# Low light video denoising and enhancement using kalman filter and tone mapping technique

[1]K. Narasimha Prasad, [2]P.Madavi

[1]M.Tech-ECE-DECS, [2]Assistant Professor
Priyadarshini College of engineering and technology, Nellore, AP, India

***Abstract*—In this paper, a new perception for noise reduction and enhancement of extremely low-light video. The proposed manner works directly on the color filter array (CFA) raw video for attain low memory consumption. A motion adaptive temporal filtering based on a Kalman structured updating is proposed for noise elimination. Dynamic range of denoised video is elevated by balancing of RGB histograms employing Gamma correction with adaptive clipping thresholds. Certainly, residual noise is eliminated by employing a nonlocal means (NLM) denoising filter.**

**Index Terms — Noise reduction, tone mapping, nonlocal means, and low-light video**

---

## 1.INTRODUCTION

Over the last several decades, there have been substantial improvements in modern digital cameras including resolutions and sensitivity. Despite these improvements, quality of videos in low-light conditions is still limited. Firstly, low-light videos have poor dynamic range. To capture images of high dynamic range, most consumer cameras often rely on automatic exposure control, but longer exposure time results motion blur. Secondly, image sequences captured in low-light conditions often have very low signal-to-noise ratio (SNR).

However, the estimation of a transmission term in the hazy image acquisition model by using a dark channel prior (DCP) becomes unreliable in very low-light conditions and requires large computation loads.

Fig: Example of a moderately low-light video (left) and an extremely low-light video (right).

Only a few of the approaches provide experimental results for videos in very low-light conditions. illumination level is below 0.1 lux, when the level of noise becomes relatively much higher than the signal, thus conventional denoising and tone-mapping techniques cannot achieve satisfying enhancement performances. In this paper, the proposed method, which is improved from the previous work, is aimed to develop a novel framework to enhance video from extremely low-light environments.

It consists of an effective motion adaptive temporal filter based on the Kalman filter framework, a tone-mapping by histogram adjustment with adaptive clipping, and spatial noise reduction with the NLM denoising filter. Since a denoising step may cause additional artifacts, all steps are implemented to handle the Bayer pattern color filter array (CFA) raw video.

Two major characteristics of low-light video is high level of noise (i.e., low SNR) and low dynamic range. Some pixels may have higher dark current than others, thus produces a fixed pattern noise (FPN). FPN increases with exposure time, so it is signal-dependent and more noticeable at higher intensities. Also, the shapes of histograms of every color channels are almost identical. as the illumination level is decreased, the height of histogram moves towards the zero, thus it becomes an “L-shape”.

