A NEW APPROACH FOR VERY DARK VIDEO DENOISING AND ENHANCEMENT

Qing Xu\textsuperscript{1,2}, Hailin Jiang\textsuperscript{1}, Riccardo Scopigno\textsuperscript{3}, and Mateu Sbert\textsuperscript{4}

\textsuperscript{1} School of Computer Science and Technology, Tianjin University, Tianjin 300072, China
\textsuperscript{2} Department of Electronics, Politecnico di Torino, 10129 Torino, Italy
\textsuperscript{3} Istituto Superiore Mario Boella, 10129 Torino, Italy
\textsuperscript{4} Graphics and Imaging Laboratory, Universitat de Girona, 17071 Girona, Spain

ABSTRACT

The paper presents a novel three-stage algorithm for very-low-light video denoising and enhancement. The proposed technique invokes twice, in the first and the third stages, the well-known Non-Local Means (NLM) method for spatial and temporal denoising: it is well adapted to the application, leading to the definition of a novel NLM tool. The intermediate stage performs a custom tone adjustment specifically aimed at enlarging the dynamic range of very dark videos. The overall approach transforms very dark videos into more watchable ones, effectively reduces very high noise, and all in all, produces high quality restored image sequences outperforming the recent state-of-the-art results. Additionally, the first and third stages can be combined as a two-step filtering scheme for normal-light videos: the novel denoising solution achieves a heavy noise removal, while reducing motion blur artifacts and preserving image details.

Index Terms— Video signal processing, Very low light video, Non-local means, Tone mapping, Noise reduction.

1. INTRODUCTION

Noise removal and video enhancement play a critical role in many applications – such as surveillance – involving videos taken under very poor light conditions: they set a very challenging problem due to poor dynamic range and high noise level. While processing of very dark videos is expected to benefit from the adoption of the most flexible available algorithms, their specific adaptation to the case of low dynamic range videos remains largely untouched.

Quite a few video denoising algorithms have been proposed in literature: among them Non-Local Means (NLM) \cite{1} and its extensions \cite{2,3} are examples to behave very well under several conditions – potentially also with dark videos. Tone mapping is an additional topic relevant to the issue being investigated: it has been an active research field \cite{4} and a complementary solution, aimed at widening the low dynamic range of images by transforming their intensities. Finally, few reported attempts have been targeted specifically on very-low-light video denoising and enhancement: recently some related techniques \cite{5,6} have achieved very encouraging results.

In this paper a three-stage approach is proposed: its target is to broaden the very low dynamic range of very dark video and to remove the extremely high noise, while preserving motion and image edges. The proposed approach adopts the existing video processing tools and leads to their \textit{ad hoc} adaptation by the definition of a new and computationally lighter space-time NLM. The comparative analysis of results shows a large improvement upon the state-of-the-art methods \cite{5,6}. As a corollary of the three-stage approach, a reduced solution, shortcutting first and third stages, is introduced to handle heavy noise for normal light-level video. Despite this solution involves two logical steps, the overall scheme results in a better denoising performance and also in a more efficient process, running 20 times faster than the original NLM \cite{1}.

The remainder of the paper is structured as follows: in sect. 2, the three-stage method is detailed. In sect. 3, the rationale of the two-step scheme is briefly described. Results are discussed in sect. 4 and sect. 5 concludes the paper.

2. THE THREE-STAGE PROCESSING SCHEME

The rationale of the three-stage denoising and enhancing can be summarized as follows.

– Since noise in very dark videos gets amplified by tone mapping \cite{5,6}, the approach first starts with a spatial-temporal filtering aimed at noise attenuation. Notably this stage leads to a novel and computationally lighter NLM method: spatial and temporal integrals are separately done in the smaller filter windows, and then \textit{a posteriori} merged by adaptive weights which are dependent on the extent of motion in the video, to prevent motion blur and to preserve image edges.

– Then the approach invokes a custom tone mapping technique which accomplishes the task of widening the low dynamic range of the only areas where it is required.

– Finally the approach does filtering in the YCrCb color space...
Fig. 1. A frame depicting a very dark video “Hand” (by a Samsung camera SNC-B2315) processed by the proposed three-stage algorithm and, for the purpose of comparison, by [5] and [6]. From left to right, top to bottom: original, after space-time filtering (sect. 2.1), after tone mapping (sect. 2.2), after the processing of sect. 2.3 – final by proposed (a profit on the fast moving fingers is clear), by [6] and by [5].

for the left and raised noise, to enhance the visual appearance of the final results. In order to keep the computational load bounded, classic 3D-NLM [1] is performed only on the luminance component together with the use of smaller search window, while the classic Yaroslavsky neighborhood image filter [1] separately processes the Cr and Cb channels.

