ISSN: 0976-9102 (ONLINE)
DOI: 10.21917/ijivp.2022.0389

VIDEO ENHANCEMENT USING DEEP LEARNING

J. Jasmine
Department of Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology, India

Abstract

In recent years, technologies have evolved from simple mobile phones to complicated surveillance monitoring systems capable of capturing and processing video clips. During the video acquisition process, the recorded quality degrades, which is inevitable. Poor illumination and the wrong aperture or shutter speed settings are to blame. This constraint frequently results in photographs with poor quality or images with low contrast and a noisy backdrop. In addition, a video low contrast might be caused by a faulty imaging instrument or a lack of knowledge on the part of the operator. As a result, the available dynamic range is underutilised during video acquisition. As a result, the video finer details are obscured, and the image may appear washed out or strange. Contrast enhancement techniques, which enhance the image visual quality, help to mitigate these issues. The current work aims to address the issues outlined above by employing two distinct approaches. Video compression and contrast enhancement are two different techniques that can be used in conjunction with each other. To increase the quality of videos, the Deep Learning-based Adaptive Cumulative Distribution Based Histogram Enhancement (DLACDHE) technique is applied. The video frames can be more effectively analysed using this hybrid technique. To further reduce noise, the Non-Divisional Median Filter is used. When analysing the sounds, the concept of neighbourhood similarity is employed. Using the proposed DLACDHE approach, the study found that it outperforms other methods. In light of the findings, we may conclude that the proposed strategy provides superior contrast enhancement to the already used approaches.

Keywords:

Automatic Fish Detection, Fish Classification, Fish Species Recognition, Fish Database, Feature Extraction

1. INTRODUCTION

When it comes to style and image processing, digital video is now a key player. It no secret that video upgrades are getting a lot of attention these days in the world of computer vision. Video-controlled methods, such as survey, recognition, scission, and recall, will benefit from better visual representations of retreat. In addition to examining background data, human visual examination is not expensive [1] and is useful for learning object etiquettes.

Investigations, general identification, criminal justice, civil or military videos are just a few of the many uses for digital video. Diverse usages are also possible. Public locations, manufacturing facilities, home research systems, and so on [2] all have a slew of extra video cameras.

In most cases, the quality of a video camera footage is dependent on the surrounding environment. This study unit predicted entire efficiency, illumination, and climatic circumstances, but most of these cameras were not meant to be slow-light. The bad video camera quality makes video unusable in dangerous conditions such as dark nights, soaky rain, essential snow, and fog [3].

Over the course of several decades, digital cameras have greatly improved their capabilities, including resolution and sensitivity. In spite of this, the current digital camera is still unable to capture high-quality film in low light conditions. For high dynamic range footage, these cameras rely on automated exposure adjustment, although the longer exposure period can blur movement [4]–[7].

While low-light image sequences are often low-to-noise quantitatively, this isn't always the case. Due to a reduction in the noise volume on the far side of the signal, conventional noise removal techniques cannot be applied when the lighting is extremely dim [8]. Video illumination that is inexpensive and can be improved rapidly can be a drawback. Though most units appear to be aimed at boosting the quality of low-light video, nearly all of them accept footage shot in extremely dim lighting situations.

As determined by the domain in which the imagery is processed, there are two basic methodologies for image processing: spatial and frequency-based domains. [9]. In this category, spatial domain refers to the image plane. Direct pixel management methods are used in this area.

In frequency-based domain processing approaches, the image spatial frequency spectrum is transformed into a new representation. A variety of method combinations from these two categories can be employed to improve existing procedures, and this can be done in both fields with the same improvement process. With the same image processing, a variety of video enhancing approaches have been proposed. All these solutions, however, do not have a set of universal design criteria for video improvement algorithms. There is also no general theory of video enhancement [10]. Existing video improvement approaches can be divided into two basic categories: those based on the spatial domain and those based on transformations.

Pixels are directly manipulated in spatial video enhancement. For real-time deployments, space-dominant technology has a simple concept and a minimal complexity, making it ideal. There is a general lack of resilience and imperceptibility in the methods used. To represent the frequency analysis of mathematical functions or signals, the term transformed domain video enhancement is often used, as are image transforming coefficients such as transform Fourier transform, DWT, and DCT. The basic idea behind this technique is to boost the video quality by fiddling with transformation coefficients.

2. LITERATURE SURVEY

The MHE (Multi-Histogram Equalization) frame proposed and tested was designed to boost contrast and maintain natural-looking images. Using our methods, we were able to preserve more of the original image luminosity while also improving the contrast of imagegraphs with a more natural appearance. PSNR values generated by images generated by using the MMLSEMHE technique fall within the acceptable range. If the results are good, the MMLSEMHE technique can improve contrast and brightness while maintaining a natural look to the images [12].