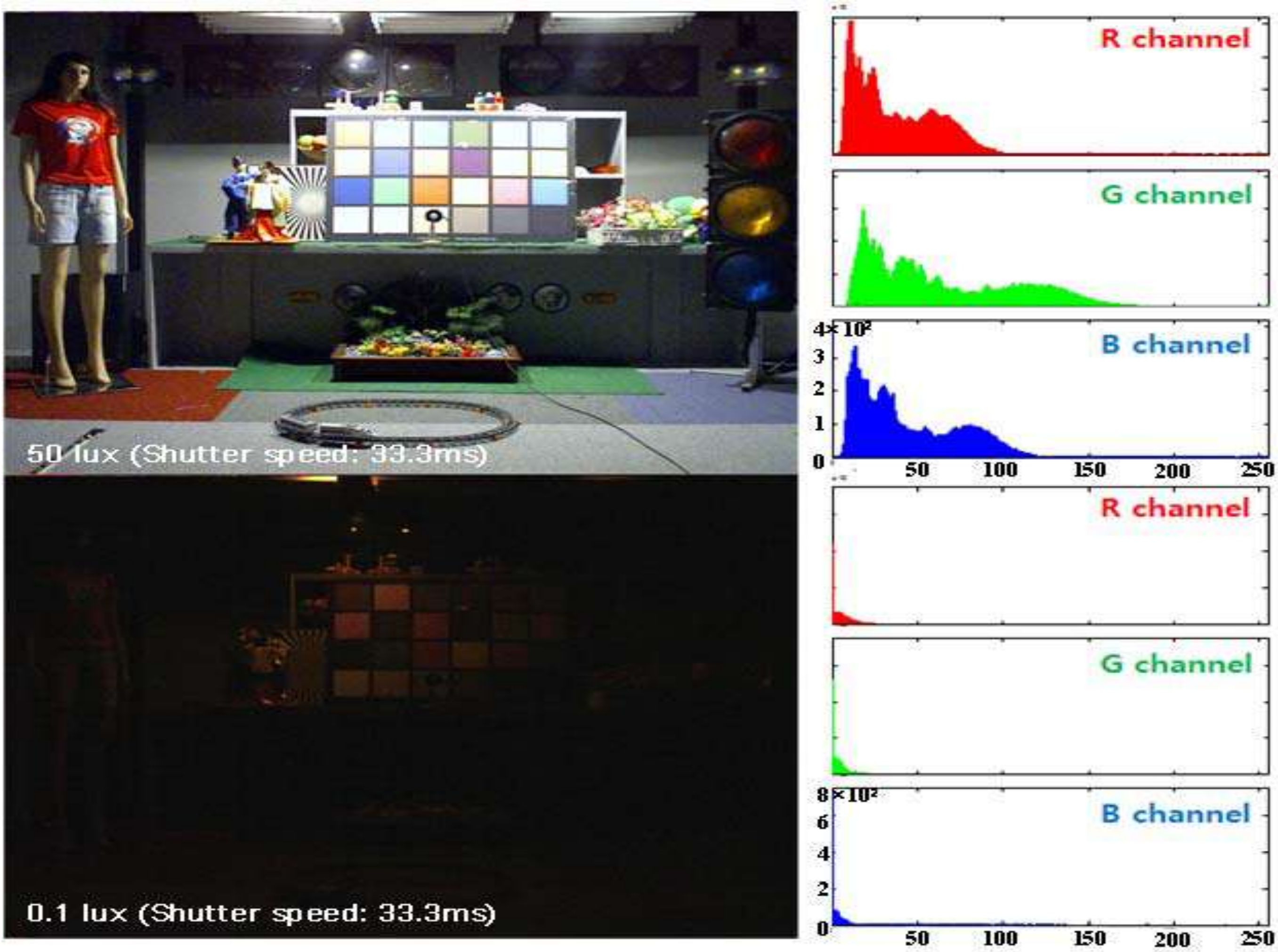



fig: Color histograms of videos captured in normal lighting (top) and extremely low lighting (bottom) conditions. Most of pixels have very small intensity ranging only approximately 5% of maximum intensity.

**2. Proposed method**:

The input video is converted into frames. For the identification of the movements in the video the current video frame is subtracted from the previous frame to obtain the difference image. The difference image represents the movements in the video frames and they were useful for the removal of the ghost effects while employing noise reduction. The difference image is also helpful for the identification of the noise locations in the frames which can be then eliminated.

The kalman gain were helpful in the estimation of the temporal noises in the images. As the result of the updatation step the temporal noises from the images were removed. Tone mapping process enhances the images and produces more accurate clear image compared to the video with dim and brighter illumination. Gamma correction is employed for the tone mapping of the images. In Gamma correction the histogram of the images were normalized to particular intensity. The clipping of the histogram decreases the intensity of the brighter image pixels and increases the intensity of the dim image pixels.

Tone mapping process is employed comparing the images with clipping and without clipping. The clipping process enhances the image further compared with tone mapping without clipping process. For clipping of the signals clipping thresholds were setted for the images based on the higher range of the histogram. The spatial noise in the tone mapped images were then removed using Non Local Means filter.

The performance of the process is measured by the calculation of the performance metrics like PSNR, SSIM, GCF and NIQE. PSNR values indicates the noise ratio in the input video frame and the resulting denoised video frame. The PSNR value must be high. SSIM value indicates the similarity between the input video frame and the resulting denoised video frame.

