Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 45 No. 7
July 2024

Enhancing CCTV Video Quality for Improved Segmentation and Detection of Suspicious Human Activities in Crowd and Non-Crowded Public Spaces: A Pre-Processing Approach

Ahmed R. A. Shamsan[*1], Suresha M[2], S. P. Raghavendra[3], Abdullah Y. Muaad[4], Mohammed A. S. Al-Mohammadi[5]

[1]Research Scholar, Department of Computer Science & M.C.A., Kuvempu University, Shimoga, India.
[2]Associate Professor, Department of Computer Science & M.C.A., Kuvempu University, Shimoga, India.
[3]Assistant Professor, Department of M.C.A., JNNCE, Shimoga, India.
[4]Research Scholar, Department of Of Computer Science, Mysore University, Mysore India
[5]Research Scholar, Department of Computer Science & M.C.A., Kuvempu University, Shimoga, India.

Abstract : Recently, reliance on surveillance systems to detect Suspicious Human Activities (SHA) in public and crowded places has increased. Despite the significant progress in surveillance cameras, climatic conditions and low light still challenge the quality of videos captured by these cameras. Moreover, the improvement methods vary with the diversity of challenges, such as low light and climatic conditions.
Therefore, researchers and developers resort to applying initial pre-processing for these videos to enhance the quality and contribute to their analysis. In addition, due to the lack of SHA datasets, in this work, we first collect a unique data set for SHA, which indicates the existence of a fight between a group of individuals in public places. Secondly, we propose a model to enhance the quality of videos using different techniques such as Adaptive Histogram, Gaussian Blur, Grayscale, and Median Filter. These techniques are considered the initial processing stage and prepare the dataset for detecting SHA in public and crowded places when the number of people is more than eight. In the future, we plan to apply segmentation techniques that enhance the performance of SHA detection.

Keywords: Adaptive Histogram, Gaussian Blur, Video Enhancement, Suspicious Activities, Crowd, and Public Areas.

1. Introduction

Surveillance cameras, proposed in 1870, became essential after the introduction of closed-circuit television. The growing use of video in various domains, such as social security in crowds and public areas, necessitated quality improvements. With artificial intelligence, real-time video analysis became possible [1]. Figure.1 illustrates the primary obstacles that surveillance systems encounter.

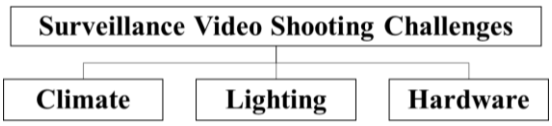

Figure 1 **Different Challenges Facing Video Surveillance**

Surveillance systems face environmental and hardware challenges that affect video quality, including weather conditions and lighting issues. Modern computer vision algorithms can address these issues [2], while video enhancement techniques like median, sharpening, and Gaussian filters address noise reduction and low-light issues. Machine learning and deep learning techniques use large-scale training sets and advanced neural networks to learn image and video degradation processes[3].

One of the significant challenges is the lack of available datasets on suspicious human activities, especially fighting between a group of people. For that reason, we have collected our won dataset, namely(Group Fight Dataset) GFD, from different resources such as the hockey dataset [4], U.C.F. Crime Dataset [5], and YouTube. This data is concerned with suspicious movements in crowded environments. The primary contributions to this work are listed as follows:

- Collecting a new dataset of suspicious human activities in a crowd focusing only on CCTV footage.

- Prepare the new dataset for further processing by removing the audio layer, resizing, and cropping all videos.
- Enhancing videos by using different enhancement techniques.
- Compare the results of original videos and enhancement videos using different evaluation metrics.

The paper's content is structured as follows: Section II presents the related work. Section III details the materials and methods. Section IV delves into experimental analysis, Section V presents the experimental results and discussion, and Section VI addresses limitations and future work. The conclusion is summarized in section VII.

2. Related Work

Computer vision relies on video enhancement for surveillance applications, with traditional techniques like Gaussian filters and sharpening filters being used[6]. However, machine learning and deep learning techniques have become more advanced[7], such as histogram equalization and regression models [8]. Deep learning approaches like convolutional neural networks have shown impressive results in addressing the challenges of video data's increasing complexity and variability[9].

Data matching and histogram shaping are introduced in[10] to enhance CCTV surveillance video images, improving image quality in normal broad daylight conditions. An anisotropic guided filter refines dark channels and transmission, enhancing brightness and contrast while preserving details and minimizing overexposure[11]. A new CLAHE-based method called CLAHE_LUV removes lighting inconsistencies, particularly in low-light and dark frames of video[12]. A network for multimedia surveillance is established using discrete wavelet transform (DWT) and single value analysis (SVD), which combines SVDDWT with averaging and contrast filtering to improve low-light video frames with distortion[13]. CLAHE and Gaussian Filter methods are used to improve CCTV image quality by resizing surveillance camera screenshots to 512X512 and converting them to grayscale.

In real-time monitoring, the suggested system detects suspicious activities with 97.65% accuracy and uses 75% less RAM[14]. Dynamic correction prioritizes questionable pixels, while correction-based frame differential detects suspicious items in consecutive frames in real time. High-resolution capture suspicious frames and tires matter, saving memory by diverting non-suspicious frames[15].

