Paper: jc20-4-8010:

Temporal-Spatial Filtering for Enhancement of Low-Light Surveillance Video

Fan Guo, Jin Tang, Hui Peng, and Beiji Zou[†]

School of Information Science and Engineering, Central South University
Changsha, Hunan 410083, China
[†]Corresponding author, E-mail: bjzou@csu.edu.cn
[Received July 7, 2015; accepted May 24, 2016]

A new surveillance video enhancement method is proposed to improve the visual effect of videos captured in low-light conditions. The proposed technique, called temporal-spatial (TS) filtering, uses adaptive temporal filtering and nonlocal mean filtering to smooth the transmission map in the temporal and spatial dimensions and thus yields restored video sequences with significantly reduced noise, improved details and good spatial and temporal coherence. The main advantage of this work is that the performance of contrast enhancement, noise reduction and temporal-spatial coherence can be significantly improved using the proposed framework, which adopts a strategy that applies the same transmission map to a series of video frames. Comparative study and quantitative evaluation demonstrate that the proposed method is better than previous techniques in terms of reducing noise and improving contrast.

Keywords: enhancement, low-light, transmission map, noise reduction, contrast improvement

1. Introduction

Most outdoor vision applications, such as vision-guided robotic cameras, mobile camera devices, real-time surveillance, security, and navigation cameras, require the robust detection of image features. However, these camera systems capture the input video in different complex lighting environments, and the input image frames of these videos are likely to possess many lighting variations in different regions of a single image frame. Therefore, they usually have poor visual quality and lack fine features due to the limitation of the capturing devices or improper illumination or exposure conditions. The human eye can adjust to light variations and change its mode of operation while adapting locally with the decrease in light intensity from day to night. This is in contrast to cameras, where the exposure is set globally [1]. Besides, sensor noise is a common phenomenon and, in practice, is also a significant issue in lighting enhancement. Therefore, the ultimate aims of this paper are dealing with the common defects appearing in real-life videos: (1) a dark video due to poor illumination, (2) a video with an underexposed foreground and background, and (3) a low-light video with sensor noise.

There are two ways to improve the visual effects of the low light video, one is using infrared sensors and the other is using image processing techniques. For the first kind of methods, although the infrared sensors used in these methods [2] can greatly enhance the low-light visibility, they require that visible objects must have a temperature higher than that of their surroundings, and the infrared systems are usually more expensive and consume more power than conventional image enhancement algorithms. Thus, a more feasible method is the image processing methods.

There are two distinct problems for low light enhancement using image processing methods, one is the low-light noise and the other is the poor contrast. Noise suppression is essential prior to contrast enhancement. Many works have been done to solve the two problems. For noise reduction, Malm et al. [3] presented a methodology for adaptive enhancement and noise reduction for very dark image sequences with very low dynamic range. Chatterjee et al. [4] proposed a framework for combined denoising and demosaicing of low-light images, specifically targeted at reducing the splotches and noise effects in such images. Yin et al. [5] proposed a novel framework for low light image enhancement and noise reduction. In the method, noise reduction and contrast enhancement are performed in different color spaces. He et al. [6] proposed a guided filter based on the local linear model. The filter performs well in detail enhancement with relatively low computational complexity. However, the color of their enhancement results is not visually pleasing in some cases. For contrast enhancement, Retinex algorithm [7] is the traditional method to improve image visual effect when lighting conditions are not good. Besides, Wang et al. [8] presented an image enhancement method based on weighted and threshold histogram equalization. Sheet et al. [9] proposed brightness-preserving dynamic fuzzy histogram equalization, which uses the fuzzy statistics of the input images to improve their brightness-preserving and contrast enhancement abilities. Panetta et al. [10–12] proposed a novel multi-histogram equalization method that utilizes the human visual system (HVS) to segment the image, which allows a fast and efficient

correction of non-uniform illumination. Although these methods may effectively enhance various degraded images and appropriate for different cases, they are unreliable when dealing with various noises in low-light frames. Moreover, the above enhancement methods all deal with single image, thus they can only be used to process each frame individually without considering temporal and spatial coherence. Thus, they may produce flickering in the restored video sequences.

