









Real-World Video Denoising for Visual Inspection in High-Dose Radiological Environments

Andrew Young,^{a*} Zhouxiang Fei,^{a} Jaime Zabalza,^{b} Graeme M. West,^{a} Paul Murray,^{b} and Stephen D. J. McArthur^{a}

^aUniversity of Strathclyde, Institute for Energy and Environment, Glasgow, United Kingdom

^bUniversity of Strathclyde, Institute for Sensors, Signals and Communications, Glasgow, United Kingdom

Received May 21, 2024

Accepted for Publication September 23, 2024

Abstract — *In high-dose radiological environments, where precision and safety are of utmost importance, the ability to acquire accurate and clear visual information is of paramount importance for ensuring safety and reliability in critical industrial processes. However, these environments inherently introduce significant challenges due to the adverse effects of radiation on imaging equipment. As a consequence, inspection videos captured within such high-radiation environments often contain a significant amount of noise. This noise substantially complicates the task of identifying and assessing potential defects in vital components. It also diverts attention and resources toward investigating false positives created by noise, leading to inefficiencies, and for industrial processes on the critical path, this can further prolong the outage. Addressing this noise is essential not only for precision, but also for ensuring safety, reliability, and efficiency in critical industrial processes.*

In this paper, we present a custom-designed filter utilizing a priori information about camera position and trajectory to remove the noise from the inspection videos, making the defects easier to manually identify. As the camera movement is in one direction at a constant speed, the proposed approach uses this temporal and spatial information to accurately remove the noise. This approach applies to a subset of visual inspection problems throughout the nuclear industry, as well as many other industries where there is knowledge available about the camera speed and direction of travel. The proposed approach is compared with three accepted/well-known approaches, median filtering, bilateral filtering, and fast nonlocal means denoising, and an additional state-of-the-art deep learning model is also used for comparison. It was found that the proposed approach produces the most accurate video denoising in terms of visual quality and the retainment of the defect features throughout the videos tested.

Keywords — *Video denoising, image processing, in-core reactor inspections, MCR pressure tube replacements, nuclear power plant.*

Note — *Some figures may be in color only in the electronic version.*

*E-mail: andrew.young.101@strath.ac.uk

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

I. INTRODUCTION

Visual inspection plays a crucial role in the field of nuclear power, enabling the examination and assessment of various components, systems, and structures within nuclear facilities.^[1] While various inspection techniques exist, including in situ visual inspections, remote boroscopic inspections,^[2] and robotically deployed inspections,^[3] they all face a common challenge: radiation-induced noise

in imaging equipment. This noise significantly complicates the identification and assessment of potential defects in critical components,^[4] often leading to false positives and inefficiencies in industrial processes. The ability to acquire clear, accurate visual information in these high-radiation environments is essential to ensuring the safety, reliability, and efficiency of nuclear facilities.

In the majority of inspection applications, the collected visual data are analyzed and used to assess the condition of nuclear components, identify defects, plan maintenance or repair activities, and ensure the safe and reliable operation of nuclear facilities. These inspections contribute to the overall safety, efficiency, and regulatory compliance of nuclear facilities by providing vital information for decision making and preventive maintenance. Inspecting the inside of a nuclear reactor core with any type of camera is challenging due to the high-dose radiological environment. This can lead to artifacts, distortions, and fluctuating brightness in the videos, making it difficult to obtain clear and reliable images. Over time, radiation-induced damage to camera components can further degrade image quality.^[5] To address these challenges, specialized cameras and shielding techniques have been developed to mitigate the impact of radiation noise and provide clearer images for analysis while ensuring the safety of operators and equipment in this extreme environment.^[6]

In this work, we present and evaluate a novel denoising technique developed for real-world data in high-dose radiological environments, where acquiring accurate and clear visual information is crucial to ensuring safety and reliability in critical industrial processes. The primary contributions of this paper include the development of a custom-designed filter that utilizes a priori information about camera position and trajectory to remove noise from inspection videos, making defects easier to manually identify. This approach is particularly valuable in nuclear and other industries where camera speed and direction of travel are known.

Our method addresses the significant challenges posed by radiation-induced noise in imaging equipment, which often leads to complications in identifying potential defects in vital components and can result in inefficiencies due to false positives. The proposed technique leverages temporal and spatial information, taking advantage of the camera's constant speed and unidirectional movement to accurately remove noise. We provide a comprehensive evaluation and benchmarking of our approach against three well-established methods: median filtering, bilateral filtering, and fast nonlocal means denoising (FNLM), as well as a state-of-the-art deep learning model.

Our analysis demonstrates that the proposed method outperforms existing techniques in terms of visual quality and retention of defect features across various test videos. [Section II](#) provides essential background on video denoising techniques that are crucial for understanding existing methodologies. The proposed approach is introduced in [Sec. III](#), followed by its evaluation and benchmarking in [Sec. IV](#). [Section V](#) concludes the paper and outlines directions for future work.