3. Materials and Methods

According to the existing research for enhancing surveillance videos, our method is structured in almost four stages, as depicted in Figure 2. In the following sub-section, we describe each of these stages.

3.1 Video Enhancement Techniques

Surveillance videos often face challenges like inadequate lighting, background noise, and limited object distinction[16] [17]. As illustrated in Figure 2.We have created a new method to improve these issues by converting images into grayscale[18], applying median filtration[19], using a Gaussian filter[20][11], and implementing the limited variability adaptive histogram equation (CLAHE). Grayscale conversion simplifies video frames, enhancing the visibility of details in low-light scenarios[21]. Median filtering reduces noise while preserving edges from sensor imperfections or

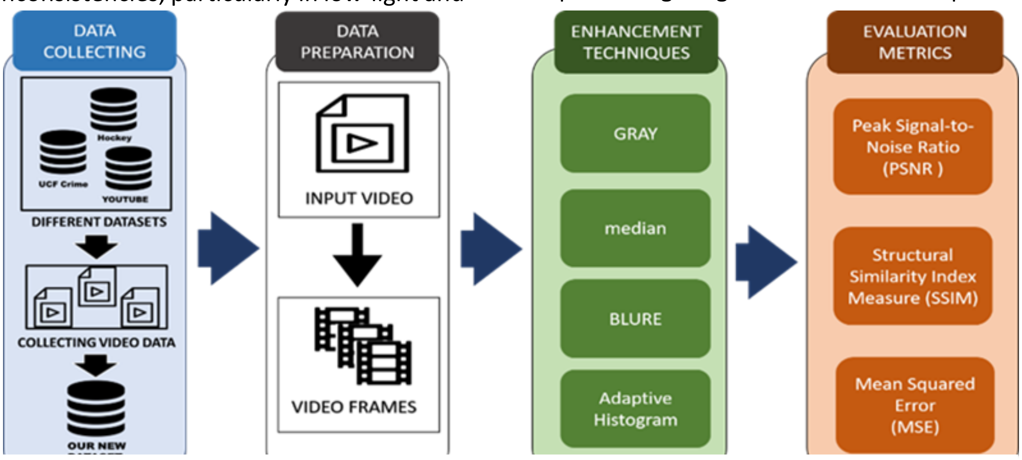

Figure 2 General Architecture of Videos Enhancement

environmental factors[22]. Gaussian blurring smooths out images by applying a Gaussian kernel, reducing high-frequency noise and small-scale details[23]. CLAHE improves the contrast and visibility of details in surveillance videos[24]. The method is applied to our collected dataset described in Section 4.2. Moreover, MSE, PSNR, and SSIM are used to evaluate the quality and accuracy of our enhancement method [25]. as shown in the results and discussion section.

3.2 Data Collection

The lack of an available dataset for video surveillance based on human suspicion is one of the big challenges in video enhancement thus motivating us to collect our dataset. The dataset was collected from different resources such as the UCF-Crime dataset [5] [26], the Hockey dataset [4], and finally, a set of videos from YouTube. This data set concerns suspicious movements that indicate a fight between a large group of people. CCTV cameras capture 90% of the videos collected in our dataset.

3.3 Data Preparation

Video processing has different steps. Starting with preparing videos for further processing which fundamental step. To facilitate video analysis and object recognition, however, there are general initial processing steps such as adjusting brightness, contrast, and digital noise level. In our proposed method, we have used four techniques starting with resizing videos by 320x240, video cutting, and removing the sound layer from each video to reduce the total size of each video in our dataset. Moreover, the Video Frames Extracting Techniques have been used via deep learning-based object detection techniques, cluster-based analysis, the Generalized Gaussian density method, the General-Purpose Graphical Processing Unit, and Histogram difference techniques for extracting essential frames from videos. These techniques extract frames that represent the video's content for object recognition, detection, tracking, and summarization. The choice of technique is determined by the application and the required level of precision and performance[27].

4. Experimental Analysis

We compare the performance of our proposed model to CCTV video enhancement work in this section. A dataset from various sources is used for evaluation. We explain all datasets in Sub-Sect 4.2. The compression task uses three evaluation metrics and four enhancement methods: grayscale, median filter, gaussian blur, and contrast limited adaptive histogram equalization.

4.1. Experimental Environment

Google Colab, Python 3.8, NumPy, pandas, TensorFlow, Keras, Sklearn, and Scikit-learn were used to run the experiments on a system with the specified data. Tesla T4 has 40 multiprocessors, 7.5 capability, and 14 GB memory. Datasets, baseline models, evaluation metrics, and experimental results will be covered in subsequent subsections.

4.2. Datasets

Our datasets for this issue are unidentified. We chose to collect our dataset. We analyze Hockey and UCF-Crime benchmark datasets to find relevant videos. YouTube videos related to our issue are also collected. UCF-Crime Dataset Includes 128 hours of unedited, real-world surveillance footage, totalling 1900 films. Pictures are taken from every tenth frame of every movie to evaluate anomaly detection systems [5][26].and the Hockey Dataset contains NHL game footage. Each fight or non-fight video clip has 50 frames at 720 x 57 pixels.

