Real Time Night Vision Surveillance using Improved Dark Channel Prior

M.V. Naga Bhushanam¹, G. V. Pradeep Kumar²
Department of ECE, CBIT(A), Hyderabad, India

Abstract—Videos taken under low lighting conditions usually result in severe loss of visibility and contrast and are uncomfortable for observation and analysis. Night vision cameras that cater to the needs are expensive and less versatile. To be cost effective and extract maximum information from videos taken in low lit conditions, video enhancing techniques must be used. Though there are many night vision enhancement techniques available in literature, this paper particularly emphasizes about Improved Dark Channel Prior algorithm and its results. This approach suits well for real time night video enhancement. It has been found that a pixel-wise inversion of a night video appears very similar to the video obtained during foggy days. The same idea of haze-wise inversion approach is used to boost the visual quality of night videos. An improved dark channel prior model is presented that is integrated with Gaussian Pyramid operators for local smoothing. The experimental results show that the proposed method can boost the perceptual quality of detailing in night videos.

Keywords—Video Enhancement, Gaussian Pyramid, Local Smoothing, Haze Removal, Improved Dark Channel Prior

I. INTRODUCTION

A Dark Channel Prior can be assumed as an envelope which tries to “hide” undesired aberrations that try to degrade the quality of an image or a video. DCP can be utilized in night vision enhancement as a lowly lit video is similar to the inverse of a very hazy video. Hence, parameters like depth of the haze and actual light present during the time of capture play a major role. Proper modelling of these parameters helps in obtaining better results. Typically, Dark Channel Prior is based on the property of “dark pixels”. They are pixels that have a very low intensity in at least a colour channel. The DCP-based dehazing techniques are composed of four major parameters: atmospheric light estimation, transmission map estimation, transmission map refinement, and image reconstruction. An in-depth analysis of the DCP is performed in this thesis. Outdoor images have poor visibility under inclement weather due to the absorption and dispersal by atmospheric particles in haze. Poor visibility has a negative impact not only on consumer photography but also on applications of computer vision for outdoor environments such as object detection and video surveillance. Since haze-free images are visually appealing and can greatly enhance the output of computer vision tasks, Haze removal, which is called dehazing, is considered an important process. Under different weather conditions multiple images from the same scene are captured to be used as reference images with clear weather conditions. Such methods, however, have limitations in online image dehazing applications with multiple reference images, and may involve special imaging sensors. That leads the researchers to focus a single reference image on the dehazing process. Single image-based methods are based on the characteristics typical of haze-free images. Hence, DCP and dehazing go hand in hand.

A real-time lowly lit video is very similar to an inverse of a very hazy video. This is the key assumption of this project and dark channel prior algorithm helps achieve it. Hence, every real time lowly lit video is cut up into many frames, enhanced and stitched back together as a video feed. This work models every image using image degradation model. This paper is organized as follows; Section II emphasizes on literature work. Section III briefly deals with methods such as histogram equalization and contrast stretching that are alluded to in this project's length and throws a limelight at the concept of enhancing night vision. Section IV focuses on the algorithm used and discusses the parameters used in the Improved DCP model. The results and the intermediate changes to the input image are discussed in Section V. This work is concluded in Section VI.

II. LITERATURE REVIEW

Digital video has become an essential part of everyday life. It is well-known that in recent years, video enhancement has gained significant attention as an important subject in computer vision. The goal is to enhance the visual quality of the video or to provide a "better" representation for future automated video processing, such as analysis, identification, segmentation, and recognition [6]. There are numerous applications that acquire, process, and use digital video, such as surveillance, general identity verification, traffic, criminal justice systems, civil or military video processing etc. Carrying out video enhancement understanding under low quality video is a challenging problem because of the following reasons:

A. Due to low contrast, extract moving objects from the dark background becomes difficult. Most colour-based methods will fail on this matter if the colour of the moving objects and that of the background are similar.
B. The signal to noise ratio is usually very low due to high ISO (ISO is the number indicating camera sensors sensitivity to light). Using a high ISO number can produce visible noise in digital photos. Low ISO number means less sensitivity to light.