II. BACKGROUND

II.A. Video Denoising

In digital image processing, images often contain unwanted noise that can significantly degrade the quality of the image. Noise can appear as random variations in pixel values, resulting from factors such as sensor limitations, transmission errors, or environmental conditions.^[7,8] The primary goal of video denoising is to reduce or eliminate noise and unwanted artifacts that may degrade the visual quality of the videos. Video denoising in certain scenarios can be challenging due to various factors. One of the primary challenges arises from the diverse nature of the noise itself, which can manifest in different forms, such as Gaussian noise, salt-and-pepper noise, or even more complex patterns.^[8]

Each type of noise may require specialized techniques for effective removal, making it challenging to develop a universal denoising solution. Spatial filtering techniques are an important part of video denoising approaches.^[7] These methods operate on individual frames independently and aim to attenuate noise while preserving important details. Temporal filtering techniques, on the other hand, leverage the temporal coherence present in a video sequence.^[7] They utilize information across multiple frames to enhance denoising performance. Additionally, machine learning-based approaches, such as deep neural networks, have demonstrated significant potential in video denoising by learning complex mappings between noisy and clean video data sets.^[8]

In this section, three well-known conventional methods, median filtering, bilateral filtering and FNLM, are briefly introduced along with a state-of-the-art deep learning-based approach, all of them used for performance comparison.

II.A.1. Median Filtering

Median filtering is a widely used technique in video denoising^[9,10] that aims to reduce noise and enhance the

visual quality of video sequences. Similar to its application in image processing, median filtering in video denoising involves replacing each pixel value in a video frame with the median value of its neighboring pixels. This technique is particularly effective in scenarios where videos suffer from noise, such as salt-and-pepper noise, which appears as white and black pixels. However, its performance can be limited in scenarios involving complex noise patterns or when the noise level is high, potentially leading to the loss of fine details and reduced effectiveness in preserving video sharpness.

Median filtering successfully eliminates these outliers without blurring or distorting the important motion details of the video. With its ability to enhance video quality and reduce noise artifacts, median filtering plays a vital role in various domains, including surveillance systems,^[11] video compression,^[12] and video restoration.^[13]

II.A.2. Bilateral Filtering

Bilateral filtering is another technique used in video denoising to preserve edges and fine details while effectively reducing noise.^[14] Unlike traditional filtering methods that consider only the spatial neighborhood of a pixel, bilateral filtering takes into account both the spatial and intensity similarities between pixels. It applies a weighted average to the neighboring pixels based on their spatial proximity and similarity in intensity values. However, it can be computationally expensive and may struggle with preserving fine details in areas of high contrast or with large amounts of noise.

By incorporating this additional intensity similarity factor, bilateral filtering can differentiate between noise and important image features, thereby preserving edges and fine details during the denoising process. This makes bilateral filtering suitable for denoising videos with complex motion and fine details, such as those encountered in video surveillance^[15] and medical imaging applications.^[16]

II.A.3. FNLM Filtering

Fast nonlocal means denoising filtering is a technique used for video denoising that offers superior noise reduction while preserving fine details and textures.^[17] FNLM filtering is based on the concept of nonlocal means denoising, which exploits redundancy in natural images to remove noise effectively. Unlike traditional denoising techniques that only consider local neighborhoods, FNLM considers the similarity of patches from different

regions of the image to determine the denoised value of a pixel. This allows the algorithm to capture and utilize information from distant and structurally similar patches, resulting in more accurate noise estimation and reduction.

However, its computational complexity can be high, potentially limiting its efficiency and scalability for large data sets or real-time processing. FNLM filtering is used throughout many different domains, such as video surveillance,^[18] microscopy,^[19] and medical imaging,^[20] where preserving fine details and textures is crucial for accurate analysis and interpretation.

II.A.4. Recurrent Video Restoration Transformer

The recurrent video restoration transformer (RVRT) is a recent deep learning architecture designed to enhance video restoration tasks by effectively processing local neighboring frames in parallel within a globally recurrent framework. This novel approach aims to integrate the advantages of existing methods, addressing their respective limitations in model size, effectiveness, and efficiency.^[21] By dividing the video into multiple clips, RVRT utilizes previously inferred clip features to estimate subsequent ones, achieving a balanced trade-off between high model efficiency and performance.

Within each clip, frame features are updated through implicit feature aggregation, while guided deformable attention facilitates accurate alignment across different clips by predicting and aggregating relevant features from the entire inferred clip. Extensive experiments across various video restoration tasks, including super resolution, deblurring, and denoising, demonstrate that RVRT achieves state-of-the-art performance on benchmark data sets. It effectively maintains a well-balanced model size, testing memory, and run time, showcasing its efficiency and effectiveness in comparison to other approaches.^[22]

II.B. Motion Estimation

Motion estimation is a fundamental process in video analysis and processing, essential for tasks such as video compression, stabilization, and denoising. It involves determining the movement of objects or the camera between successive frames to align them accurately.

Traditional motion estimation methods of motion estimation include block matching^[23] and optical flow.^[24] Block matching divides an image into small blocks and searches for the best match in the subsequent frame to estimate motion vectors. While effective, this approach can be computationally intensive and struggles

with issues such as occlusions and variable motion patterns. Optical flow methods estimate motion based on the apparent velocity of pixel intensities, providing a more detailed motion representation but also demanding significant computational resources.

More recent developments have introduced feature-based methods and deep learning approaches. Feature-based methods track distinctive points or features across frames, offering improved robustness against noise and occlusions.^[25] These methods can provide accurate motion estimates with reduced computational cost compared to block matching. Deep learning techniques have further enhanced motion estimation accuracy by learning complex motion patterns from large data sets. However, these methods can be computationally expensive and require substantial training data.^[25]

III. METHODOLOGY

III.A. Algorithm

Video denoising is crucial to improving the visual quality of videos captured under noisy conditions. In scenarios where camera movement is consistent and predictable, motion estimation techniques can be leveraged to effectively denoise the video. This is because such techniques exploit the temporal redundancy present in consecutive frames. By calculating the motion between frames, these techniques can align corresponding pixels across frames, allowing for more accurate averaging or filtering of noisy pixels.