To overcome these problems, a strategy is adopted for the lighting enhancement problem in this paper. That is, one transmission map is applied to a series of video frames without creating significant errors in the enhanced video. Compared with existing frame-by-frame methods, the proposed method calculates the transmission map only once for the whole video sequences, so the lighting enhancement speed is fast. The primary contributions of this work are as follows: (i) A lighting enhancement (LE) model is built for restoring low-light video, (ii) a temporal-spatial filtering (TSF) approach is proposed for maintaining the temporal and spatial coherence in enhanced video frames, and (iii) an image degradation model combined with a nonlocal mean filter is used to achieve both noise reduction and contrast enhancement. Experimental results demonstrate that the better quality results can be obtained by using the proposed algorithm.

2. Proposed Model

The lighting enhancement (LE) model as a conceptual model is proposed for analyzing and enhancing low-light video. Poor lighting scenarios often lead to low-light or underexposed videos in surveillance applications.

2.1. Pixel-Wise Inversion Prior

Although the light captured by the camera blends with atmospheric light in hazy, while low-light conditions do not necessarily involve atmospheric scattering of light rays, and it is dark mainly due to the lack of photons differently from hazy conditions, we can't find any similarity between hazy image and low-light image from the perspective of physical mechanism. However, a prior called "pixel-wise inversion prior" can be observed from the perspective of image appearance or our experience. The pixel-wise inversion prior is a kind of statistics of the low lighting images. It is based on the key observation — most results of the pixel-wise inversion of low-light images look similar to hazy images.

A low-light color image set from several image search engines is collected to verify how good the pixel-wise inversion prior is. Experiment results show that the inverted low-light images are visually and statistically similar to hazy images. An illustrative example is shown in **Fig. 1**. One can clearly see that the images captured in hazy conditions are similar to the inverted output of low-light images. Here, the "inversion" operation can be written as:

$$Iv_b^c(x,y) = 255 - I_b^c(x,y), \quad c \in \{R,G,B\} \quad . \quad . \quad (1)$$

Fig. 1. Comparison between inversion results and hazy images. (a) Original low light images. (b) Inverted images. (c) Hazy images.

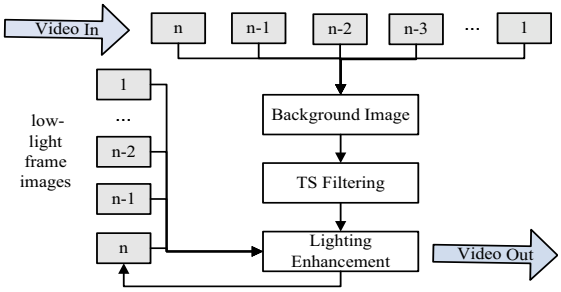


Fig. 2. ULE model for processing low-light video.

where c denotes the three RGB color channels, $Iv_b^c(x,y)$ is the intensity of the inverted image at pixel (x,y) and $I_b^c(x,y)$ is the intensity of the input color image. We call the above statistical observation or knowledge the pixel-wise inversion prior.

2.2. Enhancement Model

After inversion operation, low-lighting videos have a strong similarity to hazy videos. Therefore, the image degradation model widely used in image dehazing can be applied to low lighting enhancement. The image degradation model can be expressed as follows [13, 14]:

$$Iv^c(x,y) = Jv^c(x,y)t(x,y) + A\,[1 - t(x,y)] \quad . \quad . \quad (2)$$

where Iv^c is the observed intensity and input image, Jv^c is the intrinsic intensity of the object and enhanced image, t is the medium transmission (also called transmission map) describing the portion of the light that is not scattered or reaches the camera, and A is the airlight, generally assumed to be globally constant. Lighting enhancement aims to recover Jv^c, A and t from Iv^c.

