns}, respectively. The parameter θ denotes a small constant.

Subsequently, the mean directional α is employed based on the pixel orientation within the neutrosophic image. Equations (13) through (15) are introduced for this purpose, with equation (13) defining the updated value of the truth-membership set.

$$\bar{T}(\alpha) = \begin{cases} T(i,j), & I(i,j) < \alpha \\ \bar{T}_\alpha(i,j) \times F(T) & I(i,j) \ge \alpha. \end{cases} \tag{13}$$

where $\bar{T}_\alpha(i,j)$ denotes the intersection of the truth-membership values of pixels within a neighborhood window of size w in the neutrosophic image Im_{Ns}, with pixel (i,j) located at the center of this window. Equation (14) specifies the updated value of the indeterminacy set.

$$\bar{I}_\alpha(i,j) = \frac{\bar{\delta}_T(i,j) - \bar{\delta}_{T_{\min}}}{\bar{\delta}_{T_{\max}} - \bar{\delta}_{T_{\min}}} \tag{14}$$

$$\bar{\delta}_T(i,j) = \left|\bar{T}(i,j) - \bar{\bar{T}}(i,j)\right| \tag{15}$$

$$C = A \cap B = \{\min(T_A, T_B), \max(I_A, I_B), \max(F_A, F_B)\} \tag{16}$$

where $\bar{\bar{T}}(i,j)$ denotes the intersection of the truth-membership values of pixels within a neighborhood

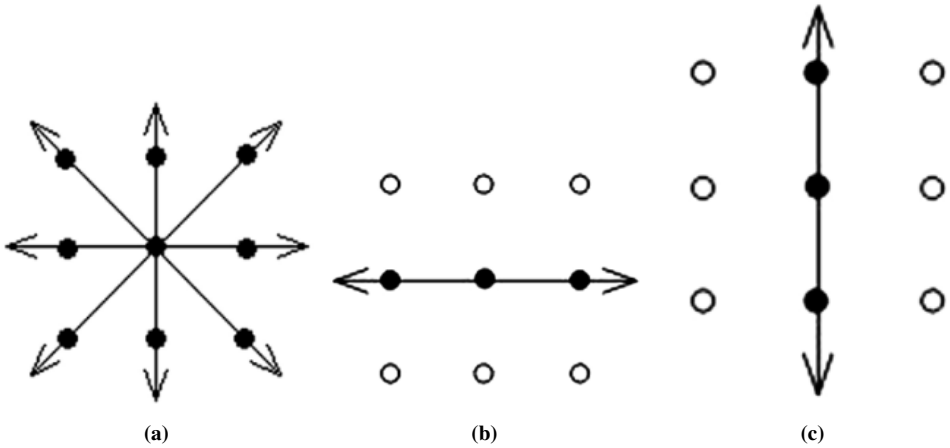


Figure 7. Pixel orientations: (a) Omnidirectional, (b) Horizontal, (c) Vertical.

https://doi.org/10.57647/fomj.2026.0701.03

window of size w in $\bar{T}_\alpha(i,j)$, with pixel (i,j) positioned at the center of this window. As described in equation (16), the intersection of truth-membership values of two points in neutrosophic sets is defined as the minimum truth value of those points.

Since the values of the truth and indeterminacy sets play a critical role in contour determination, their updated values are computed according to equations (13) through (15). The falsity set remains unchanged. By adjusting the constant α, the sharpness and smoothness of the detected contours can be effectively controlled. Based on the above considerations, the updated neutrosophic sets are formulated as given in equation (17).

$$\bar{p}(\alpha) = \{\bar{T}(\alpha), \bar{I}(\alpha), F\} \tag{17}$$

The mean directional α is iteratively applied to generate more precise edges. The gradient magnitude of the pixels in $\bar{T}(\alpha)$ is utilized to determine the degree of pixel membership to edge pixels. For this purpose, a threshold on the gradient magnitude is required to decide whether a pixel belongs to an edge or not. Consequently, equation (18) is introduced to define this thresholding criterion.

$$E_{NS}(Z) = \begin{cases} 1, & \|\nabla(\bar{T}(\alpha))\| \geq \nabla_{th} \\ 0, & \|\nabla(\bar{T}(\alpha))\| < \nabla_{th} \end{cases} \tag{18}$$

where $\|\nabla(\bar{T}(\alpha))\|$ denotes the norm of the gradient of $\bar{T}(\alpha)$, and ∇_{th} is the threshold value used to identify edge pixels. Equations (12) and (18) are applied iteratively; the stopping criterion is met when the extracted edges remain unchanged between iterations. The contour detection algorithm for neutrosophic images is outlined as follows:

Neutrosophic image contour detection algorithm

Start

1)Input image Im, and set iteration counter $t = 0$.