This approach is particularly advantageous given that the same surface or object in the video will appear in several frames with noisy pixels in different positions. This methodology section outlines the steps involved in our proposed video denoising technique, which employs motion estimation, frame realignment, and pixelwise temporal median filtering to produce high-quality denoised videoframes.

III.A.1. Video Denoising Algorithm

Initially, the frames are extracted from the video at the desired frame rate, resulting in a sequence of frames. The video denoising process involves selecting a set of frames before and after the frame under analysis, forming a temporal window. The colored shapes (circle, triangle, and square) in Fig. 1 represent distinct objects or features in the frame. Note that noise is not shown in this figure to clearly illustrate object movement.

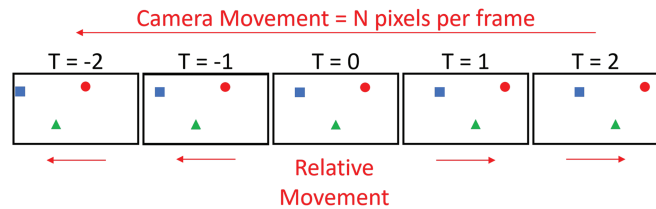


Fig. 1. Frame relative movement for the frame under analysis based on camera movement. In this example, objects in the frame move horizontally in the left to right direction.

In this approach, prior knowledge of consistent and predictable camera movement is leveraged, eliminating the need for motion estimation on a per-frame basis, a computationally intensive task in many conventional methods.^[26] This can be computationally intensive and may not always yield accurate results, especially in cases where noise is prevalent. In contrast to other approaches, this methodology uses the information that the consistent camera movement allows for certain assumptions about the relative positions of frames within a specific time-frame. As a result, motion vectors are calculated based on this known camera movement pattern. These motion vectors describe how each frame within a predefined temporal window should be displaced to align with a central reference frame. By doing so, this significantly reduces the computational burden associated with estimating motion for each frame independently.

These vectors are then used to realign each frame within the temporal window to the coordinate system of the central frame, ensuring correct registration and compensation for the known camera movement. Temporal median filtering is then applied to these realigned frames within the temporal window. This involves pixelwise median calculation for each pixel location, computing the median value of the corresponding pixel across the realigned frames within the temporal window.

Frame reconstruction is then achieved by combining the pixelwise median values obtained in the previous step (see Fig. 2). For example, the pixels in the 100th column of the central frame are median filtered with the 96th column in the previous frame and the 104th column in the next frame if the camera movement is four pixels in the horizontal direction. The colored shapes (circle, triangle, and square) represent the same objects as in Fig. 1, now aligned across frames.

The use of temporal median filtering is considered appropriate, as it effectively reduces random noise by computing the median pixel value across multiple frames. Given the consistent camera movement and the realignment process, the risk of systematic noise patterns, such as lines, overlapping in successive frames is minimized. With no

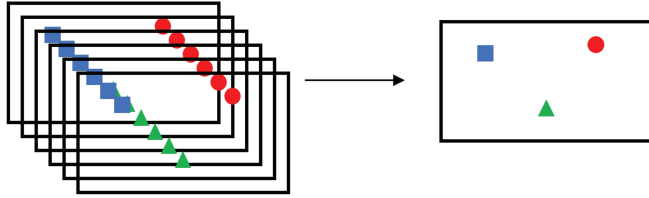


Fig. 2. After correcting for relative movement, the median image can be produced to remove noise.

instances of such overlap observed in the case study data, the median filter was deemed suitable. However, should noise characteristics change, or for other applications, alternative filters, such as bilateral or nonlocal means, may be explored for potential additional benefits.

III.A.2. Start and Stop Detection

In addition to the denoising methodology described previously, it is important to consider the specific context in which this technique is applied. In the following case study, the videos that were analyzed contained a dynamic scenario where the camera underwent three distinct phases. Initially, the camera was stationary, which was followed by the visual inspection of the asset, and finally, another stationary phase at the end of the video. Failure to account for these distinct phases can lead to inaccuracies in the denoising process. The proposed video denoising approach contains a start and stop detection mechanism.

This critical step involves identifying the transitional points within the video where the camera transitions from a stationary state to an inspection phase and then back to stationary. Accurate detection of these phases is crucial to ensuring the precise application of the denoising technique.

This is achieved by calculating the mean-squared error (MSE) between consecutive frames. The MSE is a measure of the average squared differences between corresponding pixels of two consecutive frames. Higher MSE values indicate greater differences between frames, which signifies camera movement or significant changes in the scene. By analyzing variations in MSE values over the sequence of frames, the algorithm identifies points where these variations exceed certain thresholds, indicating transitions from stationary to inspection phases and vice versa. These identified transitional points serve as crucial markers for accurately applying the denoising technique at appropriate stages of the video. Table I gives the steps of the video denoising algorithm, and Fig. 3 provides the algorithm flowchart.

In the initial stationary phase, where the absence of camera movement allows for stable conditions, denoising

is achieved using a temporal window of consistent width equal to the duration of the stationary period. This conventional time-filtering method is straightforward and relies on temporal averaging to reduce noise effectively.