4.2.1. Collected Dataset Group Fighting Dataset (GFD)

Suspicious human activities are rare in public areas, often leading to violence. Data sets vary in suspicious activities, with surveillance cameras affecting video quality. This study collected a unique dataset of suspicious movements in crowded areas, using UCF-Crime[26], Hockey [4] datasets, and YouTube videos. CCTV cameras captured 90% of the videos, highlighting the importance of monitoring and tracking such activities in public spaces. The dataset aims to provide a unique perspective on fighting between groups in crowded areas. Table 1 describes our dataset.

Table 1 GFD Collected Dataset Description

Source	No. Of Videos	Average Number of Frames			Total Duration by Seconds
		1 To 100	100 To 1000	Above 1000	
Hockey Dataset	61	√			67
UCF-Crime	11		√	√	1492
Youtube	20		√	√	941
Total	92	√	√	√	2500

4.3. Evaluation Metrics

P.S.N.R., S.S.I.M., and M.S.E. are metrics used to evaluate video quality, measuring similarity between original and processed videos. They are used in assessing visual quality of videos, comparing different processing techniques, compression algorithms, or codecs. The goal is to identify methods that produce high-quality videos or preserve details in high-motion scenes. M.S.E. and P.S.N.R., based on statistical processing of pixel difference, do not correlate highly with subjective V.Q.A. scores. S.S.I.M. and MS-SSIM, which consider H.V.S., achieve better correlation with V.Q.A. scores. They compare structural features extracted from videos, not pixel values [28]. P.S.N.R. compares a signal's maximum power against corrupting noise. Higher values indicate greater compressed image or video quality. P.S.N.R. is calculated as Eq. (1).

$$PSNR = 10.log_{10}\left(\frac{MAX_I^2}{MSE}\right) \quad (1)$$

The mean squared error (MSE) is the average squared difference between the original and compressed video frames, and MAX_I^2 is the image's maximum pixel value [29]. The S.S.I.M. (Structural Similarity Index Measure) evaluates the structural similarity between two video frames, considering luminance, contrast, and structure. It's used to evaluate image quality post compression or processing S.S.I.M. is calculated as Eq. (2).

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \quad (2)$$

Where x and y are two images being compared, μ_x and μ_y are their means, σ_x^2 and σ_y^2 are their variances, σ_{xy} is their covariance, and c_1 and c_2 are constants to avoid division by zero [2]. The Mean Squared Error (M.S.E.) [30] is a metric for evaluating image and video quality, calculated by averaging squared differences between original and enhanced images. Lower M.S.E. values indicate better quality, but it's not always a reliable indicator of visual image quality. M.S.E. is calculated as Eq. (3).

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \quad (3)$$

Where $I(i,j)$ represent the pixel value of the original video frame at position (i,j), $K(i,j)$ represent the pixel value of the compressed video frame at position (i,j), and m and n represent the dimensions of the video frames. [31]

In summary, PSNR., SSIM., and MSE. are all metrics used to evaluate the quality of video frames after compression or processing, with higher PSNR. and SSIM. values and lower M.S.E. values indicating better quality.

5. Experimental Results and Discussion

Python has several image processing and computer vision libraries. In this research we have used different library to enhance the quality of videos

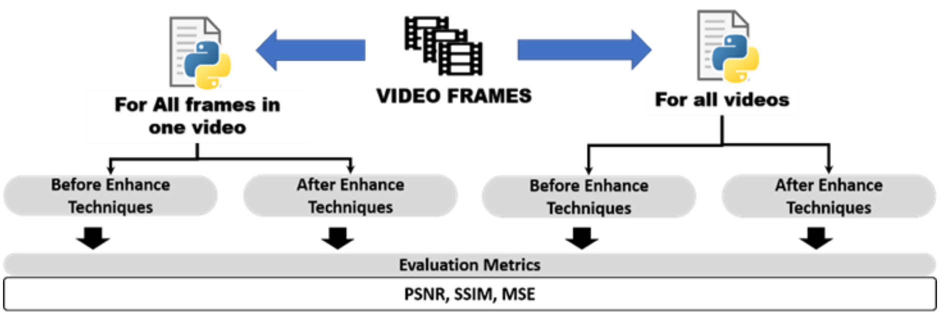

Figure 3 Experiment Flow Steps

such as OpenCV. The proposed model follows different steps as illustrated in Figure 3. In this section, we explore and discuss all results of our proposed model for all datasets with different scenarios.

5.1. Hockey Dataset

5.1.1. Results Before and After Enhancement Using All Frames

Before applying our Enhancement method, we evaluated video frames using PSNR, SSIM, and MSE. With 40 frames per video, we selected specific frames to display their values and compare them with the improvement results. Table 2 and Figure 4. presents the results before and after an experiment on HCKY01 video from a hockey dataset, which extracts video files into frames and applies an enhancement method. The selected frames and outcomes are displayed. The percentage of improvement increases and the margin of error reduces, with PSNR (24.38-25.92), SSIM (69.30-70.41), and MSE (44.11) as the lowest values.