The SSIM value must be within one. GCF - Global Contrast Factor is a measure for the analysis of the comparison of the contrast in input video frame and the resulting denoised video frame. NIQE - Natural Image Quality Evaluator is a distance metric for the model statistics and a factor for the comparison of the quality of the input video frame and the resulting denoised video frame.

Some advantagesare: The overall performance indicated that the performance of the process is well improved which is due to the application of two different noise reduction system and a efficient contrast enhancement system.

The prediction and the updation steps employed in the kalman filter identifies the position exactly where noise is present and hence the original color information's were preserved.

## 3. SYSTEM ARCHITECTURE

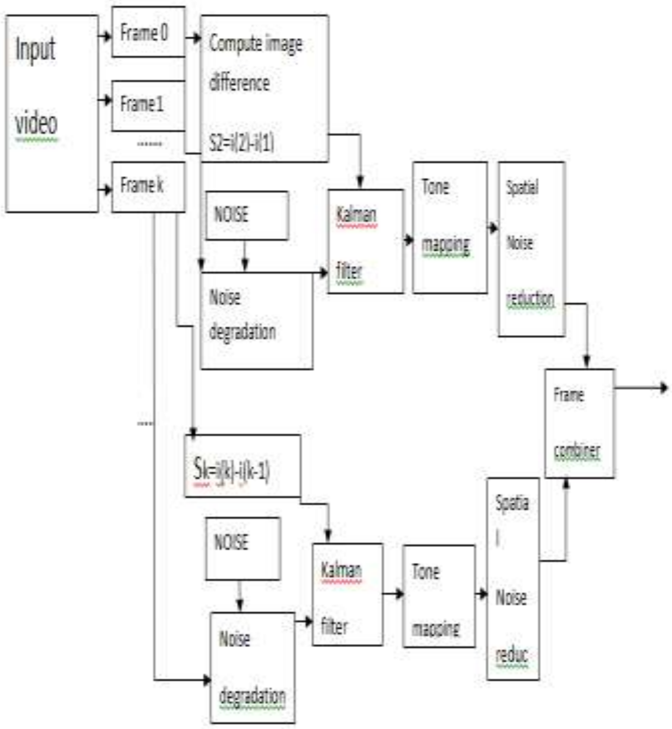



Fig: System architecture

## 4. Overall Framework

Since noise in a low-light video can be amplified by stretching dynamic range, severe noise should be suppressed before the tone-mapping step. A spatio-temporal filtering can suppress most of noise in a low-light video. Moreover, because noise cannot be eliminated perfectly with recent denoising methods, remaining noise level is raised by tone-mapping.
Therefore, proper noise reduction strategies should be applied both before and after tone-mapping process.

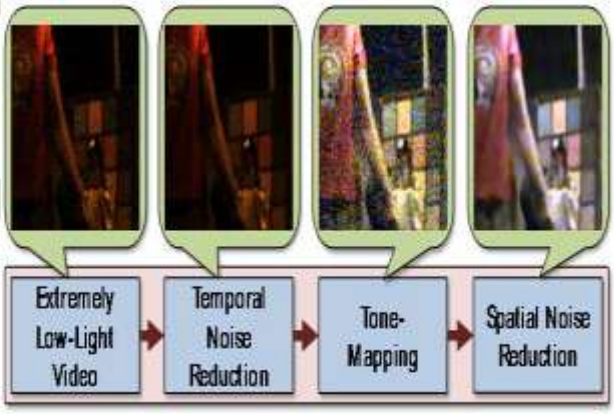



Fig: overall framework of proposed low light video enhancement

An effective motion adaptive temporal filtering, which is developed by modifying the Kalman filter approach, is applied at the very first. And then, the narrow dynamic range of denoised signal is widened by Gamma correction of each RGB histogram with low and high intensity levels clipped by appropriate thresholds. Lastly, the remaining amplified noise after having been through the former two steps is filtered by spatial noise reduction. As it is apparent that visual features in the enhanced video are more distinguished than those in the initial input video, the patch-based nonlocal means filter can remove the remaining noise effectively while preserving edges.