Upon detection of the transition from a stationary to an inspection state, the known movement vector of the camera is used following the video denoising process, as described in the previous section. Upon the camera returning to a stationary state, the temporal window width can once again be adjusted to match the dimensions of the stationary region. This approach allows for the stationary frames and the frames from the inspection phase to be processed differently, and allows for a more adaptive and context-aware denoising strategy that addresses the specific characteristics of each phase.

III.B. Metrics for Performance Evaluation

The performance of the denoising techniques was evaluated using a variety of metrics to assess the quality and fidelity of the denoised videos compared to the original noisy videos. The five metrics used were the peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), blind/referenceless image spatial quality evaluator (BRISQUE),^[27] natural image quality evaluator (NIQE),^[28] and perception-based image quality evaluator (PIQE).^[29]

A widely used metric for assessing the quality of denoised videos, PSNR measures the quality of the denoised video by comparing it to the original noisy video, providing a numerical value that represents the level of noise reduction. SSIM compares the structural similarity between the denoised and original videos, considering brightness, contrast, and structure. BRISQUE quantifies image quality by modeling the statistical properties of natural images, offering insights into perceived image quality. NIQE measures image quality by analyzing various statistics and texture features, providing a comprehensive assessment of image fidelity. PIQE calculates a quality score that indicates how an image is likely to be perceived by human observers. It offers valuable insights into the perceptual quality of the denoised videos.

IV. CASE STUDY: CALANDRIA TUBESHEET BORE INSPECTION VIDEOS

IV.A. Data

The data used for this case study were a specific subset of the data gathered by Bruce Power (Ontario,

TABLE I
Steps of the Video Denoising Algorithm

Step	Description
1	Extract frames from the video at the desired frame rate, resulting in a sequence of frames.
2	Select a set of frames before and after the frame under analysis, forming a temporal window.
3	Calculate motion vectors based on the known camera movement pattern, describing how each frame within the temporal window should be displaced to align with a central reference frame.
4	Realign each frame within the temporal window to the coordinate system of the central frame using the calculated motion vectors.
5	Apply temporal median filtering to the realigned frames within the temporal window, $\text{Median}(x,y) = \text{median}(P^1_{(x,y)}, P^2_{(x,y)}, \dots, P^N_{(x,y)}) \cdot (1)$
6	Reconstruct the frames by combining the pixelwise median values obtained in the previous step.

Canada) during the inspection of Calandria tube sheet bores (CTSBS) during their major component replacement program,^[30] which included over 1000 individual 360-deg scans (videos) of the inside of the CTSB surface. These scans were conducted using Ahlberg PTZ620 high-definition radiation-resistant cameras.^[31]

Within these inspection videos, there was a significant amount of noise, primarily caused by the high-dose radiological environment. Additionally, other sources of noise, such as camera motion, fluctuations in lighting conditions, and equipment vibrations, also played a role in degrading the inspection video quality. Python was utilized for the analysis and processing of these data, enabling the implementation and evaluation of the proposed video denoising techniques.

The scans were all captured with a constant and predictable camera movement; therefore, this provided a rich database of inspection videos that contained significant noise and the prior information required for the proposed technique to be effectively tested. In addition to the noise, it should also be noted that there were saturated regions on the left and right regions of the images (see Fig. 4).

IV.B. Denoising Analysis

IV.B.1. Quantitative Comparison

In the comparison of the proposed denoising approach, a comprehensive analysis was conducted by first testing three widely recognized and established video denoising techniques: median filtering, bilateral filtering, and FNLM. These three approaches were selected as reference points to assess the effectiveness and performance of the proposed denoising approach.

Comparison with a state-of-the-art deep learning-based video denoising model (RVRT) was also performed.

The initial comparison was performed by evaluating the PSNR and SSIM metrics for each denoising technique. In addition to this, three no-reference image-quality metrics were calculated for each denoising technique. These metrics were BRISQUE, NIQE, and PIQE. Table II presents the results for each metric relating to the average results for several frames of five randomly selected videos from the data set.

The results of the comparison revealed that while the visual quality of the videos produced by the proposed denoising technique was better than the traditional methods, the quantitative results of the PSNR and SSIM scores showed a worse performance. This is a common issue when comparing denoising techniques^[32] with objective metrics, such as those described previously, and will not necessarily align with the engineer's perception of visual quality.

When compared to the deep learning-based RVRT model, the performance of the proposed denoising technique was similar. While RVRT demonstrated impressive results, the proposed technique showed comparable performance in terms of the metrics calculated. In fact, these two were consistently the top two denoising approaches out of all the approaches tested. Again, as with the traditional approaches, when comparing visual quality to an engineer, the proposed approach outperformed the state-of-the-art deep learning approach, as discussed in the following.

IV.B.2. Qualitative Comparison

In addition to quantitative analysis, a series of in-depth qualitative evaluations were undertaken to determine the visual quality of the denoised videos. These