Table 2 Result Before Enhancement for Each Frame in The Video

Video	Frames	Before PSNR	**After PSNR**	Before SSIM	**After SSIM**	Before MSE	**After MSE**
	1	24.7	**25.9**	70	**73.3**	59.8	**54.4**
	5	24.3	**25.7**	73	**76.6**	58.3	**49.9**
	10	24.3	**25.8**	73.1	**76.5**	54.8	**44.1**
	15	24.1	**25.6**	70.2	**74.7**	65.9	**59.1**
HCKY01	20	23.5	**24.9**	70.1	**75.4**	67.1	**60.1**
	25	23.3	**24.6**	70.7	**75.4**	63.7	**58.2**
	30	23.2	**24.5**	71	**75.7**	65.6	**55.6**
	35	23.2	**24.5**	71.3	**75.6**	62.7	**51.3**
	40	23.1	**24.4**	70.9	**75.6**	62.5	**54**

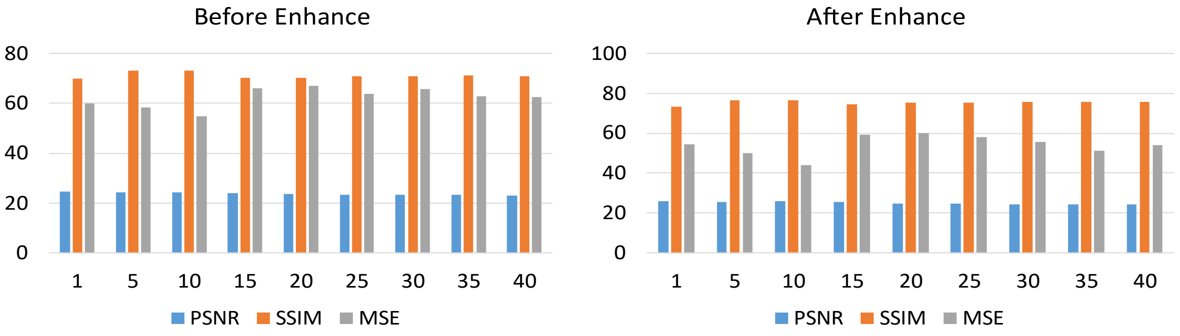

Figure 4 The Results Before and After an Experiment on HCKY01 Video

5.1.2. Results Before and After Enhancement Using Average of All Video Frames

Table 3 displays the average PSNR (22.21), SSIM (70.86), and MSE (65.50) enhancement results for five Hockey dataset video files before and after the proposed method. Figure 5 shows the average enhancement outcomes for video file frames based on PSNR (23.48), SSIM (75.62), and MSE (54.99) ratios. A random selection of five Hockey dataset video files shows a variation in outcomes after improvement.

Table 3. The Average Result for All Frames in Each Video Before and After Enhancement

Video	Frames	Before PSNR	**After PSNR**	Before SSIM	**After SSIM**	Before MSE	**After MSE**
HCKY01	40	23.1	**24.4**	70.9	**75.6**	62.5	**54**
HCKY02	40	22.5	**23.5**	70.9	**75.6**	67	**59.7**
HCKY03	40	20.9	**21.8**	70.9	**75.6**	65.5	**59.2**
HCKY04	40	22.2	**23.2**	70.9	**75.6**	63.5	**59.7**

HCKY05	40	20.6	**21.5**	70.9	**75.6**	66.5	**61.8**

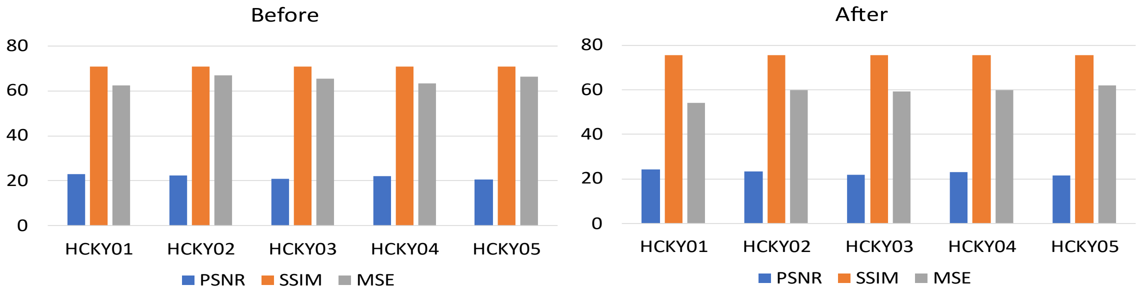

Figure 5 Average Values for All Frames in Each Video Before and after Using the Enhancement Method

5.2. UCF Crime [Fight] Dataset

5.2.1. Results Before and After Enhancement Using All Frames

Table 4 and Figure 6 show PSNR, SSIM, and MSE values for individual frame clusters in each video before and after our enhanced technique. Frames with the highest overall frame rates are chosen. Before improvement, PSNR, SSIM, and MSE ratios are 19.9-20.0, 61.8-64.0, and 73.5-77.3, respectively. The video frame extraction enhanced results are also shown. The enhancement method produces various outputs for each video frame. The results of the specified frames are in Table 3. Also, the improvement rate has grown since Table 3. The margin of error reduced when PSNR was 19–20.0, SSIM was 61.8–65.2, and MSE was 73.2.