C. The information carrying video signal is a degraded version of a source or original video signal which represents the three-dimensional continuous world. These degradations can be a result of the acquisition process, or the rate and format conversion processes.

D. Environmental information affects the way people perceive and understand what has happened. Hence, dealing with moving tree, fog, rain, behaviour of people in night-time video are the difficult because they lack background context due to poor illumination.

E. Inter-frame coherence must also be maintained i.e., the moving objects region as weights in successive images should change smoothly.

F. One pixel from a low-quality image may be important even if the local variance is small, such as the area between the headlights and the taillights of a moving car.

G. The poor quality of the used video device and lack of expertise of the operator [6].

Researchers have been trying to reduce many of these shortcomings using various methods, algorithms, and mathematical models. It is deduced from empirical data that the quality of a dimly lit video depends on its dynamic range [2],[7],[8]. According to the application, it can either be increased or decreased. Dynamic range compression followed by contrast enhancement helps in multi-sensor image fusion [2]. Efficient dynamic range compression plays an important part in logarithmic image processing model. It helps in shadow compensation for still coloured images [8]. The nature of the photoreceptor also plays a crucial role in the amount of information captured in an image. An underexposed image can be tweaked adaptively and independently by varying exposure at every photoreceptor. Temporal integration expands the image’s dynamic range and reduces noise simultaneously [7]. Also, moving objects in a video contribute unwanted artefacts and noise. Smoothing kernels adapt to local spatial and temporal intensity structure in adaptive enhancement. This method hence preserves and enhances fine spatial detail. This hence prevents motion blur and leading to minimum information loss [5]. Night vision enhancement has flourished under thermal cameras. A non-linear enhancement of fused visible video with infrared video helps in better vision [1]. The need to eliminate noise and preserve the naturalness of a video is of utmost priority. The method of lightness decomposition using retinex caters to this. It states that the overall lightness of an image is a summation of reflex lightness and ambience illumination [4]. Ambience illumination generally caters to unnecessary artefacts and noise and hence is not required to preserve naturalness. The ability to see in dim light is inspired by nocturnal animals. It is modelled that an optimum summation of spatial and temporal visuals should enable night vision [5]. The most successful model of night vision yet is image degradation model. This method models a dim lit image as an inverse of a hazy image. An effective prior should be used to remove haze from a single input image. A depth map of haze is also obtained from this model which describes the nature of haze which is analogous to noise [9]. The most important constituents of image dehazing are transmission map and air-light estimation. Proper estimation of these two parameters can improve the detailing of the output video. One can remove unwanted artefacts in depth map of hazy image using weighted least mean square-based edge preserved smoothing technique. The complexity of this method varies between O(n) and O(nlog,n) [10]. Air-light estimation be well modelled using atmospheric scattering model. It caters well for both real and artificial daze [11]. This method calculates atmospheric scattering for every frame and hence consumes a lot of time in computation. A faster computation of all these parameters can be carried out using temporal correlations between subsequent frames. Inverting a low-lit video and optimising image dehazing algorithm can build a fast and efficient algorithm [3]. More details of the image can be preserved by local smoothing and Gaussian pyramid parameters. This ensures to build a better and an improved dark channel prior [12]. An improved dark channel prior helps in perceiving better details in lowly lit night-time video and avoids excessive enhancement phenomenon.

III. OVERVIEW OF NIGHT VISION ENHANCEMENT TECHNIQUES

Most of the image processing techniques used for night vision are based on adjusting the image contrast to make the objects easier to detect in the scene. These include linear contrast stretching, histogram equalization, contrast enhancement dependent on wavelets and Thermal Imaging etc. Many versions of the conventional histogram equalization are in use. They can be loosely categorized into three groups, Regional, Adaptive and Block-based.

Contrast stretching (often called normalization) is a basic image enhancement technique that aims to increase contrast in an image by 'stretching' the range of intensity values it contains to cover a specified range of values, e.g., the maximum range of pixel values allowed by the image form in question. This differs from the more advanced equalization of the histogram, as it can only apply a linear scaling function to the pixel values of the image. Consequently, the 'enhancement' is less harsh. (Most implementations accept a gray-level image as their input and produce another gray-level image as their output.)
Thermal imaging is a method of using infrared radiation and thermal energy to collect information about objects, often in low visibility conditions, to formulate photographs of them.