The low-light video processing approach using the proposed LE model is presented in **Fig. 2**. Combined with the image degradation model (see Eq. (2)), temporal-spatial filtering is implemented as a way to improve the temporal-spatial coherence of the input low-lighting video. As can be seen in **Fig. 2**, in this framework all frames are added to the filter at the same time and a background image is first estimated with knowledge of the whole video sequences. Then, various noises are eliminated by the nonlocal filter in the temporal and spatial domains. Therefore, the trans-

mission map obtained by the TSF can benefit from information in the whole video frames while also ensuring that low-lighting enhancement is temporally consistent. One can apply the same transmission map to any video frame without creating significant errors in the final enhanced frame.

Given a low-light video, the image degradation model and nonlocal filtering are applied to improve poorly exposed areas and handle the noise caused by sensors. This enhancement algorithm is discussed in Section 3.

2.3. Surveillance Video Characteristics

In a surveillance camera system, the camera is usually fixed in a higher location, so the background of each frame is unchangeable, and the difference in transmission map between a foreground object and the background behind it is usually small. For all the surveillance video that has the above characteristics, the strategy that adopting the same transmission map for each frames of the input video can be applied here. The map is suitable for all these video frames, and there is no need to estimate transmission map for each frame.

Specifically, the proposed strategy estimates the transmission map of a background image, applied to every video frame, because calculating the transmission map for each frame is time consuming. Generally, the errors in a transmission map do not create significant errors in the restored video frame. To calculate the transmission map that can be used for all instances, this map should contain only an outlier of the scene. Therefore, moving foreground objects and image details can be regarded as noise. As the fast nonlocal mean filter [15] is efficient for image de-noising, the filtering is thus conducted to calculate the transmission map. We will discuss this de-noising filter in the following section.

3. Proposed Algorithm

The proposed temporal-spatial filtering (TSF) seeks out the background image and the transmission map to enhance lighting conditions according to the image degradation model. Two major factors affect the TSF: how to make the background image robustly extracted from the input video and how to ensure that the transmission map contains only an outlier of the scene without redundant details. TSF is implemented by transitioning between temporal and spatial filtering while choosing parameters based on the scenarios.

3.1. Temporal Filtering for Background Estimation

For simplicity, the static part of the scene is defined as the background part, and the moving object in the scene is defined as the foreground part. Given a background pixel in the n-th frame B_n, the temporally filtered background pixel can be represented by

$$B_{n+1}(i) = \alpha B_n(i) M_n + (1-\alpha) B_1(i) M_n \quad \ldots \quad (3)$$

where $B_n(i)$ is the current intensity of the background pixel at position i, $B_{n+1}(i)$is the updated background pixel intensity, $B_1(i)$ represents the first frame modeled as the initial background image, $\alpha \in [0, 1]$ is the filtering coefficient, and M_n is a binary value, that determines if the distribution adapts to local change. M_n is a key element that tunes the filter to the desired temporal frequency; thus, either a slow or fast object can be detected from the background. M_n can be defined as follows:

$$M_n = \begin{cases} 1 & \text{if } (|C_n - C_{n-1}| < Th) \\ 0 & \text{others} \end{cases} \quad \ldots \quad (4)$$

where C_n and C_{n-1} represent the intensity of the current and previous frames, respectively, and Th is a threshold. When $|C_n - C_{n-1}| < Th$ the positions of the estimated background pixels are indicated for the current frame and the first frame as the new and old background. A low updating coefficient α is then given to the new background, and high α is given to the old background. The two parameters Th and α are given by experiences. For updating coefficient α, slow background updating speed is described by small value of α, and quick speed is described by high value. In our experiments, α is fixed to 0.01 to make a trade-off between stability and quick update. The value of Th is application-based, and we set it to 35 for all the images reported in this paper. For other images, the value of Th should be estimated according to the practical conditions to prevent artifact "trails" forming behind moving objects.