Table 4 Result Before and After Enhancement for Each Frame in The Video

Video Name	Frames	Before PSNR	**After PSNR**	Before SSIM	**After SSIM**	Before MSE	**After MSE**
UCF02	1	19.9	**23.5**	63.9	**83.8**	74.5	**67.5**
	500	19.9	**23.7**	64.2	**83.9**	73.2	**66.8**
	1000	19.9	**23.7**	64	**83.8**	73.5	**66.8**
	1500	19.9	**23.6**	61.8	**83**	74.3	**67.8**
	2000	19.9	**23.5**	65.2	**84**	74.2	**65.9**
	2500	19.9	**23.6**	63.4	**83.5**	74.5	**68**
	3000	19.9	**23.6**	62.8	**83.3**	75.9	**70.6**
	3100	19.9	**23.6**	62.6	**82.9**	77.3	**71.9**

5.2.2 Results Before and After Enhancement Using Average of All Videos

Table 5 and Figure 7 present a comparison of the improvement results achieved by implementing the proposed strategy. The table displays the PSNR ratio of 18.5, the SSIM ratio of 67.4, and the MSE ratio of 89.2 for a random sample of five video files from the UCF Crime dataset collection.

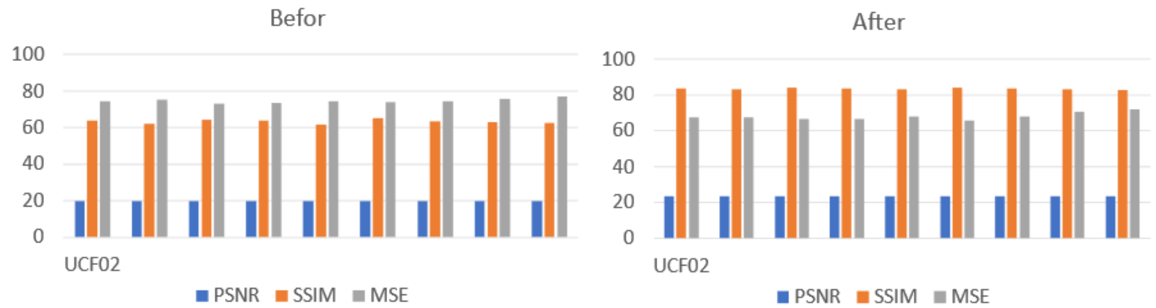

Figure 6 Average Values for All Frames in The Video Before and After Enhancement

Table 5 The Average Result for All Frames in Each Videos Before and After Enhancement

Video	Frames	Before PSNR	**After PSNR**	Before SSIM	**After SSIM**	Before MSE	**After MSE**
UCF01	804	18.9	**20.8**	75.4	**90**	89.2	**90.4**
UCF02	3100	19.9	**23.6**	62.5	**83**	76.4	**70.6**
UCF03	2208	19.4	**20.3**	82.8	**87.9**	80.8	**78.6**
UCF04	3565	22.1	**24.1**	76.9	**75.6**	62.4	**71.2**
UCF05	4456	18.5	**19.7**	67.4	**78.2**	78.2	**71.5**

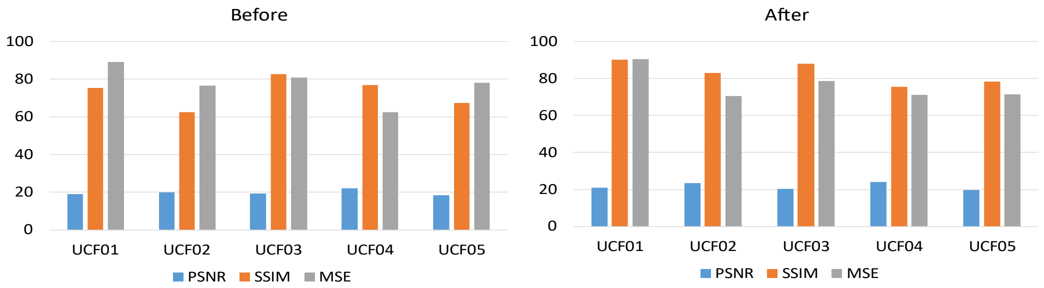

Figure 7 Average Values for All Frames in The Videos Before and After Enhancement

5.3. Our Collection from Youtube

5.3.1. Results Before and after Enhancement Using All Frames

The video file has many frames, so we assigned PSNR, SSIM, and MSE values to groups of frames. Table 6 and Figure 8. show video frames with high PSNR, SSIM, and MSE. Using estimated video frame rates, PSNR is (22.5 to 23.1), SSIM is (61.8 to 63.9), and MSE is (50.9 to 55.1). This data is necessary before optimizing. Moreover, it shows the video file experiment results in frames. Our method yields multiple results in each video frame and shows selected frames and results. The improvement percentage has also increased. PSNR, SSIM, and MSE (48.9) became the lowest error values, decreasing the margin of error.