A. Histogram Equalization

Histogram Equalization is a method used in computer image processing to enhance contrast in images. This is accomplished by effectively spreading the most frequently used intensity values, i.e., extending the image's intensity range. Usually, this method increases the global contrast of images when closely contrasting values represent its usable data.

This allows areas with lower local contrast to gain a greater contrast. A picture in which the pixels seem to uniformly occupy the full spectrum of possible gray rates has a strong contrasting appearance. Such an image shows a lot of detail at the gray-level and has a high dynamic range. The equalization of histograms has thus proved to be a very effective tool for night vision.

B. Contrast Stretching

Contrast stretching is a typically a linear transformation which can be applied to images taken under poor illumination conditions. The basic idea behind it is to increase the dynamic range of the gray levels in the image being processed. Contrast stretching is also referred to as normalization. This operation works as follows. Initially, the upper and lower-pixel value limits over which the image is to be normalized is specified.

Usually, these limits are the minimum and maximum possible pixel values of the original image. For example, for an 8-bit gray level image the lower and upper limits might be 0 and 255. Let n represent the lower limit and m the upper limit.

The normalization procedure then reads every pixel in the image to find the lowest and highest pixel values currently present. Say these are called l and u respectively. Then each pixel P is scaled using the following function equation, shown in the equation 1.

\[
P_{\text{out}} = \left( \frac{P_{\text{in}} - l}{m - n} \right) \left( \frac{l - n}{u - l} \right) + n \quad (1)
\]

C. Thermal Imaging

Thermal imaging is based on the science of infrared energy which is emitted from all objects (otherwise known as "heat"). This energy from an object is also known as the "heat signature," and the amount of radiation emitted tends to be proportional to the object’s overall heat.

Thermal cameras or thermal imagers are sophisticated instruments consisting of a sensitive heat sensor, capable of recording minute temperature changes. As they collect the infrared radiation from objects in a particular environment, they can begin mapping an image based on the temperature measurement differences and inflections.

Thermal images are generally grayscale: with white representing heat, black representing colder regions and various shades of grey indicating temperature gradients between the two.

IV. DARK CHANNEL PRIOR ALGORITHM

The Dark Channel Prior algorithm is based on the property of “dark pixels,” which have a very low intensity in at least one colour channel, except for the sky region. In outdoor environment these objects form shadows which provides dark channels.

Three factors play a vital role in the low intensity of dark channel prior.
- shadows of moving stationary objects.
- physical objects, which are dark.
- coloured objects with low reflectivity.

A real-time night video enhancement approach using an improved dark channel prior model based on atmospheric scattering model is developed. It has low computational complexity and straightforward motivation.

First, the input night video is pre-processed through pixel-wise inverting. Then, the global atmospheric light is calculated which indicates the concentration of fog for the whole video and use the improved dark channel prior to estimate medium transmission for the inversion of each input frame. With the global atmospheric light and transmission, the enhanced inversion is calculated. Finally, post-processing is applied to obtain the enhanced videos.

Dong observed that a pixel-wise inversion of a night video has quite similar appearance with the video acquired at foggy days. They examined the statistical similarity between the pixel-wise inversion of the night video and the foggy video and claimed that it is conceivable to use haze removal algorithm to enhance the night video.

For a colour input image \( I(x) \), the pixel-wise inversion of its channel \( c \in \{r,g,b\} \) can be calculated as shown in the equation 2.

\[
I_{\text{inv}}^c(x) = 255 - I^c(x) \quad (2)
\]

Based on the above observation, the image degradation model proposed is introduced, which is formulated as shown in the equation 3.
$I_{\text{inv}} (x) = J_{\text{inv}} (x) t(x) + A(1 - t(x))$ \hspace{1cm} (3)

where $J_{\text{inv}}(x)$ is the scene radiance, $t(x)$ is the medium transmission, and $A$ is the global atmospheric light. Once knowing $A$ and $t(x)$, $J_{\text{inv}}(x)$ can be obtained from $I_{\text{inv}}(x)$ as shown in the equation 4.

$$J_{\text{inv}}(x) = \frac{I_{\text{inv}}(x) + A t(x)}{t(x)} + A \hspace{1cm} (4)$$

Then, $J_{\text{inv}}(x)$ is converted to get the final enhanced frame. To achieve this, it is needed to estimate the image degradation model (Eq. 3) including the global atmospheric light and medium transmission parameters.