The advantage of the temporal filtering is its capability to quickly respond to the background change without affecting foreground objects. The filter is continuously updated with time, and the background image in accordance with the practical situation can then be obtained.

3.2. Nonlocal Spatial Filtering for Noise Reduction

To reduce noise, fast nonlocal image filtering [15] is used in the proposed TSF. The filtering builds on the separable property of neighborhood filtering to greatly reduce computational complexity. Nonlocal filtering for a pixel s with a non-negative weight w is as follows:

$$u(s) = \frac{1}{Z(s)} \sum_{t \in N(s)} w(s,t) v(t) \quad \ldots \quad (5)$$

$$w(s,t) = g_h\left(S_d(s+P) - S_d(s-P)\right) \quad \ldots \quad (6)$$

$$S_d(p) = \sum_{k=0}^{p} \left(v(k) - v(k+d)\right)^2, \quad p \in \Omega \quad \ldots \quad (7)$$

where v is the original noisy image, u is the restored image. $Z(s)$ is the normalization constant, $N(s)$ corresponds to a set of neighboring sites of s, $s \in \Omega$ (Ω is a discrete regular grid), g_h is a continuous non-increasing function, S_d corresponds to the discrete integration of the squared difference of image v and its translation by d, and parameters p and k are used to control the smoothness of the result. **Fig. 3** shows the relationship between the parameter values and the smooth effect when either parameter

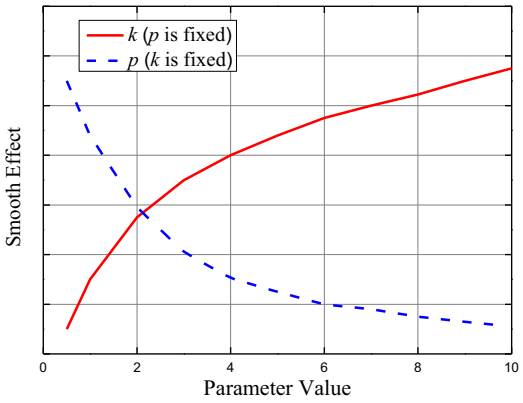


Fig. 3. A plot showing the relationship between parameter value and smooth effect.

(p or k) is fixed. The smaller the value of p or the higher the value of k, the smoother the filtering result will be. The values of p and k are set to 5 and 10, respectively, for all the transmission map estimation results reported in this paper. That's because the value of p indicates the neighborhood size for weight computation, and parameter k indicates the size of the searched region. Results on a large quantity of testing images with a size of 640×480 demonstrate that when the neighborhood size for weight computation is set to 5 and the size of the searched region is set to 10, a relatively good image smooth effect can be obtained. The two parameter values for the images with different sizes should be estimated according to the practical conditions.

With the parameter values, nonlocal filtering can effectively smooth out the redundant details and various noises while preserving edges. Therefore, this method can be used in transmission map estimation and in the noise reduction of the final results.

3.3. Implementing TSF

The TSF is proposed to enhance low-light videos while preserving temporal and spatial coherence. TS filtering is described in detail in the following steps.

Step 1. The temporal filter (Eqs. (3) and (4)) is adopted to extract the background image $I_b^c(x,y)$ from the input low lighting video and airlight A in Eq. (2) is estimated for the extracted background image. Inspired by the defogging algorithm proposed by Loza et al. [16], we first calculate the saturation component S of the background image.

$$S(x,y) = 1 - \frac{\min\limits_{c\in\{r,g,b\}} I_b^c(x,y)}{\max\limits_{c\in\{r,g,b\}} I_b^c(x,y)} \quad \ldots\ldots\ldots \quad (8)$$

where $I_b^c(x,y)$ is the background image, c is the three color channel. Then the intermediate image $D(S)$ is inferred as:

$$D(S) = \min_{(k,l)\in P(x,y)} (1 - S(k,l)) \quad \ldots\ldots\ldots \quad (9)$$