To improve contrast, numerous strategies for preservation are reviewed in [11]. The term input refers to the amount of light entering the camera so that the final image appears realistic. The ultimate goal of these methods is to keep the input intact as long as possible.

In model [10], the authors examined various histogram equalisation methods and evaluated the contrast improvement of PSNR with tools such as AMB Error (Absolute Mean Brightness Error) (AMBE). Better suitable for use in consumer electronic devices, Brightness Preserving Dynamic Histogram Equalizing (BPDHE) and Gain-Controllable Clipped Histogram Equalizing (GC-CHE) preserve the original brightness of histograms.

A mixture of Dynamic Echalation and Clipped Equalization of histogram techniques, Quadrant Dynamic with Automatic Plateau Limit Histogram Equalization (QDAPLHE), has been created in [9]. PSNR ranges from 12dB to 35dB, which is critical for consumer electronic items. Qualitative and quantitative analysis yield the best results for these products.

In [8], a Contrast Enhancement (CE) approach is suggested that keeps the histogram localization intact while enhancing contrast. The proposed method, known as histogram-based location preserving CE (HBLPCE), addresses an optimization problem in order to compute an input image histogram of intensity. The goal of the optimization problem is to find a minimal square solution to the locality constraints. According to the findings, HBLPCE performs well when used with images in terms of both computational efficiency and other quality indicators.

The model in [7] proposes a new method of automatic image improvement utilising real-coded PSO, where the number and intensity of edge pixels and the picture entropy measurement are used to define an appropriate fitness function. By raising the total number of pixels on the edges, it was possible to see additional details in the images. The algorithm is tested on four images. In the table below, the results are compared to those obtained using the Genetic Algorithm (GA). PSO-based image enhancement is clearly superior to GAs in terms of image quality and computing efficiency. For example, fine-tuning PSO settings to reduce particle counts and increasing the maximum number of repetitions are two strategies to expand the scope of this picture improvement method based on particle size optimization (PSO).

3. PROPOSED METHOD

There are some rules that must be followed in order to use DL-based image enhancement. The augmented image requires a relatively high intensity of the edges in order for the proposed technique to work well. It follows from this that an image that lacks natural contrast would have a fitness criterion proportional to the number and intensity of pixels in the edges. Based on the standard deviation and the mean value of the pixels, it employs a local enhancement strategy. It doesn't interact with humans in any way. An objective fitness criterion is used, and it based on a clumping factor and the number of edges in the image.

- It uses a local enhancement technique based on the standard deviation and the mean value of the pixels.
- It has no interaction with the humans.
- It uses an objective fitness criterion that is proportional to the number of edges in the image and to a clumping factor of the intensity transformation curve.

To ensure that the histogram of the image has a more uniform distribution, it is necessary to apply a fitness criterion. Fitness function optimization is what the proposed DL goal is all about. To accomplish these goals, the DL focuses on the following:

- Relatively increase the number of pixels on the image edges.
- The general intensity of the edges should be increased.
- Maximize the entropic measure in order to achieve a histogram that closely resembles a uniform distribution.
- Continue when the maximum number of iterations or the minimum error requirement is not met.

The particle optimization using the adaptive cumulative distribution-based histogram enhancement technique (DACDHE) and video quality enhancement procedure has been developed extremely successfully by the proposed algorithm. The Fig.1 depicts the working procedure of DACDHE. A non-divisional median filter is used to remove any unwanted noise from the video frames, and then the images are improved using the DACDHE method to improve their quality. This method uses the histogram contrast enhancement process to boost the overall video quality, making it easier to spot patterns.