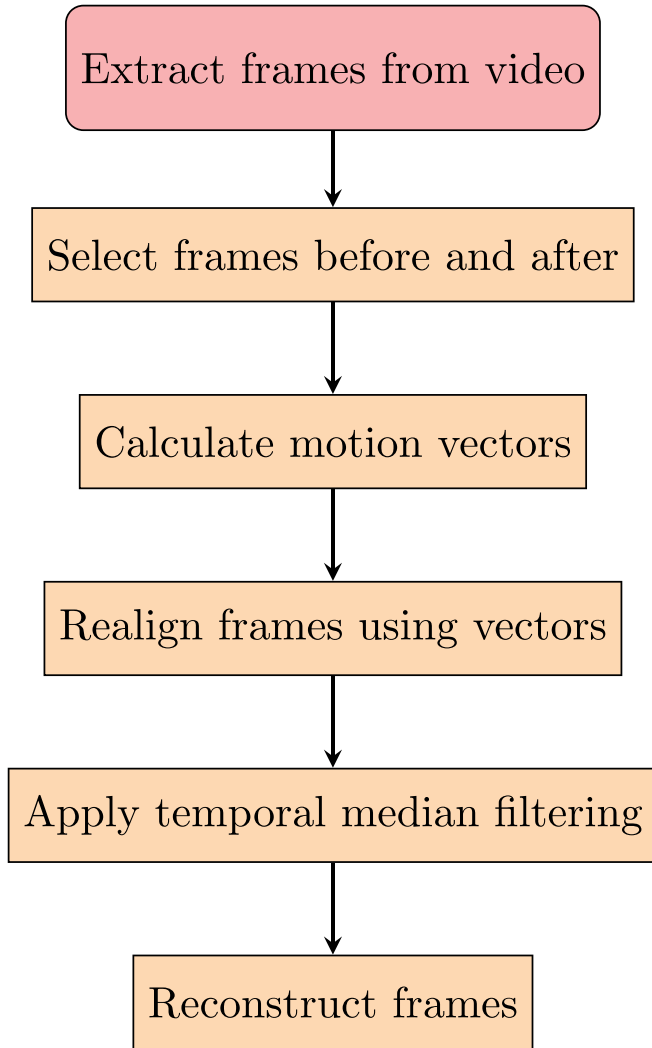


Fig. 3. Video denoising algorithm flowchart.

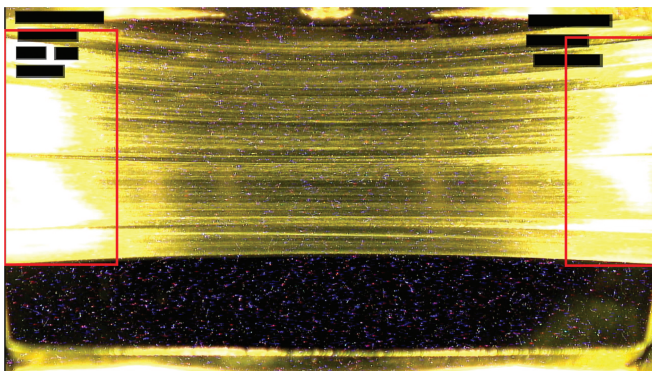


Fig. 4. Representative frame gathered from the CTSCB visual inspection. The red boxes highlight the regions at the left and right ends of the image where the white pixels correspond to saturated pixels.

evaluations were designed to capture aspects of the denoising process that might not be fully represented through numerical metrics alone. Expert and nonexpert evaluators reviewed the denoised videos from each method and provided subjective feedback regarding artifacts, blurriness, and overall visual appeal. This qualitative analysis provided valuable insights into the perceptual quality of the denoised videos, complementing the quantitative metrics obtained from the PSNR, SSIM, BRISQUE, PIQE, and NIQE evaluations. The images used in the qualitative analysis are shown in Fig. 5.

The results of the analysis are shown in Table III. While the quantitative metrics did not fully capture these qualitative improvements, the subjective observations from the analysis highlighted the practical significance of the proposed method. From the study, there were 10 evaluators, and for all but three cases, the evaluators selected the proposed approach as the best denoising approach. In this nuclear inspection application, the engineers were looking to precisely identify features on the CTSCB surface. The removal of the majority of the noise, even with the added drawback of adding a small amount of blurring, still allowed for a more effective interpretation of critical features on the CTSCB surface during the inspection.

IV.C. Noise Distribution Analysis

Following the quantitative and qualitative analyses of the denoising technique, an investigation into the distribution of noise within the videoframes was undertaken. The purpose of this investigation was to understand the spatial patterns of noise across the entire video, providing insights into the behavior of noise in different regions of the frames. Fig. 6a shows the binary classification of the noisy pixels in a single frame, where noisy pixels are white and the rest of the pixels are black. This was repeated for the entire video, and a histogram of the noise for both the x and y axes are shown in Fig. 6b.

The histogram for the x -axis showed a skew toward higher noise distribution in the central region of the frames, indicating a concentration of noise in the middle area. This concentration was due to consistently saturated pixels on both the left and right sides of the frame, as observed in the regions of Fig. 6a where no noisy pixels were detected. The y -axis histogram, however, exhibited a dent in the central region, suggesting a lower concentration of noise in the middle portion of the frames, with relatively higher noise toward the top and bottom edges. This pattern was consistent with the