## 5. Temporal Noise Reduction

Since noise in low-light video is regarded as a zero-mean Gaussian after eliminating FPN, it can be suppressed easily with a simple averaging along the temporal direction. However, a simple temporal averaging may result artifacts when motion exists in video sequences. To reduce noise while preventing “ghost effects”, many approaches utilizing an adaptive spatio-temporal filter have been proposed. Bennett introduced an adaptive spatio-temporal filter using a combination of temporal or spatial bilateral filters in presence of motion.

An adaptive anisotropic filter using a 3D structure tensor was presented to smooth a low-light video. Recently, various patch-based video denoising algorithms were proposed. A block-matching and 3D collaborative filtering (BM3D), which is considered as one of the state-of-the-art denoising technique, was extended to handle video signals. Also, a 3D NLM filter and its variants are developed for video denoising. Han and Chen introduced a dynamic nonlocal means (DNLM) denoising method based on Kalman filtering theory. In Kalman filter framework, the state transition and observation equations for a video can be defined as follows:

$$\begin{cases} X_t = M_t X_{t-1} + S_t, \\ Y_t = X_t + V_t, \end{cases} \text{where} \begin{cases} S_t \sim N(0, Q_t^{-1}), \\ V_t \sim N(0, C_t^{-1}), \end{cases} \qquad (1)$$

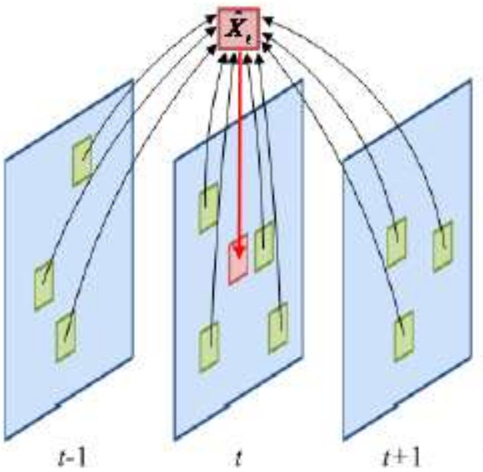


Fig: 3D NLM

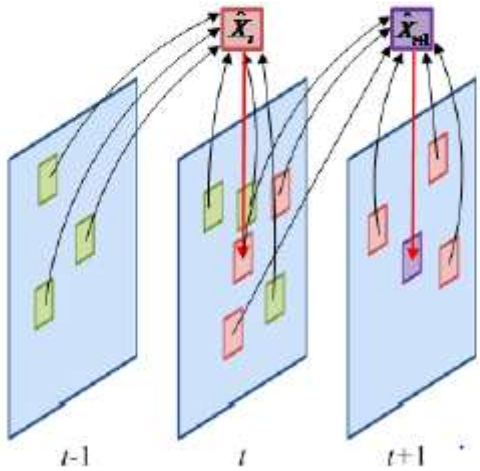


Fig: DNLM

Where Xt and Yt are clear image frame and noisy image frame at time t respectively. They are rearranged as column vectors in lexicographic order. Since Mt denotes a motion matrix and St Accounts for the difference between the previous and current frame, Qt -1 reflects the amount of motion estimation error. Vt represents the Gaussian noise during acquisition of current frame. Thus, Ct -1 denotes the variance of measurement noise. DNLM exploits a weighted translational motion model under an assumption that most complex motion fields can be represented by a linear combination of global translations.
The estimation flows of 3D NLM and DNLM are illustrated in above figures
Accuracy of finding similar patches is a critical factor for performance of NLM-based approaches. However, it becomes very hard to find a similar patch and assign a proper weight to it when an illumination level is extremely low.

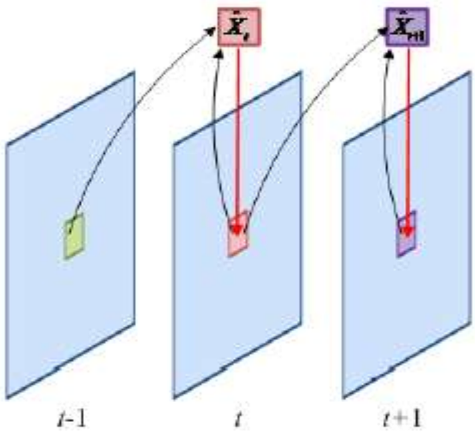


Fig: Proposed temporal noise reduction.