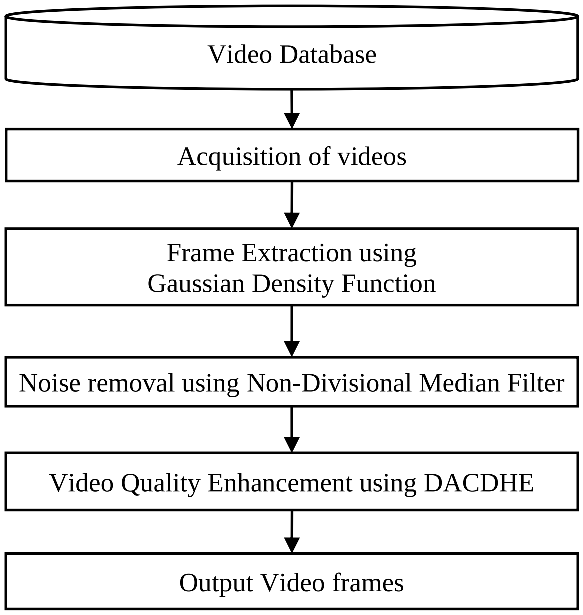

Fig.1. Architecture of the Proposed Study

The suggested system initially accepts videos from the standard video database. A Multi Wavelet with a Gaussian Density Function is used to extract the frames from the video. The sounds are eliminated using a non-divisional median filter after the frame extraction. After the noise is removed, DADCHE is used to optimise the video frame quality using the DL algorithm, which then improves the quality of the remaining frames. The output video files are then obtained from the receiver end once the video frames have been merged.

3.1 DADCHE

Video quality can be improved by using the DACDHE programme. Video frames are loaded, and the histogram values of each frame are examined to enhance contrast. The distribution function Rayleigh is employed to enhance the contrast of the video frame image throughout this procedure.

The centre point value is derived from the particles, and the frames are separated into four subsections once the video frame information is acquired. To make the border of the frames more uniform, the histogram value is estimated from the divided subsection and the frame standard deviation is determined. Entropy values are used to normalise or smooth the entire video frame. After determining the histogram values of a picture, this process is repeated until the best possible image histogram values are achieved.

When the histogram value of the framework is obtained, the cumulative distribution of each frame histogram distribution is checked since it works with multiple random variables in real-time.

As a result of the aforementioned procedure, improved histogram values can be used to get new values for each individual pixel. This process is repeated until the enhanced video frames are reorganised in the video to study the mechanism of pattern recognition.

3.1.1 *Non-Divisional Median Filter*

The next step is to remove noise from the non-divisional median filter that looks at each pixel in the image according to the concept of similarity. In this process, the intensity concept is followed by each pixel to be examined;

$$v(i) = u(i) + n(i) \tag{1}$$

where, $v(i)$ is denoted as the observed frame value, $u(i)$ is true value of each pixel, $n(i)$ is denoted as the noise associated with each pixel

After getting value from video frame, again Gaussian noise is examined based on the assumptions such as noise associated pixels are independent of each other which are identically distributed of Gaussian values of variance σ^2 and mean value (0).

According to the basic assumptions, each pixel similarity values is examined by computing neighborhood pixel values. Similar neighborhood pixels give a great weight, $w(p,q_1)$ and $w(p,q_2)$, while much unlike neighborhoods give a lesser weight such as $w(p,q_3)$. Each pixel p of the NDM filtered image is calculated by using following equation:

$$NL(V,p) = \sum_{q \in V} w(p,q) V(p) \tag{2}$$

where V is defined as the noisy image, and weights $w(p,q)$ meet the subsequent conditions $0 \le w(p,q) \le 1$ and $\sum_q w(p,q)=1$. Along with this, each pixel present in the frame is also called as the weighted average of whole video frame pixels. The frame weighted value has been depending on the similarity of neighborhood pixels of p and q. Considered N_i be the center of the square neighborhood on pixel i with a user-defined Radius Similarity (R_{sim}). Then the weighted sum of square value is estimated using Eq.(7) for computing the neighborhood similarity between two pixels.

$$d(p,q) = \left\| V(N^p) - V(N^q) \right\|^2_{2,F} [1,2] \tag{3}$$

where F is defined as the neighborhood filter employed to the neighborhood's squared difference. The weights is defined as follows

$$w(p,q) = \frac{1}{Z(p)} e^{-\frac{\max\left(d^2 - 2\sigma^2(p,q)\right)}{h}} \tag{4}$$

where σ is as defined as the standard deviation of the noise and $2\sigma^2$ are set to 1. $Z(p)$ is defined as the normalizing constant which is defined as follows

$$Z(p) = \sum_q e^{\frac{d(p,q)}{h}} F[1,2] \tag{5}$$

where h is defined as the weight-decay control parameter. As earlier mentioned, F is known as the neighborhood filter with R_{sim}. The weights of F are computed is as follows

$$F = \frac{1}{R_{sim}} \sum_{i=m}^{R_{sim}} 1/(2 \neq i|1)^2 \tag{6}$$

where m is defined as the distance the weight from the neighborhood filter's center. The F provides higher weight to pixels near the neighborhood center, and provide lower weight to pixels near the neighborhood edge. Based on the above process, each pixel average value is generated for getting the noise free frames. This process is repeated continuously until to eliminate noise from entire video frame. The generated video noise free frames are fed into the next video enhancement process.