2.1. First stage — spatial and temporal filtering

The static 3D NLM for movie denoising [1] is defined as

\[ NLM_S[u(\vec{x}, t)] = \frac{1}{\mathcal{G}} \int \int G_{\rho}(\vec{x}, t), \]

where \( u(\vec{x}, t) \) is the observed value at a 2D pixel index \( \vec{x} \) in frame \( t \), \( C(\vec{x}, t) \) is the normalizing factor, \( h \) is the decay parameter, and \( G_{\rho} \) is a 2D Gauss kernel of standard deviation \( \rho \). The integral is done both in the spatial search window \( \Omega_S \) (around \( \vec{x} \)) and in the temporal search window \( \Omega_T \) (around \( t \)), and the search window \( (\Omega_S, \Omega_T) \) is a 3D space. Finally

\[ (G_{\rho}(\vec{x}+\vec{s}, t)-u(\vec{y}+\vec{s})^2) \mid ds = \int^{(G_{\rho}(\vec{x}+\vec{s}, t)-u(\vec{y}+\vec{s})^2) \mid ds \]

defines a similarity function between \( u(\vec{x}, t) \) and \( u(\vec{y}, s) \), based on the similarity of their neighborhoods (\( \Omega_C \) is a 2D neighborhood domain defining the comparison window).

Although the original NLM [1] is effective in video denoising, it does not directly consider the motion character of video and this may create “ghosting” artifacts if the contribution from a pixel in the presence of motion is excessive. Additionally 3D NLM [1] manages space and time integrations tightly, in a computationally heavy way. To overcome these limitations an alternative spatio-temporal filter is here proposed: a pure spatial and a pure temporal NLM denoised results are computed and combined by proper variable weights depending on the motion content of the video. This stage is carried out in the RGB color space (independently for the three color channels), by distinct NLM filters which exploit separate versions, respectively in the space and time domains, of the similarity function:

\[ NLM_S[u(\vec{x}, t)] = \frac{1}{\mathcal{G}} \int \int G_{\rho}(\vec{x}, t), \]

\[ G_{\rho} = \int \int_{\Omega_S(\vec{x}, t)} \int \int_{\Omega_T(\vec{x}, t)} e^{-\frac{(G_{\rho}(\vec{x}+\vec{s}, t)-u(\vec{y}+\vec{s})^2+(0) \mid ds \]

where \( h_S \) and \( h_T \) are the decay parameters, and \( C_S \) and \( C_T \) are the normalization factors. In the present study \( |\Omega_S| = 11 \times 11, |\Omega_T| = 5, |\Omega_C| = 7 \times 7, h_S=h_T. Clearly the separation of \( NLM_S \) and \( NLM_T \), and the use of smaller \( |\Omega_S| \) and \( |\Omega_T| \) alleviate largely the computational cost of this stage.

Finally the space-time filter is defined by the combination of the separate filters by proper weights:

\[ NLM_{ST}[u(\vec{x}, t)] = \omega_S \cdot NLM_S[u(\vec{x}, t)] + \omega_T \cdot NLM_T[u(\vec{x}, t)], \]

where \( h \) is also a parameter (in this study \( h \) varies in [15, 30]), and the gaussianian parameter \( \rho = 1 \). \( D_S \) and \( D_T \) express respectively the average similarities of spatial and temporal neighborhoods centered around pixel \( (\vec{x}, t) \). Notice that \( D_T \) indicates how large the motion at pixel \( (\vec{x}, t) \) is (the larger the motion, the higher the \( D_T \) becomes, and vice versa). So the weights \( \omega_S \) and \( \omega_T \) are used to dynamically configure spatio-temporal filtering, and temporal filtering becomes more relevant only when motion is small. As a result, averaging of the two separate filters with adaptive weights contributes to preventing motion blur and blurring of image edges.

2.2. Second stage — tone mapping

The goal of the second stage is to enlarge the dynamic range of dark image areas and meanwhile to keep the other areas slightly affected. The tone mapping technique is operated on luminance. For each pixel’s luminance \( I \), the tone mapped
luminance $I_{TMO}$ is given by

$$I_{TMO}=TLOG \cdot TEXP \cdot I_{\text{max}},$$

$$TLOG=\frac{\log(I+1)}{\log(I_{\text{max}}+1)},$$

$$TEXP=\frac{\log 10}{\log(5+((I+1)/I_{\text{max}}))},$$

where $I_{\text{max}}$ is the maximum luminance considered (set to 255), and $b$ is a parameter (more details follow). Equation (5) follows the basic logarithmic approach in image processing, while (6) embodies the nonlinear adjustment of the dynamic range, as inspired by [4]. Fig.2 illustrates the different mapping functions $TLOG \cdot TEXP$ for $b$ varying in the range $[0.6, 2]$. In this paper, according to the inspiration from [4], lower slopes of mapping functions for low intensities are preferred and $b = 1.25$. Especially considering the contrast enhancement of dark image areas, a different parameter is used in (5) instead of $I$; the parameter adopted is: $I_a = I + a(I - I_{\text{avg}})$, being $I_{\text{avg}}$ the average luminance of the neighborhood window of the considered pixel (the window is set $21 \times 21$ – pixel wide). $I_a$ helps enhance contrast between the interest pixel and its neighborhood and sharpen the images; $a = 0.15$.