**A. Estimating the Global Atmospheric Light**

The pixel-wise inversion of each frame of the input video can be obtained using Eq. 2. Then the next task is to estimate the global atmospheric light. In real-world night video, noise interference or camera shake may lead to unstable atmospheric light value and bring in colour drift and unpleasant scintillation. Therefore, estimating $A$ frame-wise is not a befitting scheme for video enhancement, always be time-consuming and they cannot be applied in real-time systems. Dong calculates atmospheric light only once the first frame of a Group of Pictures (GOP), and for successive GOPs, they obtain updated atmospheric light through a weighted combination way. However, the noise may confuse the atmospheric light since they only have once calculation process for a GOP. Intuitively, one can find a few dark patches whose inversion always contain atmospheric light in each original frame. Hence, it is assumed that the pixels which correspond to atmospheric light located in a large smooth area, and for a period of time, the atmospheric light has no change. Furthermore, considering the influence of noise, global atmospheric light $A$ is calculated. Firstly, the minimum in three colour channels is calculated as shown in the equation 5.

$$I_{\text{inv}}(x) = \min_{c=r,g,b} \left( I_{\text{inv}}(x) \right) \hspace{1cm} (5)$$

In $I_{\text{inv}}(x)$, there are always a lot of noise points and some interference areas. Hence, median filter is used here to smooth it and reveal the atmospheric light areas. Then, the pixel is chosen whose minimum intensity in all colour channels is the highest in the smoothed $I_{\text{inv}}(x)$. In the proposed approach, $A$ is calculated in the first five frames and take their average as the global atmospheric light for an input video.

**B. Estimating Medium Transmission using an Improved Dark Channel Prior**

After obtaining the global atmospheric light, the transmission can be estimated. In a general way, it is assumed that the pixels on the same object should have the same or similar depth values. In addition, as showed in Eq. 4, if local details of the transmission map are closer to the input image, the restored image will lose details more easily. Hence, to maintain $t(x)$ locally as smooth as possible an improvement in DCP is made. First, the constraints of the DCP are chosen as shown in the equation 6.

$$I_{\text{dark}} (x) = \min_{c=r,g,b} \left( \frac{I_{\text{inv}}(x)}{A} \right) \hspace{1cm} (6)$$

It is called as $I_{\text{dark}} (x)$ coarse dark-channel-map (DCM).

According to the physical characteristics of foggy image formation and DCP, strict constraint can be obtained as shown in the equation 7.

$$0 \leq I_{\text{dark}} (x) \leq I_{\text{dark}} (x) \hspace{1cm} (7)$$

where $I_{\text{dark}} (x)$ is the optimized dark-channel-map. For the aim of smoothing the coarse DCM and accelerate processing process, Gaussian pyramid operation is introduced here. In a Gaussian pyramid, subsequent images are weighted down using a Gaussian average and scaled down. Each pixel containing a local average corresponds to a neighbourhood pixel on a lower level of the pyramid. This technique is used especially in texture synthesis.

An image gaussian pyramid for $I_{\text{dark}} (x)$ is established and the following processes are all applied on the top layer. Then, it is processed using median filter. Median filtering is a nonlinear method used to remove noise from images. It is widely used as it is very effective at removing noise while preserving edges. It is particularly effective at removing ‘salt and pepper’ type noise. The median filter works by moving through the image pixel by pixel, replacing each value with the median value of neighbouring pixels. The pattern of neighbours is called the "window", which slides, pixel by pixel over the entire image 2 pixel, over the entire image. The median is calculated, as shown in the equation 8 by first sorting all the pixel values from the window into numerical order, and then replacing the pixel being considered with the middle (median) pixel value.

$$I_{\text{med}}(x) = \text{Med}_s(I_{\text{dark}}(x)) \hspace{1cm} (8)$$

where $S$ is the scale of the median filter. To make $t(x)$ locally smooth while maintaining the edges, it is needed to wipe off the details in $I_{\text{med}}(x)$.
First, local standard deviation is used to estimate the local details approximatively. Moreover, the median filter can enhance the robustness of local details approximation.