3.1.2 *DL based ACDHE:*

DACDHE, which is based on particle optimization, is used as the final stage in this project to increase video quality. First, a noise-free video frame was loaded, and then the histogram value of each frame was evaluated to boost the contrast in this manner. The Rayleigh distribution function is used to assess picture contrast in the video frame.

$$\text{Rayleigh } g = g_{min} + \left[2\alpha^2 \ln\left(\frac{1}{1-p(x)}\right)\right]^{0.5} \tag{7}$$

where, g_{min} is the minimum pixel value and $p(x)$ is the cumulative probability distribution and a *is* the clip limit value. Using the DL method, this histogram estimation procedure can be made to work better. During this process, the size of the video frame determines the number of pixel parcels. This is accomplished by maintaining the location, speed, and inertia values of the particulate matter.

Each frame will be broken into four portions once all the video frames have been taken into account. In order to smooth out the frame borders, a histogram value and a standard frame deviation are derived from a subset of the data. The entropy values are then determined in order to normalise or flatten the entire video framework. When the histogram values are determined, the parameters are consecutively adjusted to improve the histogram frame values.

3.2 DEEP AUTO ENCODER

Due to the several nonlinear processing levels, the deep autoencoder described above is able to derive faithful codes for feature vectors. But the code that is extracted this way is transformation-adjusted. In other words, when the input feature vector is changed, the extracted code will be affected in an unpredictable manner. A predictable change in code can be useful in situations when the observed content is invariant to the underlying alteration.

To alter an audio or visual object, the auto-encoder uses capsules, which are separate networks that work together to identify and extract a single parameterized feature. Both the input vector and the intended output vector are received by transforming auto-encoders; e.g., translation of an image and frequency shift of speech are examples (the latter due to the vocal tract length difference). The worldwide shift is presumptively represented explicitly. Capsule outputs make up the coding layer in the transforming autoencoder.

Capsules learn to extract distinct entities during the training phase so that the final output is as close to the intended aim as possible.

4. EXPERIMENTAL RESULTS

The collected videos are analyzed in terms of frame by frame using proposed system and able to enhance the quality of the video frames with effectively and the performance of the system is examined using following metrics.

4.1 PERFORMANCE METRICS

The following performance metrics are used to evaluate the performance of the proposed system.

4.1.1 Absolute Mean Brightness Error (AMBE):

This metric is used to measure how accurately enhance the video quality when compared to the original video. This AMBE method ensures that justification of video brightness value which is estimated as follows.

$$AMBE = E[Y]\text{-}E[X] \quad (8)$$

where, $E[X]$ is represented as the enhanced video mean contrast, $E[Y]$ is original video mean contrast value. The metric used to examine that; the enhanced video quality should maintain the original video quality.

4.1.2 Peak Signal to Noise Ratio (PSNR):

The PSNR metric used to how effectively enhance the video, mostly the PSNR value must be high which indicates the quality of video output.

$$\text{PSNR}=10\log_{10}[(L\text{-}1)^2/MSE] \quad (9)$$

If the videos are in color image, PSNR value must be calculated separately R, G and B. Finally, the values are aggregated to obtain the PSNR value.

4.1.3 Entropy:

This metric is used to measure richness of the enhanced video quality which is estimated as follows.

$$Entropy[p] = \sum_{k=0}^{L-1} P(k)\log_2 P(k) \quad (20)$$

4.2 DISCUSSION

The first 10 frames of AMBE-based experimental findings are shown in Table.1 based on the experimental results of the suggested approach and the two other conventional ways. For each of the three procedures, the average AMBE value is calculated and presented in Table 1.

Table.1. Absolute Mean Brightness Error (AMBE)

Methods	RSWHE	BPDHE	DACDHE
F1	17.64	17.17	12.50
F2	17.59	17.13	12.02
F3	18.52	17.02	12.12
F4	17.41	16.91	6.56
F5	18.65	16.99	12.05
F6	19.45	17.14	5.30
F7	19.42	17.07	12.09
F8	17.41	17.00	11.80
F9	18.65	16.89	11.88
F10	18.38	17.01	7.72

The proposed approach outperforms the two standard approaches in terms of Absolute Mean Brightness Error by a factor of 43% and 39%, respectively. Video quality, as assessed by PSNR, should be high, even when using the DACDHE approach to achieve a minimum AMBE value. Table 2 displays the PSNR value that was obtained.