Table 6 Results Before and After Enhancement for Each Frame in The Video

Video Name	Frames	Before PSNR	**After PSNR**	Before SSIM	**After SSIM**	Before MSE	**After MSE**
	1	22.5	**25**	61.8	**70**	53.5	**48.9**
	20	22.6	**25.1**	62	**69.8**	52.9	**49.7**
	40	22.7	**25.3**	63	**70.1**	52.9	**48.2**
	60	22.9	**25.4**	63.6	**70.3**	52	**46.9**
YTBCOL1	80	23.1	**25.6**	63.3	**70.5**	51.2	**47**
	100	23.1	**25.6**	63.3	**70.3**	50.9	**45.8**
	120	23.1	**25.6**	63.6	**70.2**	53.1	**47.5**
	140	23.1	**25.7**	63.9	**70.4**	51.9	**47.4**
	149	23.1	**25.6**	62.6	**69.3**	55.1	**52.6**

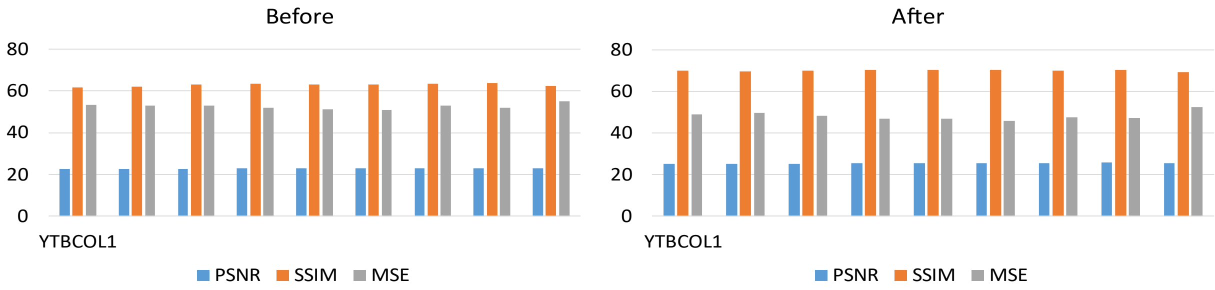

Figure 8 Results Before and After Enhancement for Each Frame in The Video

5.3.2. Results Before and After Enhancement Using Average of All Videos

Table 7 displays the average enhancement outcomes before and after our technique, depending on the average video frame count. PSNR (23.2), SSIM (67.7), and MSE (57.9) ratios and five YouTube video files randomly selected are shown.

In addition, Table 7 also displays the proposed method's average enhancement based on video frame count. PSNR (24.9), SSIM (70.9), and MSE (55.0) ratios and five YouTube video files selected randomly are shown in the table. Enhancement improved duplicate video files in Figure 9.

Table 7 The Average Result for All Frames in Each Video Before and After Enhancement

Video	Frames	Before PSNR	**After PSNR**	Before SSIM	**After SSIM**	Before MSE	**After MSE**
YTBCOL1	149	23.1	**25.6**	62.6	**69.3**	55.1	**52.6**
YTBCOL2	356	21.5	**22.2**	63.4	**70**	59.1	**56.1**
YTBCOL3	432	23.8	**26.9**	64.9	**69.5**	40.9	**38.8**
YTBCOL4	872	22.8	**24.9**	67.7	**70.9**	57.9	**52.8**
YTBCOL5	449	23.2	**25**	66.9	**72.4**	57.4	**55**

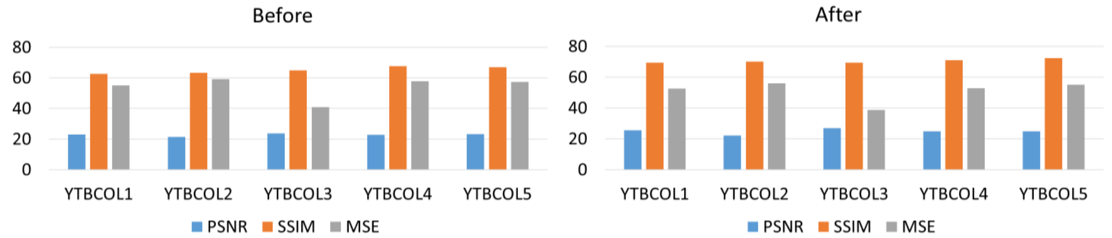

Figure 9 Average Values for All Frames in The Video Before and After Enhancement

5.4. Qualitative Results Compression Between Three Datasets

The GFD dataset was created by combining data from various sources, and the improved technique led to higher video quality. Table 8 displays the average results of multiple data sets, showing an increase in PSNR by 1.618%, an increase in SSIM by 9.224%, and a decrease in MSE by 3.526%. Figure 10 compares the results of 3 datasets, which are the resources of our collated GFD dataset.