Such calculation process can be expressed as shown in the equation 9.

\[
I_{\text{detail}}(x) = \text{Med}_x \left( I_{\text{med}}(x) - eI_{\text{dark}}(x) \right)
\] (9)

Then, local details can be removed from \( I_{\text{med}}(x) \) using the equation 10.

\[
I_{\text{smooth}}(x) = I_{\text{med}}(x) - I_{\text{detail}}(x)
\] (10)

Thinking about the constraint in Eq. 7, coarse DCM can be further optimized as shown in the equation 11.

\[
I_{\text{dark}}(x) = \begin{cases} 
\mu eI_{\text{dark}}(x) & \text{if } \mu eI_{\text{dark}}(x) < I_{\text{smooth}}(x) \\
I_{\text{smooth}}(x) & \text{otherwise}
\end{cases}
\] (11)

where \( I_{\text{dark}}(x) \) is called the optimized DCM, and \( \mu \) is a constant parameter that controls the strength of the constraints. In this work, \( \mu \) is set to 0.95 and then, calculate the transmission with the optimized DCM as DCP using the equation 12.

\[
t(x) = 1 - \omega I_{\text{dark}}(x)
\] (12)

where \( \omega \) is a parameter to adjust the value of \( t(x) \). In this paper, it is set to 0.98 as it gives the best results.

C. Obtaining the Enhanced Result

Knowing the transmission map and global atmospheric light, the inversion the enhanced result can be obtained, \( J_{\text{inv}}(x) \) as Eq. 3. Unlike Dong et al.’s method, it is not needed to adjust the transmission map to regulate enhanced results. The transmission map \( t(x) \) is adaptive to low lighting and high lighting situations in night videos is achieved. Once after obtaining \( J_{\text{inv}}(x) \), the inversion operation is performed on it to get the final enhanced video frame as shown in the equation 13.

\[
J^e(x) = 255 - J_{\text{inv}}^e(x)
\] (13)

D. Gaussian Pyramid

The Gaussian pyramid is a technique in image processing that breaks down an image into successively smaller groups of pixels, in repeated steps, for the purpose of blurring it. It is named after German mathematician Johann Carl Friederich Gauss. This type of precise mathematical blurring is used extensively in artificially intelligent computer vision as a pre-processing step. For instance, when a digital photograph is blurred in this way, edges of objects are much easier to detect, enabling a computer to identify them automatically.

The "pyramid" is constructed by repeatedly calculating a weighted average of the neighboring pixels of a source image and scaling the image down. It can be visualized by stacking progressively smaller versions of the image on top of one another. This process creates a pyramid shape with the base as the original image and the tip a single pixel representing the average value of the entire image.

V. RESULTS AND DISCUSSION

The Dark Channel Prior algorithm is a method of Image Enhancement which specializes in Haze removal and Night Vision enhancement. A lot of sample images have been tested using this algorithm and the results cater to the expectations.
Dark Channel Prior algorithm is applied on the images shown in the Fig.1,3,5,7,9 11,13 and 15. The enhanced images of the Fig.1,3,5,7,9 11,13 and 15 are shown in the Fig.2,4,6,8,10 12,14 and 16, respectively.

A. Video Processing using Dark Channel Prior
A video is analogous to a fast-moving train of many images put in a cascade. So, a video can be divided into many frames or images and can be processed individually and reconstructed into a video. Hence, a real-time video can be processed by disintegrating and re-integrating its images.
Dark Channel Prior algorithm is applied on the different frames of a video, shown in the Fig.17,19,21,23 and 25. The enhanced images of the Fig.17,19,21,23 and 25 are shown in the Fig.18,20,22,24 and 26, respectively.