TABLE II
Quantitative Results*

Video Name	Denoising Approach	PSNR	SSIM	BRISQUE	PIQE	NIQE
Video 0	Original			38.73	47.88	5.05
Video 0	Bilateral	33.24	0.80	48.87	69.72	33.24
Video 0	FNLM	35.06	0.88	38.66	74.64	35.06
Video 0	Median	33.51	0.72	46.05	64.48	33.51
Video 0	Proposed	34.26	0.75	33.63	43.57	34.26
Video 0	RVRT	32.72	0.70	22.25	60.63	32.72
Video 1	Original			24.94	30.40	4.26
Video 1	Bilateral	33.89	0.81	47.33	74.22	33.89
Video 1	FNLM	34.11	0.82	40.67	76.02	34.11
Video 1	Median	34.26	0.76	41.28	64.71	34.26
Video 1	Proposed	35.28	0.79	18.40	33.60	35.28
Video 1	RVRT	33.32	0.71	30.68	62.59	33.32
Video 2	Original			21.34	36.01	4.64
Video 2	Bilateral	33.05	0.79	43.07	72.44	33.05
Video 2	FNLM	33.39	0.80	35.24	76.94	33.39
Video 2	Median	33.47	0.74	43.79	67.64	33.47
Video 2	Proposed	34.68	0.80	33.11	37.56	34.68
Video 2	RVRT	32.92	0.73	38.00	65.80	32.92
Video 3	Original			37.61	47.07	5.11
Video 3	Bilateral	34.09	0.81	48.08	75.87	34.09
Video 3	FNLM	35.67	0.87	43.26	76.22	35.67
Video 3	Median	34.55	0.72	48.07	63.40	34.55
Video 3	Proposed	32.81	0.57	36.46	36.52	32.81
Video 3	RVRT	32.97	0.66	34.41	50.19	32.97
Video 4	Original			27.08	33.44	5.40
Video 4	Bilateral	32.09	0.80	39.53	71.08	32.09
Video 4	FNLM	33.07	0.84	39.40	70.40	33.07
Video 4	Median	32.24	0.70	42.26	64.73	32.24
Video 4	Proposed	32.97	0.73	40.52	53.05	32.97
Video 4	RVRT	32.17	0.75	31.83	60.00	32.17

*For both the PSNR and SSIM, higher score values reflect a better image quality. For BRISQUE, PIQE, and NIQE, lower score values reflect a better perceptual quality of the image.

expected behavior based on the frame's characteristics and orientation.

The results showed that the noise was uniformly distributed across the frame for the entire duration of the video, except in the expected saturated regions. This uniform distribution indicated a consistent noise pattern, confirming that the proposed approach will produce a consistent level of quality throughout the video.

V. CONCLUSIONS AND FUTURE WORK

In conclusion, this paper presented an approach to denoising inspection videos, especially in contexts where prior knowledge of camera movement is available. Leveraging movement vectors, the technique showed

great potential in enhancing content quality, particularly in scenarios featuring both stationary and moving camera footage. However, it is important to recognize the inherent limitation of this method, which relies heavily on prior information about camera movement.

This approach streamlines the denoising process by capitalizing on the anticipated camera movement, thereby enhancing efficiency and accuracy. This allows for a set of frames to be processed that is already aligned based on prior knowledge, making subsequent denoising and filtering steps more effective in producing high-quality, noise-free videoframes. This approach is particularly advantageous in scenarios where consistent camera movement is a defining characteristic, as it simplifies the denoising workflow and yields superior results compared to traditional, frame-by-frame motion estimation methods. It is

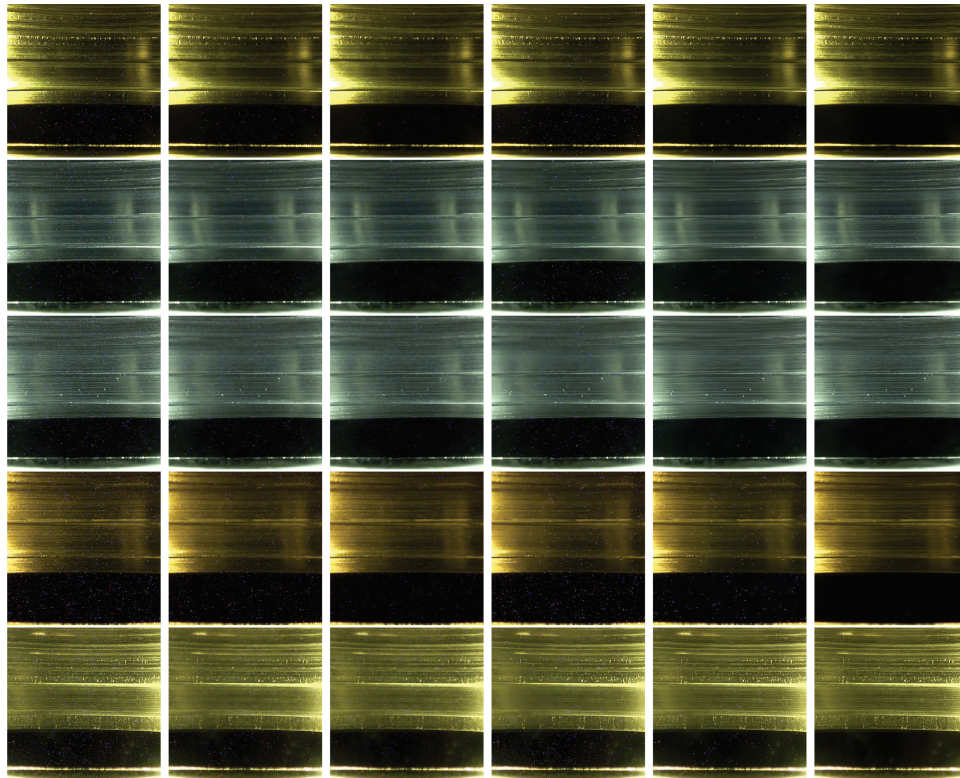


Fig. 5. (left to right) Original frame, bilateral filtering, median filtering, FNLM, RVRT, and the proposed approach for five different videos (each row). The quality difference in the images can be appreciated by zooming in.