Table 8 Comparing The Evaluation Result for Our Dataset Which Collected from Different Resprouts

Video Name	Average Values for All Frames in The Video Before Enhancement			Average Values for All Frames in The Video After Enhancement		
	PSNR	SSIM	MSE	PSNR	SSIM	MSE
UCF Crime	19.5	65.9	67.4	21.1	75.1	65.5
hockey	22.6	72.8	59.2	20.5	64.8	63.1
YTBCOL	21.8	66.2	62.2	23.6	73.1	58.7

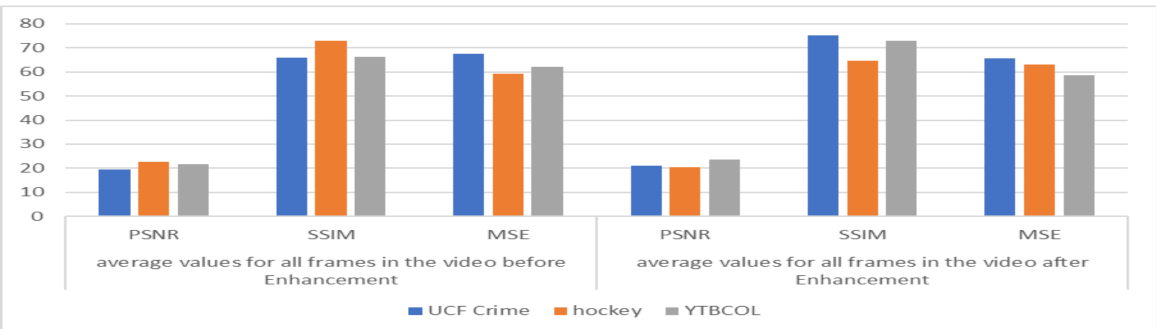

Figure 5 Comparative Between Results Of 3 Datasets Which Are the Resources Of Our Collated Dataset

6. Limitations and Future Work

Video enhancement techniques like Grayscale, Median Filter, Gaussian Blur, and Adaptive Histogram CLAHE can improve CCTV footage for detecting suspicious behavior in crowded areas. However, these methods have limitations such as loss of color information, blurring effects, and challenges in local contrast enhancement. Future research needs to address issues like lighting, weather, and camera views. Deep learning methods like CNN and RNN can enhance video quality and identify suspicious behavior by learning complex patterns. The lack of diversified and large-scale datasets for detecting suspicious activities is a challenge. Future efforts should focus on creating extensive datasets with various situations and environmental variables. More CCTV cameras in public spaces can enhance community safety by increasing surveillance and aiding in video quality and behavioural studies.

7. Conclusion

Identifying suspicious activities of individuals within a video in public and crowded places is challenging. Many reasons, such as climatic conditions and lighting, are the most significant factors that degrade the quality of video surveillance camera capture. All these results in the loss of a great deal of information, which limits the identification of individuals or group activities and thus makes it very difficult to determine their activities. On the other side, due to the lack of datasets, we collected our GFD dataset from three available resources: Hockey Fight, UCF Crime, and YouTube. Furthermore, we apply different techniques to enhance the quality of our collection of videos. The dataset was trained and tested via the proposed model and evaluated using three evaluation metrics, namely PSNR, SSIM, and MSE. The results demonstrate that our proposed model enhances video quality and improves outcomes. This paper enhances video quality and provides researchers with numerous opportunities to tackle various tasks like segmentation and tracking.

Conflicts of Interest

The authors assert that they have no conflict of interest.

Author Contributions

The primary author developed the concept, methodology, and software, validated results, formal analysis, managed resources, curated data, prepared an original draft, and wrote, reviewed, edited, and visualized the paper. The authors' second, third, and fourth oversaw the project.

Refrences

[1] "The Evolution of Video Analytics - Security Sales & Integration." https://www.securitysales.com/surveillance/innovations-in-video-surveillance/ (accessed Feb. 21, 2024).

[2] Y. Yan, "Using the Improved SSD Algorithm to Motion Target Detection and Tracking," *Comput. Intell. Neurosci.*, vol. 2022, 2022.

[3] M. A. GÜNGÖR, "Comparison of Statistical Methods for Obtaining Image from Video Frames Based on Development of Quality Metric," *Sak. Univ. J. Sci.*, vol. 25, no. 3, pp. 741–750, 2021.

[4] E. Bermejo Nievas, O. Deniz Suarez, G. Bueno García, and R. Sukthankar, "Violence detection in video using computer vision techniques," in *International conference on Computer analysis of images and patterns*, 2011, pp. 332–339.

[5] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 6479–6488.

[6] L. Di, H. Long, and J. Liang, "Fabric defect detection based on illumination correction and visual salient features," *Sensors*, vol. 20, no. 18, p. 5147, 2020.

[7] J. Ji *et al.*, "End to End Infrared and Visible Image Fusion With Texture Details and Contrast Information," *IEEE Access*, vol. 10, pp. 92410–92425, 2022.

[8] J. Gajardo, M. Mora, G. Valdés-Nicolao, and M. Carrasco-Benavides, "Burned Area Classification Based on Extreme Learning Machine and Sentinel-2 Images," *Appl. Sci.*, vol. 12, no. 1, p. 9, 2022.

[9] I. Rodríguez-Moreno, J. M. Martínez-Otzeta, I. Goienetxea, I. Rodriguez-Rodriguez, and B. Sierra, "Shedding light on people action recognition in social robotics by means of common spatial patterns," *Sensors*, vol. 20, no. 8, p. 2436, 2020.