B. Observations and Inferences

1) **Image:** DCP algorithm brightens up the whole image. It enhances the perceptual quality of the image. The quality of the output image is directly proportional to the pixel density of the input image. More pixels, more detail, better DCP output. DCP output is rich in detail and hence one can carry out edge detection, object recognition or segmentation according to the requirement of the user. Better results are found when the light source was localized to a smaller area of the image rather than a scattered source [Fig. 11, Fig. 15]

2) **Video:** DCP algorithm takes a considerable amount of time to process a real-time video. The quality of video is improved in terms of visual perceivability. The video is illuminated very well. The output video is a decipherable data that can be used as a feed for object recognition, text recognition and many other image processing methods. The motion-blur aspect has a scope for improvement [Fig. 22 and Fig. 24]. The grainy nature of the output video is inversely proportional to the frame rate of the input video. Lesser frame rate leads to a more pixelated video. The time taken to process a video also has scope of improvement.

VI. CONCLUSIONS

In this paper a real time night video enhancement algorithm is presented which is based on the motivation that a pixel-wise inversion of a night video has similar appearance with the video acquired at foggy days.

An improved dark channel prior model is presented in this report which is combined with local smoothening and image Gaussian pyramid operators to enhance the perpetual quality of night videos. This improved dark channel prior algorithm proposed optimizes the transmittance through the guided filter and the transmission of the bright areas is corrected. Finally, the intensity and the contrast of the recovered image are adjusted. This avoids the color distortion and enhances the visual effect of image. Experimental results show that the proposed method is capable of enhancing local details while avoiding excessive enhancement. Compared with the state-of-art methods the proposed method can achieve more superior performance and real time processing speed.

REFERENCES

[1] H. Ngo, T. Li, and M. Zhang, “A visibility improvement system for low vision drivers by nonlinear enhancement of fused visible and infrared video,” in Proceedings of Conference on Computer Vision and Pattern Recognition. IEEE, 2005, pp. 25–32.

[2] T. Li, H. Ngo, and M. Zhang, “A multi-sensor image fusion and enhancement system for assisting,” in Proceedings of Conference on Applied Imagery and Pattern Recognition Workshop. IEEE, 2005, pp. 106–113.

[3] X. Dong, G. Wang, Y. Pang, W. Li, J. Wen, W. Meng, and Y. Lu, “Fast efficient algorithm for enhancement of low lighting video,” in Proceedings of International Conference on Multimedia and Expo. IEEE, 2011, pp. 1–6.

[4] B. Li, S. Wang, and Y. Geng, “Image enhancement based on retinex and lightness decomposition,” in Proceedings of International Conference on Image Processing. IEEE, 2011, pp. 3417–3420.

[5] H. Malm, M. Oskarsson, and E. Warrant, “Adaptive enhancement and noise reduction in very low light-level video,” in Proceedings of Conference on Computer Vision. IEEE, 2007, pp. 1–8.

[6] Y.B. Rao and L.T. Chen, “A survey of video enhancement techniques,” Journal of Information Hiding and Multimedia Signal Processing, vol. 3, pp. 71–99, January 2012.

[7] E.P. Bennett and McMillan L., “Video enhancement using per-pixel virtual exposures,” ACM Transactions on Graphics, vol. 24, pp. 845–852, July 2005.
[8] F. Albu, C. Florea, C. Vertan, and A. Drimbarean, “One scan shadow compensation and visual enhancement of color images,” in Proceedings of International Conference on Image Processing. IEEE, 2009, pp. 3313–3136.

[9] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior,” in Proceedings of Conference on Computer Vision and Pattern Recognition. IEEE, 2009, pp. 1956–1963.

[10] Y. Koschmieder, “Theorie der horizontalen sichtweite,” Journal of Beitr. Phys. Freien Atmos., vol. 12, pp. 171–181, 1924.

[11] M. Pedone and J. Heikkila, “Robust airlight estimation for haze removal from a single image,” in Proceedings of Conference on Computer Vision and Pattern Recognition Workshops. IEEE, 2011, pp. 90–96.

[12] X. Jiang, H. Yao, S. Zhang, X. Lu and W. Zeng, "Night video enhancement using improved dark channel prior," in IEEE International Conference on Image Processing, Melbourne, VIC, 2013, pp. 553-557.