TABLE III
Qualitative Results*

Name	Original	Bilateral	FNLM	Median	Proposed	RVRT
Video 0	0	0	0	0	10	0
Video 1	0	0	0	0	8	2
Video 2	0	0	0	0	9	1
Video 3	0	0	0	0	10	0
Video 4	0	0	0	0	10	0

*Evaluator preferences for each denoising method based on visual quality, where higher scores indicate a greater number of evaluators favored that method.

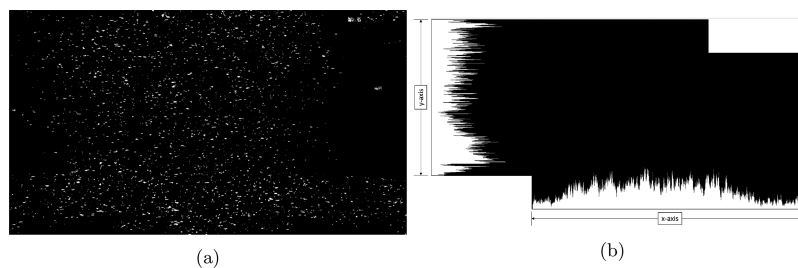


Fig. 6. (a) Binary noise classification and (b) histogram of noise pixels for both the x -axis and y -axis.

also worth noting that, while the proposed method introduces a slight amount of blurring, this trade-off appears acceptable in the context of the CTSB inspection.

The subtle blurring, as indicated by the qualitative analysis, did not compromise the overall interpretability of crucial features. Instead, it served as a compromise to achieve a cleaner, less noisy representation that aided in the identification of essential details on the CTSB surface. In addition, our approach intelligently adapted to the varying dynamics of the video, employing different denoising strategies for stationary and moving phases. By incorporating a start and stop detection process, we not only enhanced the accuracy of denoising, but also ensured that the visual quality of the video was consistently improved across all phases of the recording, delivering optimal results tailored to the specific context of the video footage.

Future work in this field should focus on two primary areas: first, the generalization of camera movement detection to make it less dependent on a priori knowledge, and second, the integration of start/stop detection mechanisms to dynamically assess the need for filter application based on camera movement presence or absence.







Acknowledgments

The authors would like to thank Bruce Power (Ontario, Canada) for providing the video footage and related information used in this work.

Disclosure Statement

No potential conflict of interest was reported by the authors.

ORCID

Andrew Young  <http://orcid.org/0000-0001-6338-6631>
 Zhouxiang Fei  <http://orcid.org/0000-0002-5003-3949>
 Jaime Zabalza  <http://orcid.org/0000-0002-0634-1725>
 Graeme M. West  <http://orcid.org/0000-0003-0884-6070>
 Paul Murray  <http://orcid.org/0000-0002-6980-9276>
 Stephen D. J. McArthur  <http://orcid.org/0000-0003-1312-8874>

References

1. Y. XU, Y. CAI, and L. SONG, “Condition Assessment of Nuclear Power Plant Equipment Based on Machine

- Learning Methods: A Review,” *Nucl. Technol.*, **209**, 7, 929 (2023); <https://doi.org/10.1080/00295450.2023.2169042>.
2. Z. FEI et al., “Enhanced Video-Level Anomaly Feature Detection for Nuclear Power Plant Component Inspections Using the Latency Mechanism,” presented at the 13th Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies (2023).
3. R. BOGUE, “Robots in the Nuclear Industry: A Review of Technologies and Applications,” *Industrial Robot*, **38**, 2, 113 (2011); <https://doi.org/10.1108/01439911111106327>.
4. M. G. DEVEREUX, P. MURRAY, and G. WEST, “Automated Object Detection for Visual Inspection of Nuclear Reactor Cores,” *Nucl. Technol.*, **208**, 1, 115 (2022); <https://doi.org/10.1080/00295450.2020.1863067>.
5. C. WANG et al., “Nuclear Radiation Degradation Study on HD Camera Based on CMOS Image Sensor at Different Dose Rates,” *Sensors*, **18**, 2, 514 (2018); <https://doi.org/10.3390/s18020514>.
6. S. XU et al., “Parallel Processing of Radiation Measurements and Radiation Video Optimization,” *Opt. Express*, **30**, 26, 46870 (2022); <https://doi.org/10.1364/OE.476102>.
7. M. C. SHEEBA and D. C. SELDEV CHRISTOPHER, “A Review on Video Denoising Methods,” *Proc. 2019 Int. Conf. on Recent Advances in Energy-efficient Computing and Communication (ICRAECC)*, pp. 1–6 (2019); <https://vciba.springeropen.com/articles/10.1186/s42492-019-0016-7> (accessed Sep. 23, 2024).
8. L. FAN et al., “Brief Review of Image Denoising Techniques,” *Vis. Comput. Ind. Biomed. Art*, **2**, 7 (2019); <https://vciba.springeropen.com/articles/10.1186/s42492-019-0016-7> (accessed Sep. 23, 2024).
9. I. PITAS and A. N. VENETSANOPOULOS, *Nonlinear Digital Filters: Principles and Applications*, Vol. 84, Springer Science & Business Media (2013).
10. Y. ZHU and C. HUANG, “An Improved Median Filtering Algorithm for Image Noise Reduction,” *Physics Procedia*, **25**, 609 (2012); <https://doi.org/10.1016/j.phpro.2012.03.133>.
11. X. LI, M. K. NG, and X. YUAN, “Median Filtering-Based Methods for Static Background Extraction from Surveillance Video,” *Numer. Linear Algebra Appl.*, **22**, 5, 845 (2015); <https://doi.org/10.1002/nla.1981>.
12. S. J. HUANG, “Adaptive Noise Reduction and Image Sharpening for Digital Video Compression,” *Proc. 1997 IEEE Int. Conf. on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, 4, 3142 (1997); <https://doi.org/10.1109/ICSMC.1997.633077>.
13. W. LIANG et al., “A Fast Defogging Image Recognition Algorithm Based on Bilateral Hybrid Filtering,” *ACM Trans. Multimedia Comput. Commun. Appl.*, **17**, 2, 1 (2021); <https://doi.org/10.1145/3391297>.