In Eq. (9), $P(x,y)$ is a local patch centered at (x, y), and the size of $P(x,y)$ is 3×3. Thus, the airlight A^c is estimated as

Fig. 4. Example of transmission map estimation. (a) Extracted background image. (b) Inverted background image. (c) Estimated initial transmission map. (d) Estimated final transmission map.

a color vector of a pixel in background image $I_b^c(x,y)$ with maximum intensity at location corresponding to a small subset of the highest intensity pixels of the dark channel $\{D_s(S)\}$. The process can be written as:

$$A^c = \max I_b^c(x,y), \quad (x,y) \in \{x_{D_s(S)}, y_{D_s(S)}\} \quad . \quad (10)$$

Step 2. The initial transmission map is obtained using the estimation method in haze removal algorithm. According to the pixel-wise inversion prior, the inverted low-light images are visually and statistically similar to hazy images. Therefore, the image degradation model widely used in haze removal can be used here for lighting enhancement.

Specifically, the "inversion" operation (1) is first applied to the extracted background image of a surveillance video. Therefore, the pixel intensities for the inverted background image $Iv_b^c(x,y)$ can be obtained. According to the image degradation model (2) and the dark channel prior [14], the transmission map can be deduced as:

$$\tilde{t}(x,y) = 1 - \omega D\left(\frac{Iv_b^c(x,y)}{A^c}\right) \quad \ldots\ldots\ldots \quad (11)$$

where parameter ω ($0 < \omega \leq 1$) is introduced to adjust output luminance. The higher the value of ω is, the brighter the enhanced result. In our experiments, ω is set to the average value of the dark channel to make it scene-dependent, that is $\omega = meanD(S)$.

Step 3. As the moving foreground object and image details of a surveillance video can be regarded as noise, to calculate the transmission map just once and use it for all, the map should contain only an outlier of the scene. To this end, the fast nonlocal spatial filter (Eqs. (5)–(7)) is applied to estimate the transmission map based on the result obtained in Step 2. **Fig. 4** shows an example of the transmission map estimation. A good balance between feature blurriness and edge preservation can be achieved with the filter to generate a robust map and improve the spatial and temporal coherence.

3.4. Scene Lighting Enhancement

For lighting enhancement, according to the proposed strategy, the scene lighting of each low-light frame can be enhanced with the same transmission map obtained by the algorithm presented above. As the term $J(x,y)t(x,y)$ in the image degradation model (2) can be very close to zero, the transmission $t(x,y)$ is restricted to a lower bound t_0. A typical value of t_0 is 0.1. The final scene $Jv^c(x,y)$

(a) (b) (c) (d)

Fig. 5. Lighting enhancement results for a video clip. (a) and (b) are arbitrarily selected input low-light frames. (c) and (d) are the lighting enhancement results obtained by our method.

can then be enhanced by

$$Jv^c(x,y) = \frac{Iv^c(x,y) - A^c}{F(x,y)\max(t(x,y),t_0)} + A^c \quad . \ . \ . \ (12)$$

$$F(x,y) = \begin{cases} \dfrac{\log\left(\frac{t(x,y)}{tMax}(\psi-1)+1\right)}{\log(\psi)}, & 0 < t(x,y) \le 0.5 \\ 1, \quad 0.5 < t(x,y) \le 1 \end{cases}$$

$$. \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ (13)$$

where $Iv^c(x,y)$ represents each inverted frame of the input–inverted low-light video. Once the transmission map $t(x,y)$ is estimated by the algorithm presented above, it can be applied to all the inverted frames. $Jv^c(x,y)$ is the enhanced frame. Parameter $F(x,y)$ is used to maintain the spatial continuity of $t(x,y)$ and visually smoothes the enhanced video. The maximum values of the transmission map $tMax$ and parameter ψ (often around 40) control the attenuation profile.