2) Transform the initial image $Im^{(0)}$ into the neutrosophic domain to obtain $Im_{NS}^{(0)}$ using equations (4-8).

3) Apply equation (12) on $Im_{NS}^{(0)}$ to generate $Im_F^{(0)}$.

4) Calculate $E^{(0)}$ using equation (18).

5) Set $t = t + 1$.

6) Apply equation (12) on $Im_F^{(t-1)}$ to obtain $Im_F^{(t)}$.

7) Compute $E^{(t)}$ via equation (18).

8) If $|E^{(t)} - E^{(t-1)}| \geq \varepsilon$, then return to Step 5.

9) $E^{(t-1)}$ is the final extracted edge map.

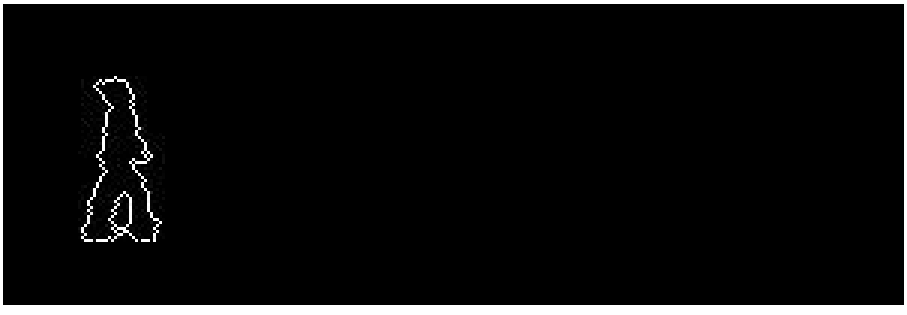
End

For each video frame treated as an input image, the above algorithm is employed to extract edges. Subsequently, edges corresponding to foreground pixels in the foreground-background segmentation map (i.e., white pixels in the segmentation mask) are considered as contours of the moving object. Edges not aligning with white regions in the segmentation mask are discarded, as they do not belong to the target moving object.

Figure 8 illustrates the extracted contour for a sample video frame. To construct the database, the moving object must be extracted in every frame. Initially, a bounding rectangle enclosing the moving object is determined.

Given that contours are employed in this research for comparison and selection of the most appropriate

(a) #41



(b) #42

Figure 8. Contour database generation: (a) A video frame, (b) Contour of the corresponding frame.

https://doi.org/10.57647/fomj.2026.0701.03

pattern, a contour database is required. To construct this database, the contour video (where only the foreground contours are displayed) is utilized. In each frame of this video, a bounding rectangle is extracted, the dimensions of which are described in the following.

As illustrated in Fig. 8 (b), the spatial coordinates of each pixel belonging to the contour are obtained. These coordinates are then sorted in ascending order based on their row and column values. The pixels corresponding to the topmost, bottommost, leftmost, and rightmost edges of the contour are identified and denoted as minRow, maxRow, minCol, and maxCol, respectively. Consequently, the bounding rectangle for each frame is defined as (minCol: maxCol) × (minRow: maxRow). Figure 9 depicts a segment of the database constructed up to this stage.

Subsequently, the foreground can be extracted in each frame based on the obtained coordinates. Figure 10 illustrates a video frame where only the foreground has been extracted. With the foreground-extracted video and bounding rectangle coordinates available for each frame, the database of the moving object is constructed. Figure 11 presents a segment of the resulting database. This dataset is referred to as the foreground database and is utilized for hole filling in the completion process.

Foreground reconstruction considering a partially damaged person

Considering a partially damaged person introduces two key issues. The first is whether the partially damaged person can be used for comparison and identification of the most suitable pattern. The second is how to retain the intact regions of the partially damaged person while completing the remaining parts of the hole. We begin by addressing the first issue, followed by a discussion of the second.