14. C. TOMASI and R. MANDUCHI, “Bilateral Filtering for Gray and Color Images,” *Proc. 6th Int. Conf. on Computer Vision*, IEEE Cat. No. 98CH36271, p. 839 (1998); <https://doi.org/10.1109/ICCV.1998.710815>.
15. L. SI et al., “Image Enhancement for Surveillance Video of Coal Mining Face Based on Single-Scale Retinex Algorithm Combined with Bilateral Filtering,” *Symmetry*, **9**, 6, 93 (2017); <https://doi.org/10.3390/sym9060093>.
16. D. BHONSLE, V. CHANDRA, and G. SINHA, “Medical Image Denoising Using Bilateral Filter,” *Int. J. Image Graph. Signal Process.*, **4**, 6, 36 (2012); <https://doi.org/10.5815/ijigsp.2012.06.06>.
17. V. KARNATI, M. ULIYAR, and S. DEY, “Fast Non-local Algorithm for Image Denoising,” *Proc. 2009 16th IEEE Int. Conf. on Image Processing (ICIP)*, p. 3873 (2009); <https://doi.org/10.1109/ICIP.2009.5414044>.
18. Z. CONG, Z. GAO, and X. ZHANG, “A Practical Video Denoising Method Based on Hierarchical Motion Estimation,” *Proc. 2013 IEEE Int. Symp. on Broadband Multimedia Systems and Broadcasting (BMSB)*, p. 1 (2013); <https://doi.org/10.1109/BMSB.2013.6621684>.
19. S.-H. KANG and J.-Y. KIM, “Application of Fast Non-local Means Algorithm for Noise Reduction Using Separable Color Channels in Light Microscopy Images,” *Int. J. Environ. Res. Public Health*, **18**, 6, 2903 (2021); <https://doi.org/10.3390/ijerph18062903>.
20. P. COUPÉ, P. YGER, and C. BARILLOT, “Fast Non Local Means Denoising for 3D MR Images,” *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2006*, R. LARSEN, M. NIELSEN, and J. SPORRING, Eds. pp. 33–40, Springer, Berlin Heidelberg (2006).
21. J. LIANG et al., “Recurrent Video Restoration Transformer with Guided Deformable Attention,” arXiv:2206.02146 (2022).
22. J. LIANG et al., “Recurrent Video Restoration Transformer with Guided Deformable Attention,” *Adv. Neural Inf. Process. Syst.*, **35**, 378 (2022).
23. S. ZHU and K.-K. MA, “A New Diamond Search Algorithm for Fast Block-Matching Motion Estimation,” *IEEE Trans. Image Process.*, **9**, 2, 287 (2000); <https://doi.org/10.1109/83.821744>.
24. M. ZHAI et al., “Optical Flow and Scene Flow Estimation: A Survey,” *Pattern Recognit.*, **114**, 107861 (2021); <https://doi.org/10.1016/j.patcog.2021.107861>.
25. H. LIU and H. WANG, “A Review on Vision-Based Motion Estimation,” arXiv preprint arXiv:2407.14478 (2024).
26. S. SHIBA et al., “Secrets of Event-Based Optical Flow, Depth and Ego-motion Estimation by Contrast Maximization,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2024); <https://doi.org/10.1109/TPAMI.2024.3396116>.
27. A. MITTAL, A. K. MOORTHY, and A. C. BOVIK, “No-Reference Image Quality Assessment in the Spatial Domain,” *IEEE Trans. Image Process.*, **21**, 12, 4695 (2012); <https://doi.org/10.1109/TIP.2012.2214050>.
28. A. MITTAL, R. SOUNDARARAJAN, and A. C. BOVIK, “Making a ‘Completely Blind’ Image Quality Analyzer,” *IEEE Signal Process. Lett.*, **20**, 3, 209 (2013); <https://doi.org/10.1109/LSP.2012.2227726>.
29. N. VENKATANATH et al., “Blind Image Quality Evaluation Using Perception Based Features,” *2015 Twenty First National Conference on Communications (NCC)*, pp. 1–6 (2015); <https://doi.org/10.1109/NCC.2015.7084843>.
30. H. BARPUGGA, R. CLUETT, and B. SINGH, “Bruce Power Major Component Replacement Automated Inspection and Calandria Tube Install Innovation,” presented at the 42nd Annual CNS Conf. And 47th CNS/CAN Student Conference (2023).
31. “PTZ620—Nuclear Visual Inspection,” Ahlberg Cameras; <https://www.ahlbergcameras.com/products/cameras/ptz620/> (accessed Sep. 23, 2024).
32. D. S. TURAGA, Y. CHEN, and J. CAVIEDES, “No Reference PSNR Estimation for Compressed Pictures,” *Signal Process. Image Commun.*, **19**, 2, 173 (2004); <https://doi.org/10.1016/j.image.2003.09.001>.