Thus the weighted averaging process with wrongly chosen patches from a spatial neighborhood may produce blurry results. In the proposed method, neighboring patches only in temporal domain are considered in the Kalman filter estimation process as depicted in above Fig to minimize motion blurs. This can be realized by defining the motion matrix in (1) as an identity matrix as above:

$$\begin{cases} X_t = X_{t-1} + S_t, \\ Y_t = X_t + V_t, \end{cases} \text{where} \begin{cases} S_t \sim N(0, Q_t^{-1}), \\ V_t \sim N(0, C_t^{-1}), \end{cases} \qquad (2)$$

Note that St now reflects a difference of intensity between previous and current frames, hence it becomes high when motion exists between previous and current frames. Thus, its covariance matrix $Q_t^{-1}$ should be adapted to the amount of motion.

Consequently, the new diagonal weight matrix computed by the sum of squared distance (SSD) between patches centered at a same spatial location of previous and current noisy frame is introduced as below:

$$W_t(i,j) = \exp\left\{-\frac{G_\rho * \left\|R_{i,j}(Y_t - Y_{t-1})\right\|_2^2}{\sigma_s^2}\right\}, \qquad (3)$$

whereRi,j denotes an operator for extracting a patch with its center pixel at (i,j). Gρ is a modified Gaussian kernel for Bayer pattern CFA image with its standard deviation ρ as in Fig.5 $\sigma_S^2$ is diagonal value of $Q_t^{-1}$ and controls the degree of ghost effect. In previoius work, a median absolute deviation (MAD) method is used for estimating σS as a noise standard deviation. However, MAD can hardly provide reliable results for estimating a noise level when the illumination level is very low. Therefore, in this paper, it is determined as being proportional to the average of squared intensity difference between current and previous input frame:

Consequently, the new diagonal weight matrix computed by the sum of squared distance (SSD) between patches centered at a same spatial location of previous and current noisy frame is introduced as below:

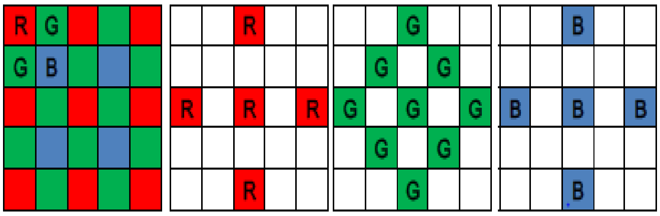


Fig: Modified Gaussian kernels for smoothing Bayer pattern image. Only pixels with same color contribute to filter a center pixel of 5×5 mask.

$$\sigma_s^2 = \delta' \cdot \left[\frac{(Y_t - Y_{t-1})^{\mathrm{T}}(Y_t - Y_{t-1})}{M}\right], \qquad (4)$$

where M is the total number of pixels in a video frame. δ′ represents a sensitivity of the motion detection. Under then assumption that the intensity difference of patches due to the motion between adjacent frames is generally larger than that caused by temporal noise, δ′ is set to values from 2 to 4 depends on the amount of motion.

When a large motion is present around a certain pixel, SSD of its patch becomes large, hence the weight in decreases. In turn, the contribution of previous estimate is desired to be decreased to prevent motion blurs Therefore, the modified prediction and update equations of Kalman filter estimation can be derived by dividing $Q_t^{-1}$ with the proposed weight term as follows:

$$\text{Prediction:} \begin{cases} \tilde{X}_t = \hat{X}_{t-1}, \\ \tilde{P}_t = \hat{P}_{t-1} + Q_t^{-1} W_t^{-1}, \end{cases} \qquad (5)$$

$$\text{Update:} \begin{cases} \hat{X}_t = \hat{P}_t(\tilde{P}_t^{-1}\tilde{X}_t + C_t Y_t), \\ \hat{P}_t = (\tilde{P}_t^{-1} + C_t)^{-1}, \end{cases} \qquad (6)$$

$$\sigma_V^2 = \delta'' \cdot \left[\frac{(Y_t - \mu_{Y_t})^{\mathrm{T}}(Y_t - \mu_{Y_t})}{M}\right], \qquad (7)$$