Table.2. Peak Signal to Noise Ratio (PSNR)

Methods	RSWHE	BPDHE	DACDHE
F1	24.74	29.34	35.45
F2	24.61	21.12	35.47
F3	24.75	30.71	35.40
F4	24.63	32.34	36.14
F5	24.62	32.00	35.40
F6	24.61	34.11	36.17
F7	24.61	34.14	35.40
F8	26.19	25.63	35.37
F9	24.74	28.42	35.40
F10	26.61	27.43	36.19

Because it uses a higher PSNR value, as shown in Table.2, the DACDHE approach is better at improving video frame quality than other methods. As a result, the PSNR value and video quality obtained by the DACDHE approach were both high. An entropy metric is used to ensure that the upgraded video contains the same amount of information as the original video. Entropy is calculated as indicated in Table.3. When using the DACDHE approach, image quality can be improved by 29% over RSWHE results and 17 percent over BPDHE results.

Table.3. Entropy

Methods	RSWHE	BPDHE	DACDHE
F1	7.83	7.05	9.55
F2	7.79	8.05	9.51
F3	7.07	8.03	9.55

F4	7.11	8.02	9.47
F5	7.14	8.04	9.49
F6	6.86	8.03	9.49
F7	6.91	7.82	9.55
F8	6.92	7.00	9.54
F9	6.90	7.03	9.58
F10	6.69	7.02	9.50

Thus, the new DACDHE is able to efficiently preserve the quality of the video. When compared to the performance of RSWHE and BPDHE approaches, the improvement percentages are 25.24% and 20.08%, respectively. Effectively, the DACDHE resulted in a significant improvement in the quality of the video.

5. CONCLUSION

Using PACDHE to improve video quality is the focus of this paper. A minimum of 17% and a maximum of 29% improvement in video quality is observed when compared to conventional methods. Contrast enhancement and histogram equalisation are merged into a single function that transforms the original image into a high-quality one.

REFERENCES

[1] Xiaopei Liu, Zhaoyang Lu, Jing Li and Wei Jiang, "Detection and Segmentation Text from Natural Scene Images Based on Graph Model", *WSEAS Transactions on Signal Processing*, Vol. 10, No. 1, pp. 124-135, 2014.

[2] Monika Xess and S. Akila Agnes, "Survey on Clustering Based Color Image Segmentation and Novel Approaches to FCM Algorithm", *International Journal of Research in Engineering and Technology*, Vol. 2, No. 12, pp. 346-349, 2013.

[3] A.J. Jadhav, Vaibhav Kolhe and Sagar Peshwe, "Text Extraction from Images: A Survey", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 3, No. 3, pp. 333-337, 2013.

[4] Chengquan Zhang, Cong Yao, Baoguang Shi and Xiang Bai, "Automatic Discrimination of Text and Non-Text Natural Images", *Proceedings of International Conference on Document Analysis and Recognition*, pp. 886-890, 2015.

[5] Qiong Xu, Zongliang Gan, Changhong Chen and Feng Liu, "Novel Chinese Text Localization Method for Natural Images through SVM Classification", *Journal of Computational Information Systems*, Vol. 9, No. 18, pp. 7291-7298, 2013.

[6] Lluis Gomez and Dimosthenis Karatzas, "Multi-script Text Extraction from Natural Scenes", *Proceedings of International Conference on Document Analysis and Recognition*, pp. 1-5, 2013.

[7] Stanley Sternberg, "Biomedical Image Processing", *IEEE Computer*, Vol. 16, No. 1, pp. 22-34, 1983.

[8] Lalit Prakash Saxena, "An Effective Binarization Method for Readability Improvement of Stain-Affected (Degraded) Palm Leaf and Other Types of Manuscripts", *Current Science*, Vol. 107, No. 3, pp. 489-496, 2014.

[9] Rapeeporn Chamchong, Chun Che Fung and Kok Wai Wong, "Comparing Binarisation Techniques for the Processing of Ancient Manuscripts", *Proceedings of International Symposium on Entertainment Computing*, pp. 55-64, 2010.

[10] Salem Saleh Al Amri, N.V. Kalyankar and S. Khamitkar, "Deblured Gaussian Blurred Images", *Journal of Computing*, Vol. 2, No. 4, pp. 132-139, 2010

[11] C. Li, S. Anwar and F. Porikli, "Underwater Scene Prior Inspired Deep Underwater Image and Video Enhancement", *Pattern Recognition*, Vol. 98, pp. 1-13, 2020.

[12] A. Jalal, A. Salman and F. Shafait, "Fish Detection and Species Classification in Underwater Environments using Deep Learning with Temporal Information", *Ecological Informatics*, Vol. 57, pp. 1-16, 2020.