For the input low-light video, once $Jv^c(x,y)$ is computed, the enhanced result $J_f^c(x,y)$ can be produced by performing inversion again. This process can be written as $J_f^c(x,y) = 255 - Jv^c(x,y)$. The obtained enhanced images usually have amplified noise. Therefore, the last step is to reduce the noise using the fast nonlocal filter again, with p and k of approximately 5 and 3, respectively. **Fig. 5** shows the lighting enhancement results with the transmission map, **Figs. 5(a)(b)** are arbitrarily selected low-light frames. **Fig. 5** shows that the transmission map shown in **Fig. 4(d)** does not result in significant errors to enhanced frames in **Figs. 5(c)(d)**. Furthermore, the same transmission map and airlight are used to enhance video frames, making color correction unnecessary. The flow chart of the entire enhancement process is shown in **Fig. 6**.

The lighting enhancement process in **Fig. 6** begins by extracting the background image from the input low-light video. The transmission map is then estimated using the proposed strategy. Finally, the scene is recovered and noise is reduced to generate the output enhanced video. In **Fig. 6**, highlighted areas show the processing steps to preserve temporal and spatial coherence.

4. Experimental Results and Performance Evaluation

Two distinct problems for low light video are discussed in this paper. One associated with low-light noise and the other is poor contrast. For noise reduction, both non-local filtering and background modeling through temporal filtering are used to achieve this goal. Therefore, in this section, we mainly evaluate the proposed algorithm from the two aspects: noise reduction and contrast enhancement. Besides, to quantitative rate various enhancement methods, statistical evaluation and computation time are also considered here.

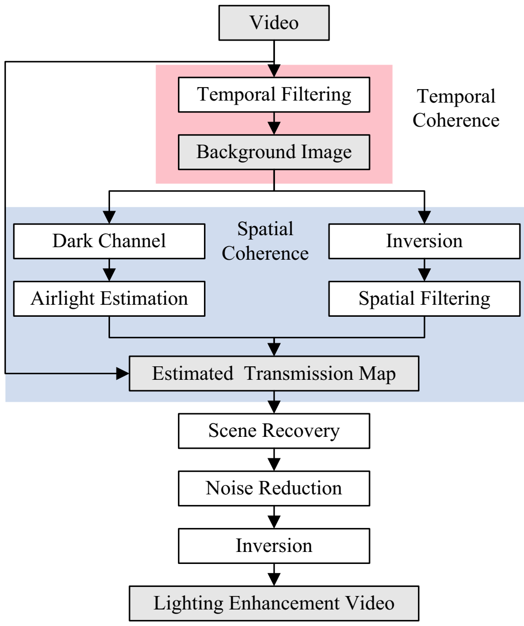


Fig. 6. Flow chart of the process for enhancing low-light video.

(a) (b) (c)

Fig. 7. Comparison with contrast enhancement algorithms. (a) Original image. (b) Enhanced by Wang [8]. (c) Enhanced by Sheet [9].

4.1. Noise Reduction Measurement

Sensor noise is a common phenomenon and significant issue in lighting enhancement. Traditional histogram-based method, such as Wang's method [8] and Sheet's method [9], share the limitation of the traditional histogram equalization. For these methods, noise is amplified with increasing contrast and brightness. An illustrative example is shown in **Fig. 7**.

From **Fig. 7**, one can clearly see that the noises in the enhanced results are also amplified with the image contrast increasing. He's algorithm [6] is a representative method for low-light image denoising and enhancement. The algorithm is mainly about edge preserving smooth, so noise filtering is an important feature for the enhancement method. We compare our method with He's work. In **Fig. 8**, one can clearly see that our algorithm produce more visual pleasing result with good de-noising capa-

(a) (b) (c)

Fig. 8. Comparison with He's work. (a) Original image. (b) Enhanced by He [6]. (c) Enhanced by our method.

(a) (b) (c)

Fig. 9. Noise reduction for a color image. (a) Original image. (b) Enhanced by Malm [3]. (c) Enhanced by our method.