\section{Third stage — filtering in the YCrCb space}

After the first two stages, it is necessary to perform a new filtering, because the noise left after the spatial and temporal filtering is raised by tone mapping. In order to enhance the visual color appearance of the final results, the third stage works in the YCrCb color space. The YCrCb color space is chosen because it is good at separating luminance from chrominance in the YCrCb color space. The YCrCb color space is chosen because it is good at separating luminance from chrominance and, additionally, it is linear with the RGB color space.

In the Y channel the 3D NLM [1] is utilized for video filtering: the smaller search window and the comparison window are fixed as $11 \times 11 \times 3$ and $7 \times 7$ respectively, and the standard deviation of Gauss kernel is set to 1. In the Cr and Cb channels Yaroslavsky neighborhood filter [1] is employed for 2D image filtering, the spatial neighborhood is fixed as $21 \times 21$ and the standard deviation of Gauss kernel is in $[10, 30]$.

\section{The two-step filtering scheme}

Considering the difficult problem of fighting heavy noise (such as high variance Gaussian noise, mixture of Gaussian and impulse noise, and high real-world noise), the combination of the new space-time filter proposed in 2.1 and the filtering process used in 2.3 is an appropriate and reasonable two-step approach. Actually the new space-time filter has a good ability to reduce strong noise and, more importantly, to prevent blurring artifacts; so it can produce smooth restored results with satisfying motion and image edges. As for the noise left after the first step, it is probably due to the extremely noisy input: the second step is then necessary to achieve more acceptable final results. Notably the proposed scheme can run much faster than the original NLM [1], as already stated. Due to space limit this scheme is mainly discussed by the results it yields.

\section{Results}

The proposed methods produce convincing results on a lot of videos, including real videos with fast moving objects and the standard test videos. Results are obtained on a Windows PC with Intel Centrino Duo 1.8 GHz CPU and 3GB RAM.

Handling very dark videos, $h_S$ and $h_T$ vary in $[10, 30]$, and the filtering parameter of the Y channel 3D NLM varies in $[10, 40]$. All the settings used in the two competitive methods [5, 6] are tuned to obtain their best possible results. Fig.1 depicts the entire three-stage algorithm. The space-time filter denoises while preserving motion and image edges; tone mapping operator stretches significantly the dynamic range of
the image; noise reduction in the YCrCb color space enhances the visual appearance of the final image. Fig.1 and Fig.3 show the largely better results by the proposed technique. Table 1 lists the computing time. Although the runtime of the proposed technique is worse than that achieved by [5], it is still manageable using modern hardware as suggested in [6].

Fig. 4. Comparison between two-step scheme and NLM for a movie at resolution $616 \times 356 \times 150$. From left to right, top to bottom: original, noisy (mixed Gaussian noise of $\sigma = 30$ with 30% random-valued impulse noise, PSNR=19.61 dB), by NLM (PSNR=28.02 dB, runtime=6293 sec), and by proposed (PSNR=31.28 dB, runtime=294 sec; better motion (birds) and details (background) are clear).

Handling normal light-level videos, $h_S$, $h_T$ and the filtering parameter of the Y channel NLM vary in $[40, 80]$. For original NLM the search and the comparison windows are $21 \times 21 \times 21$ and $7 \times 7$ respectively, as originally suggested [1]; filtering parameter is tuned to obtain its best possible results. Fig.4 and Fig.5 demonstrate clearly the improved results by the proposed two-step scheme. Table 2 provides details on the average PSNR and runtime comparisons for a few of standard sequences artificially noise-corrupted (with heavy Gaussian noise of $\sigma \geq 30$). All the PSNR data achieved by the two-step algorithm are better than those by NLM (with exception of “Flower garden”, containing high motions which can be compensated by methods suggested in [2]). Notably, the proposed scheme can run at least 20 times faster than NLM.

5. CONCLUSIONS

A new framework for very dark videos denoising and enhancement has been introduced and shown to largely improve current state-of-the-art results. The same framework can effectively handle normal light-level videos too. The proposed approach can be further speeded up and this constitutes the authors’ future work.

6. REFERENCES