Where $\mu_{Y_t}$ is the mean intensity of the curenf frame, and $\delta''$ the parameter to control the degree of filtering strength. Update equations in (6) can be rewritten as follows with a Kalman gain i.e., $K_t = \hat{P}_t C_t$

$$\hat{X}_t = (1 - K_t)\hat{X}_{t-1} + K_t Y_t. \qquad (8)$$

According to (8), observation of current frame Yt contributes more than the previous $\hat{X}_{t-1}$ estimate to the final estimate in the moving object region, thus ghost effect is minimized. On the other hand, in the static background region, previous estimates contribute sequentially to the final estimate. As a result, a significant amount of noise can be removed gradually as every new frame is injected.

**6. Tone-mapping**

After the temporal noise is reduced, dynamic range of lowlight video is required to be stretched for enhancing visibility. Various techniques for obtaining high dynamic range (HDR) image have been presented in previous research efforts. Histogram adjustment with Gamma correction is proposed in this work. Because most of pixels have very small intensity values ranging

To enhance only those pixels representing meaningful contexts of the scene, RGB histograms are stretched after clipping pixels with too low and too high intensity values. Clipping thresholds can be selected with the following rule:

$$\begin{cases} \lambda_{low,c} = \arg\max_{\lambda}(h(\lambda)), \\ \lambda_{high,c} = \arg\min_{\lambda}\left(\sum_{x=0}^{\lambda} h(x) \geq \alpha \cdot M\right), \end{cases} \quad \text{for } c \in \{r, g, b\}, \qquad (9)$$

where $\lambda_{low,c}$ and $\lambda_{high,c}$ are low and high thresholds for each color channel, respectively. h(x) is the histogram for the normalized intensity value x.

The intensity value below which the majority of pixels in video frame are accumulated is selected as the high clipping threshold. α represents the proportion of total pixels that are accumulated below a chosen high threshold. It is empirically set to values from 0.99 to 0.999 depends on lighting conditions. With these selected thresholds, low and high intensity values of an input frame are truncated as follows:

$$\begin{cases} \hat{X}_{trc,c}(i,j) = \lambda_{low,c}, & \text{if } \hat{X}_c(i,j) \leq \lambda_{low,c}, \\ \hat{X}_{trc,c}(i,j) = \lambda_{high,c}, & \text{if } \hat{X}_c(i,j) \geq \lambda_{high,c}, \end{cases} \qquad (10)$$

$\hat{X}_{trc}$ is the truncated frame. Then each pixel in the truncated frame is transformed by the proposed tone-mapping operator, T[·] as follows:

$$\begin{aligned} \hat{X}_{T,c}(i,j) &= T[\hat{X}_{trc,c}(i,j)] \\ &= \left(\frac{\hat{X}_{trc,c}(i,j) - \lambda_{low,c}}{\lambda_{high,c} - \lambda_{low,c}}\right)^{\gamma}, \quad 0 \leq \gamma < 1, \end{aligned} \qquad (11)$$

**7. Spatial Noise Reduction**

Even though a large amount of noise is attenuated by the proposed temporal noise reduction filter, the remaining noise can be exaggerated by the tone-mapping step. Unlike any other video enhancement techniques that utilize a spatiotemporal filter, the proposed method only applies a temporal filter for noise removal before the tone-mapping step. Therefore, an additional filtering in spatial domain is desired to reduce remaining noise. In this paper, the classical NLM denoising filter is modified for the Bayer pattern CFA image. Firstly, pixels of each color channel are smoothed separately with a modified Gaussian mask as in Fig. 5 to alleviate the adverse effect of the amplified noise when measuring the similarity between neighboring and reference patch. Also, only neighboring patches with the same pattern as a reference patch are considered to avoid any faulty inter-color similarity computations. Thus the final enhanced video frame $\hat{X}_E$ obtained as the NLM estimate:

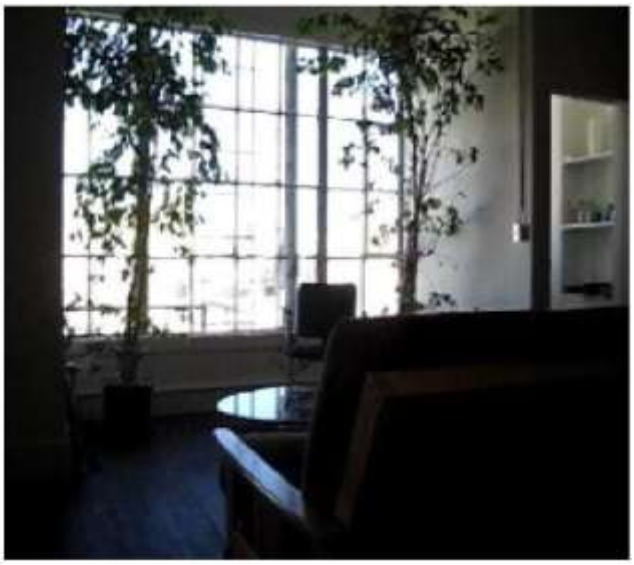
$$\hat{X}_E(i_0, j_0) = \sum_{(i,j)\in s(i_0,j_0)} w(i_0, j_0, i, j)\hat{X}_T(i_0, j_0), \qquad (12)$$

$$w(i_0, j_0, i, j) \propto \exp\left\{-G_\rho * \left\| R_{i,j}\hat{X}_T - R_{i_0,j_0}\hat{X}_T \right\|_2^2 \Big/ \beta\hat{\sigma}_T^2\right\}, \qquad (13)$$

To minimize computation complexities,

the number of similar patches that contribute to NLM estimate is limited to a certain level

## 8. RESULT ANALYSIS:



## 9. CONCLUSION:

The method also includes a sharpening feature which prevents the most important object contours from being over-smoothed. Most parameter scan be set generally for a very large group of input sequences. These parameters include: the clip-limit in the contrast-limited histogram equalization, the maximum and minimum widths of the filtering kernels and the width of the isotropic smoothing of the structure tensor and in the gradient calculations.

## REFERENCES

[1] S. W. Lee, V. Mail, J. Jang, J. Shin, and J. Paik, "Noise-adaptive spatio-temporal filter for real-time noise removal in low light level images,"IEEE Trans. Consumer Electron., vol. 51, no. 2, pp. 648-653, May. 2005.

[2] E. Bennett and L. McMillan, "Video enhancement using per-pixel virtualexposures," ACM Trans. Graphics, vol. 24, no. 3, pp. 845-852, Jul. 2005.

[3] H. Malm, M. Oskarsson, E. Warrant, P. Clarberg, J. Hasselgren, and C.Lejdfors, "Adaptive enhancement and noise reduction in very low light-level video," In Proc. IEEE International Conference on ComputerVision, Rio de Janeiro, Brazil, pp. 1-8, Oct. 2007.

[4] Q. Xu, H. Jiang, R. Scopigno, and M. Sbert, "A new approach for verydark video denoising and enhancement," In Proc. IEEE International Conference on Image Processing, Hong Kong, China, pp. 1185-1188, Sept. 2010.

[5] 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 Proc. IEEE International Conference on Multimedia and Expo , Barcelona, Spain, pp.1-6, Jul. 2011.

[6] X. Zhang, P. Shen, L. Luo, L. Zhang, and J. Song, "Enhancement andnoise reduction of very low light level images," In Proc. International Conference on Pattern Recognition, Tsukuba, Japan, pp. 2034-2037, Nov. 2012.

[7] M. Kim, D. Park, D. K. Han, and H. Ko, "A novel framework forextremely low-light video enhancement," in Proc. IEEE InternationalConference on Consumer Electronics, Las Vegas, USA, pp. 95-96, Jan. 2014.

[8] A. Loza, D. Bull, and A. Achim, "Automatic contrast enhancement of low-light images based on local statistics of wavelet coefficients," In Proc. IEEE International Conference on Image Processing , Hong Kong, China, pp. 3553-3556, Sept. 2010