[10] N. Chumuang, M. Ketcham, and T.

Yingthawornsuk, “CCTV based surveillance system for railway station security,” in *2018 International Conference on Digital Arts, Media and Technology (ICDAMT)*, 2018, pp. 7–12.

[11] L. Guo, Z. Jia, J. Yang, and N. K. Kasabov, “Detail Preserving Low Illumination Image and Video Enhancement Algorithm Based on Dark Channel Prior,” *Sensors*, vol. 22, no. 1, p. 85, 2022.

[12] D. Abin and S. D. Thepade, “Video Frame Illumination Inconsistency Reduction using CLAHE with Kekre’s LUV Color Space,” *Int. J. Eng. Adv. Technol.*, vol. 9, no. 3, 2020.

[13] N. Halder, D. Roy, and A. Mitra, “Low-light Video Enhancement with SVD-DWT Algorithm for Multimedia Surveillance Network,” *Curr. Trends Signals Process.*, vol. 10, pp. 43–51, 2020.

[14] N. H. M. H. Saputra and M. Furqan, “Application Of Contrast Limited Adaptive Histogram Equalization (Clahe) And Gaussian Filter Methods For Improvement Of Image Quality On Closed Circuit Television (CCTV),” *INFOKUM*, vol. 10, no. 4, pp. 119–127, 2022.

[15] T. Biswas, D. Bhattacharya, and G. Mandal, “Dynamic strategy to use optimum memory space in real-time video surveillance,” *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 3, pp. 2771–2784, 2023.

[16] E. Cabal-Yepez, G. Botella, A. H. Mazinan, and J. P. Papa, “Introduction to the special section on Image and Video Processing,” *Computers & Electrical Engineering*. Elsevier BV, pp. 64–67, 2017.

[17] C. Jittawiriyanukoon and V. Srisarkun, “Evaluation of graphic effects embedded image compression,” *Int. J. Electr. Comput. Eng.*, vol. 10, no. 6, pp. 6606–6617, 2020.

[18] D. Mamchur, J. Peksa, A. Kolodinskis, and M. Zigunovs, “The Use of Terrestrial and Maritime Autonomous Vehicles in Nonintrusive Object Inspection,” *Sensors*, vol. 22, no. 20, p. 7914, 2022.

[19] S. K. Satpathy, S. Panda, K. K. Nagwanshi, S. K. Nayak, and C. Ardil, “Adaptive non-linear filtering technique for image restoration,” *arXiv Prepr. arXiv2204.09302*, 2022.

[20] W. Wang, X. Wu, X. Yuan, and Z. Gao, “An experiment-based review of low-light image enhancement methods,” *Ieee Access*, vol. 8, pp. 87884–87917, 2020.

[21] Z. Li, Z. Jia, J. Yang, and N. Kasabov, “Low illumination video image enhancement,” *IEEE Photonics J.*, vol. 12, no. 4, pp. 1–13, 2020.

[22] X. Dong *et al.*, “Image Noise Removal Method Based on Improved Nonlocal Mean Algorithm,” *Complexity*, vol. 2021, pp. 1–10, Jul. 2021.

[23] Y. Sun *et al.*, “Low-illumination image enhancement algorithm based on improved multi-scale Retinex and ABC algorithm optimization,” *Front. Bioeng. Biotechnol.*, vol. 10, 2022.

[24] A. K. Rai and S. Srivastava, “A Thorough Investigation on Image Forgery Detection.,” *C. Model. Eng. Sci.*, vol. 134, no. 3, 2023.

[25] J. Lian *et al.*, “A Pulse-Number-Adjustable MSPCNN and Its Image Enhancement Application,” *IEEE Access*, vol. 9, pp. 161069–161086, 2021.

[26] W. Ullah, A. Ullah, I. U. Haq, K. Muhammad, M. Sajjad, and S. W. Baik, “CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks,” *Multimed. Tools Appl.*, vol. 80, no. 11, pp. 16979–16995, 2021.

[27] K. Archana and V. K. Prasad, “Object Pre-processing using Motion Stabilization and Key Frame Extraction with Machine Learning Techniques,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 11, 2022.

[28] J. Vlaović, D. Žagar, S. Rimac-Drlje, and M. Vranješ, “Evaluation of objective video quality assessment methods on video sequences with different spatial and temporal activity encoded at different spatial resolutions,” *Int. J. Electr. Comput. Eng. Syst.*, vol. 12, no. 1, pp. 1–9, 2021.

[29] R. G. Deshpande, L. L. Ragha, and S. K. Sharma, “Video quality assessment through psnr estimation for different compression standards,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 11, no. 3, pp. 918–924, 2018.

[30] T. O. Hodson, T. M. Over, and S. S. Foks, “Mean squared error, deconstructed,” *J. Adv. Model. Earth Syst.*, vol. 13, no. 12, p. e2021MS002681, 2021.

[31] M. Mynoddin, M. Kabir, N. Islam, R. Karim, H. Rehana, and S. Asaduzzaman, “De-Speckling of SAR Images with Fuzzy Filters along with Altered Preserved Edge Values,” *J. Comput. Commun.*, vol. 10, no. 03, pp. 10–28, 2022.