The contour of the foreground for the frame where the hole first appears, denoted as conh, is considered for comparison within the contour database. A potential question arises: can a reliable comparison result be achieved using a partially damaged contour? The answer is that the initial damaged frame typically exhibits only minor impairment, leaving the remaining portions of the contour sufficiently usable. Furthermore, experimental results indicate that employing the damaged contour for comparison still produces satisfactory outcomes in video completion. For the comparison and identification of the most suitable pattern, equation (19) is employed.

$$d(h,i) = |con_h - con_i|, \; i = 1,2,\ldots N, \quad i \neq h, h+1, \ldots h+n-1 \tag{19}$$

We consider the five most suitable contours and perform a second comparison among these candidates. The second comparison criterion is employed to evaluate the preservation of continuity. If a valid sequence of length n exists, equation (20) is used as the second comparison criterion. However, if no valid sequence of length n is available, equation (21) is applied as the alternative

Figure 9. A segment of the contour database. For ease of visualization, contours from different frames are displayed side by side.

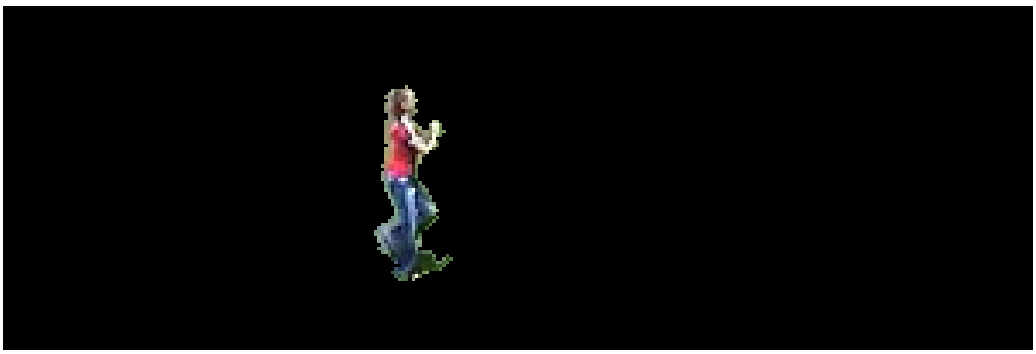


Figure 10. A frame from the foreground-extracted video.

Figure 11. A segment of the foreground database. For ease of visualization, the moving object from different frames is displayed side by side.

https://doi.org/10.57647/fomj.2026.0701.03

criterion.

$$B = \arg\min_i \Big(d(h-1, i-1) + \sum_{k=1}^{2} d(h+n+k-1, i+n+k) \Big), \; i = 1,2,3,4,5. \tag{20}$$

$$B = \arg\min_i \sum_{k=1}^{2} d(h+n+k-1, i+n+k), \quad i = 1,2,3,4,5. \tag{21}$$

In case using equation (21), an iterative procedure is done to find the most suitable template in each completion step. Therefore, the most similar contour to con_h is selected which is named as con_B.

The partially occluded moving object's data base is applied to complete the hole; the green part is replaced by its corresponding pixels' data. It is done for all hole frames. Figure 12 depict some inpainted frames using this approach. By utilizing the precise contours and bounding information of the moving objects, the proposed framework ensures more accurate reconstruction of the missing regions, thereby enhancing both spatial coherence and the perceptual quality of the completed video. This integration of the proposed method with the constructed foreground database provides a substantial improvement over existing approaches, particularly in maintaining object continuity and reducing visual artifacts. Figure 13 gives a comparative result demonstrating the superior performance of the proposed method versus the approach in [9].

4. Evaluation of video inpainting results

4.1 Qualitative Evaluation

Figure 13 presents a qualitative comparison between the results produced by the proposed method and those obtained using the approach described in [9]. Selected frames from each method are shown to visually assess the effectiveness of background reconstruction and the consistency of object removal across consecutive frames. You can visit [24] and [25] for the initial video and the two inpainted videos, respectively.