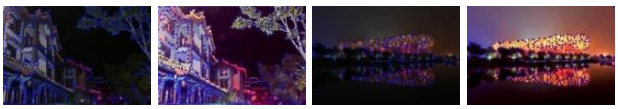
(a) (b) (c) (d)

Fig. 10. Comparison of lighting enhancement results for a video clip. (a) Original noisy frame image. (b) Local amplification region in original image. (c) Noise reduction result obtained by our method. (d) Local amplification region in our result.

(a) (b) (c)

Fig. 11. Contrast enhancement for a color image. (a) Original image. (b) Enhanced by Sheet et al. [9]. (c) Enhanced by the proposed method.

Fig. 12. Comparison of brightness enhancement results for a video clip. First row: original low contrast video frame. Second row: contrast enhancement results obtained by our method.



(a) (b) (c) (d)

Fig. 13. Harris corners comparison. Number of detected corners is mentioned in brackets. (a) Original image (547). (b) Corner detection for TS filtering (659). (c) Original image (277). (d) Corner detection for TS filtering (308).

bility. That's because the nonlocal filter integrated into the proposed method enhances the capability of the algorithm in noise reduction. However, the estimated transmission map obtained by our method is not an accurate map, which blurs or smoothes some scene object edges in the enhancement image. Nevertheless, the overall quality of the original image is largely improved by reducing the noise and enhancing the contrast, as shown in **Fig. 8(c)**.

Besides, there are many other algorithms specific for low night noise filtering that can be used for bench marking. Malm's adaptive enhancement and noise reduction method [3] is one of them. **Fig. 9** shows examples of lighting enhancement for very noisy image captured in low-light conditions using Malm's algorithm. It can be seen that the noises in the original image are all effectively suppressed by using the two methods. However, the color of our results seems more vivid than Malm's results.

Figure 10 presents the enhancement results of the proposed method for some arbitrarily selected video frames from a low-light video with sensor noise. From the local amplification picture shown in **Fig. 10**, one can clearly see that the noise contained in the original image is largely removed by the proposed method. Therefore, the results maintain a good balance of denoising smooth regions while retaining textured regions.

4.2. Contrast Enhancement Measurement

The proposed method is tested on various low-light images, since if image processing can obtain better results using the proposed method, the same is true is for video processing. Some of the results are presented in **Fig. 11**. For comparison, the simulation results of Sheet's method [9] are also presented here. It can be seen that the result obtained with our algorithm has better contrast and naturalness compared with Sheet's method [9] for low-light image.

The processing results in a video are shown in **Fig. 12**, one can clearly see that the proposed method can achieve good enhanced contrast and image details. For example, the vehicles that are hardly seen in the original frames can be clearly seen in the enhanced results.

To further verify the effectiveness of our method for contrast enhancement, some typical computer vision applications, such as Harris corner detection and face detection, are also presented here. The Harris corner detector is a popular interest point detector due to its strong invariance to rotation, scale, illumination variation and image noise. It is based on the eigenvalues of the second moment matrix [1]. **Fig. 13** depicts the number of Harris corners [17] detected on the images processed by the proposed approach. The original image shown in **Figs. 13(a)(c)** detects only a few interest points. We can

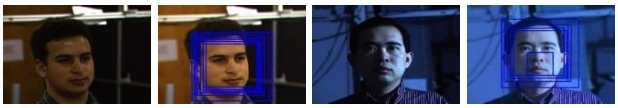

(a) (b) (c) (d)

Fig. 14. Face detection with proposed technique on CMU PIE database [19]. (a) and (c) are original images, (b) and (d) are face detection results using TS filtering.

clearly observe that **Figs. 13(b)(d)**, enhanced with the proposed algorithm, detects a higher number of corners compared to the original image especially in the brighter regions, which demonstrates that the overall image contrast of the low-light image is largely improved by the proposed method.