4.2 Quantitative evaluation

To conduct a quantitative assessment of video inpainting performance, some quantitative metrices are used which are introduced as following:

(a) Average Squared Visual Salience (ASVS)

ASVS represents the amount of error present within a hole and is expressed as a single value. This value arises from intra-region distortions and becomes apparent when the target region directly attracts the viewer's attention. Therefore, a lower intra-region value indicates better performance of the algorithm. As evident, the smaller the ASVS value, the better the inpainting of the target region [26]. Equations [22] and [23] are applied to get ASVS.

$$\text{in-region(I)} = \frac{1}{|\Omega|} \sum_{p \in \Omega} [S(p)]^2 \tag{22}$$

$$\text{ASVS(I)} = \text{in-region(I)} \tag{23}$$

where Ω denotes the target region, I is the original image, and $S(p)$ is the saliency map for the completed image.

Table 1 presents the ASVS values obtained using the proposed method in comparison with those reported by the method in [4, 9].

(b) Structural Similarity (SSIM)

The Structural Similarity Index was introduced by Wang and colleagues in [27] to provide a quantitative metric for comparing image inpainting results when the original image is available. The SSIM index is obtained from equation (24).

$$\text{SSIM}(x,y) = [l(x,y)]^{\alpha}, [c(x,y)]^{\beta}, [s(x,y)]^{\lambda} \tag{24}$$

where $l(x,y)$, $c(x,y)$, and $s(x,y)$ represent the luminance, contrast, and structural components of the two images, respectively, and are obtained from equations (25) to (27).

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{25}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{26}$$

$$s(x,y) = \frac{2\sigma_{xy} + C_3}{\sigma_x^2 \cdot \sigma_y^2 + C_3} \tag{27}$$

(a) #8 **(b)** #22

Figure 12. Some frames from the video completed using the proposed method. (a) frame number 8. (b) frame number 22.

https://doi.org/10.57647/fomj.2026.0701.03

(a) #10 **(b)** #10

Figure 13. Comparison of video completion results using the method proposed in this approach and the method in [9] (column-wise). (a) Video frames completed using the propose method. (b) Video frames completed using the method in [9].

Table 1. Quantitative comparison using the ASVS metric.

[9]	0.1374
[4]	0.1174
Proposed method	0.0961

Table 2. Quantitative comparison using the MSSIM metric.

[9]	0.9886
[4]	0.9905
Proposed method	0.9919

If in equation (24), $\alpha = \beta = \lambda = 1$, then the SSIM can be expressed by equation (28):

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_x\sigma_y + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (28)$$

where C_1 and C_2 are stability constants. Since a large number of values are obtained based on the number of windows, the average SSIM over all windows is considered as the comparison metric for an image and is calculated using equation (29):

$$\mathrm{MSSIM}(X, Y) = \frac{1}{M}\sum_{j=1}^{M} \mathrm{SSIM}(x_j, y_j) \quad (29)$$

where X is the original image, Y is the damaged image, M is the number of windows, and x_j and y_j are the pixels of the j^{th} window of the images. For two identical images, the MSSIM equals one; therefore, the closer the obtained MSSIM is to one, the greater the similarity between the two images. Table 2 compares the MSSIM values for the proposed method and method in [9] and [4].

(c) Peak Signal-to-Noise Ratio (PSNR)

One of the most important indices for quality assessment is PSNR, which is calculated using equations (30) and (31) ([28]).

$$\mathrm{MSE} = \frac{1}{M \cdot N}\sum_{i=1}^{m}\sum_{j=1}^{n} \|I(i, j) - M(i, j)\|^2 \quad (30)$$

$$\mathrm{PSNR} = 10 \cdot \log\left(\frac{255^2}{\sqrt{\mathrm{MSE}}}\right) \quad (31)$$

where MSE is the mean squared error, m and n are the dimensions of the image, and $I(i, j)$ and $M(i, j)$ are the pixels of the original and the completed images, respectively. Table 3 compares the PSNR values for the proposed method and previous approaches.

Table 3. Quantitative comparison using the MSSIM metric.

[9]	29.60
[4]	34.51
Proposed method	38.96

https://doi.org/10.57647/fomj.2026.0701.03

5. Conclusion

Image and video completion are pivotal processes in the domain of image and video analysis, particularly for restoring degraded or missing regions within visual content. While the core objective of image completion is to reconstruct damaged areas in a visually coherent manner that avoids noticeable discrepancies, video completion imposes additional complexities. Specifically, it necessitates the preservation of temporal consistency across consecutive frames. In scenarios involving moving objects, this implies that the synthesized content must not only spatially integrate with its surroundings but also adhere to the temporal dynamics of motion, ensuring continuity and realism throughout the video sequence.

To reconstruct the motion of the moving foreground object, a contour-based similarity criterion was employed to identify the most suitable data. For this purpose, both the contour database and the foreground database were utilized. To obtain more precise contours, video frames were transformed into the neutrosophic domain, where their contours were extracted. Finally, the completed background and the reconstructed foreground were integrated to generate the final video.

Transferring the image to the neutrosophic domain resulted in more accurate contours, and since these contours ultimately lead to the selection of the most appropriate pattern for hole filling, this transfer consequently improves the visual quality of the obtained video. In addition, the quantitative evaluation metrics ASVS, MSSIM, and PSNR provide numerical validation of the obtained results. Because ASVS represents the Average Squared Visual Saliency, a lower ASVS value indicates fewer visual distortions in the inpainted image; the proposed method in this paper achieves a lower ASVS value for the final result. Another quantitative parameter is MSSIM, which is the mean of SSIM and is used to evaluate image quality from the perspective of structural similarity. Simulation results show that the proposed method, by relying on more suitable data for completion, achieves a higher MSSIM value. Furthermore, the PSNR metric indicates that using more accurate contours can reduce the resulting error as much as possible, which leads to a lower error value and consequently a higher PSNR.

Suggestions for future work

Based on the findings of the video inpainting method presented in this dissertation, the following limitations and suggestions are outlined for future research:

- The videos analyzed in this study involved static backgrounds; dynamic backgrounds were not considered.
- Variations due to camera motion and changes in the size of moving objects across frames were not addressed.
- Scenarios where no intact sample of a moving object is available in other frames were not explored. In such cases, reconstructing the motion trajectory of the moving object becomes necessary.

Authors contributions
All authors contributed equally to the conception, design, execution, and writing of this work. All authors read and approved the final manuscript.

Availability of data and materials
The authors declare that the data supporting the findings of this study are available within the paper.

Conflict of interests
The authors assert that they do not have any identifiable conflicting financial interests or personal relationships that might be perceived to influence the work presented in this paper.

References

1. Yang S, Gu Z, Hou L, Tao X, Wan P, Chen X, and Liao J. Mtv-inpaint: Multi-task long video inpainting. 2025; arXiv:2503.11412. DOI: 10.48550/arXiv.2503.11412
2. Xie C, Han K, and Wong KYK. VipDiff: Towards coherent and diverse video inpainting via training-free denoising diffusion models. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) 2025 :2411–20. DOI: 10.1109/WACV61041.2025.00240
3. Koochari A and Soryani M. Exemplar-based video inpainting with large patches. Journal of Zhejiang University-SCIENCE C (Computers & Electronics) 2010; 11:270–7. DOI: 10.1631/jzus.C0910308
4. Wu X and Liu C. DiTPainter: Efficient video inpainting with diffusion transformers. 2025; arXiv:2504.15661:245–69. Available from: https://doi.org/10.48550/arXiv.2504.15661
5. Abraham AR, Prabhavathy AK, and Shree JD. A survey on video inpainting. International Journal of Computer Applications 2012; 56(9):43–7. DOI: 10.5120/8923-2761
6. Talouki AG, Koochari A, and Edalatpanah SA. Applications of neutrosophic logic in image processing: a survey. Journal of Electrical and Computer Engineering Innovations (JECEI) 2022; 10(1):243–58. DOI: 10.22061/jecei.2021.8069.474
7. Talouki AG and Majdi M. Improvement in video inpainting in presence of moving subjects. International Journal of Research in Industrial Engineering 2019; 8(4):325–38. DOI: 10.22105/riej.2020.215689.1116
8. Koochari A and Soryani M. Video object inpainting: a scale-robust method. Imaging Science Journal 2012; 60(5):272–84. DOI: 10.1179/1743131X11Y.0000000047
9. Li X, Xue H, Ren P, and Bo L. Diffueraser: A diffusion model for video inpainting. 2025; arXiv:2501.10018. Available from: 10.48550/arXiv.2501.10018

10. Talouki AG, Majdi M, and Edalatpanah SA. An introduction to various algorithms for video completion and their features: a survey. Journal of Computer Sciences and Applications 2017; 5(1):1–10. DOI: 10.12691/jcsa-5-1-1
11. Bertalmio M, Bertozzi AL, and Sapiro G. Navier-stokes, fluid dynamics, and image and video inpainting. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) 2001. DOI: 10.1109/CVPR.2001.990497
12. Ravi S, Pasupathi P, Muthukumar S, and Krishnan N. Image in-painting techniques-A survey and analysis. 9th International Conference on Innovations in Information Technology (IIT) 2013 :36–41. DOI: 10.1109/Innovations.2013.6544390
13. Wu Z, Chen K, Li K, Fan H, and Yang Y. BVINet: Unlocking blind video inpainting with zero annotations. Proceedings of the IEEE/CVF International Conference on Computer Vision 2025 :14017–27. DOI: 10.48550/arXiv.2502.01181
14. Bian Y, Zhang Z, Ju X, Cao M, Xie L, Shan Y, and Xu Q. Videopainter: Any-length video inpainting and editing with plug-and-play context control. Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference 2025 :1–12. DOI: 10.1145/3721238.3730673
15. Talouki AG and Soryani M. Contour-based video inpainting. 7th Iranian Conference on Machine Vision and Image Processing 2011 :1–5. DOI: 10.1109/IranianMVIP.2011.6121586
16. Nilsson D and Sminchisescu C. Semantic video segmentation by gated recurrent flow propagation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2018 :6819–28. DOI: 10.48550/arXiv.1612.08871
17. Wang P, Cai Z, Yang H, Swaminathan A, Manmatha R, and Soatto S. Scaling up image segmentation across data and tasks. Proceedings of the Computer Vision and Pattern Recognition Conference 2025 :4573–83. DOI: 10.1109/CVPR52734.2025.00431
18. Zadeh LA. Fuzzy sets. Information and Control 1965; 8(3):338–53. DOI: 10.1016/S0019-9958(65)90241-X
19. Smarandache F. Neutrosophy: neutrosophic probability, set, and logic: analytic synthesis & synthetic analysis. 1998; American Research Press. DOI: 10.5281/zenodo.57726
20. Zhang X, Jian M, Sun Y, Wang H, and Zhang C. Improving image segmentation based on patch-weighted distance and fuzzy clustering. Multimedia Tools and Applications 2020; 79(1):633–57. DOI: 10.1007/s11042-019-08041-x
21. Afful-Dadzie E, Oplatkova ZK, and Prieto LAB. Comparative state-of-the-art survey of classical fuzzy set and intuitionistic fuzzy sets in multi-criteria decision making. International Journal of Fuzzy Systems 2017; 19(3):726–38. DOI: 10.1007/s40815-016-0204-y
22. Talouki AG, Koochari A, and Edalatpanah SA. Image completion based on segmentation using neutrosophic sets. Expert Systems with Applications 2024; 238:121769. DOI: 10.1016/j.eswa.2023.121769
23. Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction. Proceedings of the 17th International Conference on Pattern Recognition (ICPR) 2004 :28–31. DOI: 10.1109/ICPR.2004.1333992
24. Available from: https://drive.google.com/file/d/1RJM7ty3oYnxo-tpbmI_Cm2EuXHiJCASC/view?usp=drive_link
25. Available from: https://drive.google.com/file/d/182ruC2X1vBZeVDvOicKAIRekMDvdXxis/view?usp=sharing
26. Ardis PA, Brown CM, and Singhal A. Inpainting quality assessment. Journal of Electronic Imaging 2010; 9(1):11002–8. DOI: 10.1117/1.3267088
27. Wang Z, Bovik AC, Sheikh HR, and Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing 2004; 13(4):600–12. DOI: 10.1109/TIP.2003.819861
28. Wang J, Xuan H, and Wu Z. Semantic-guided completion network for video inpainting in complex urban scene. Chinese Conference on Pattern Recognition and Computer Vision (PRCV) Singapore 2023; Springer:224–36. DOI: 10.1007/978-981-99-8552-4_18