The proposed enhancement method is also tested on face detection in low-light conditions using the Viola-Jones face detection algorithm [18]. **Figs. 14(b)(d)** show the result of the proposed algorithm with face detection on CMU PIE database [19]. Experimental results show that the face of original image cannot be recognized directly by using the Viola-Jones method [18], but the faces are detected well in the enhanced images. Note that other more robust detection methods may recognize the face directly. However, by using the same detection algorithm here, the faces on the enhanced images of the proposed technique are newly detected after enhancement, which proves that the image contrast is significantly improved by using the proposed technique.

4.3. Statistical Evaluation

The amount of improvement after enhancement is often difficult to measure objectively. There is no measure that can specify both the objective and subjective validity of the image enhancement method. Here, two factors of the enhanced images obtained by various algorithms are considered: image contrast and image color. The enhancement algorithms used for comparison are: He [6], Sheet [9] and the proposed method.

For image contrast measurement, we use three ways to evaluate the contrast improvement level. The first one is the entropy (E) measure of original and enhanced image. The second one is a blind-reference metric—measure of enhancement (EME) proposed by Panetta et al. [20–23], and the third one is a blind contrast enhancement assessment method proposed by Hautiere et al. [24]. Image entropy (E) is a quantity which is used to describe the "business" of an image. Low entropy images, such as those containing a lot of black sky, have very little contrast and large runs of pixels with the same or similar intensity values. On the other hand, high entropy images have a great deal of contrast from one pixel to the next. Thus, a higher entropy value correlates to higher quality image. **Table 1** shows the quantitative evaluation results for the three algorithms on eight images.

From the table, one can clearly see that the entropy value of the proposed method are higher than the entropy results obtained by He and Sheet, which indicates that our method may be better than these algorithms. The EME index gives an absolute score to each image on the basis of image contrast processed with Fechner's Law [25] relating contrast to perceived brightness or the well-known entropy concept, and higher value of EME corresponding to good enhancement results. From **Table 1**, we can see that similar conclusions can be drawn with respect to EME index. The blind contrast assessment method [24] is also conducted to evaluate the performance of various enhancement algorithms. Here, the color level image is first transformed to the gray level image and one uses three indicators e, r and σ to compare two gray level images: the input image and the enhanced image. Indicator e evaluates the ability of the method to restore edges invisible in the original image but visible in the enhanced image. Indicator r is the ratio of the average gradients after and before enhancement. Finally, the number of pixels, the white ones saturated after the lighting enhancement algorithm is applied, is computed. The value is normalized by the size of the image to obtain the σ index. The three indicators are evaluated for the three algorithms. For each method, lighting enhancement aims to increase contrast without losing detailed information. Therefore, good results are described by high values of e and r and a low value of σ. We thus define an empirical score ICI, which is given by the sum of the three different indicators.

$$ICI = e + \bar{r} + 1 - \sigma \quad \ldots\ldots\ldots\ldots\ldots (14)$$

For ICI, The higher value corresponding to good enhancement results. From **Table 1**, we can deduce that, the proposed method generally has a higher value of ICI than other methods, which demonstrate similar or better quality results can be obtained using the method.

For image color measurement, a model for optimal color image reproduction of natural images [26] was introduced based on the assumption that color quality of natural images was constrained by perceived naturalness and colorfulness of these images. Here, naturalness is the degree of correspondence between human perception and reality world, which is described by color naturalness index (CNI) [27]. The bigger the value of CNI is, the more natural the color image is. Colorfulness presents the color vividness degree, which is described by CCI [26–28]. When CCI is in certain range, the color of image is suitable for human. From **Table 1**, we can also see that the proposed algorithm generally has a higher value of CNI and CCI, which demonstrates that the color of our results is generally natural and vivid. Therefore, compared to other, our algorithm generally has a better enhancement effect in terms of contrast improvement and color quality. The assessment indexes are also tested on a variety of low-light video clips for various enhancement algorithms. **Fig. 15** shows index statistical results for a real captured video clip with 100 frames. To assess the contrast improvement level, the statistical results of EME and r for the testing video are shown in **Figs. 15(a)(b)**. To evaluate the color quality, the statistical results of CNI and CCI are shown in **Figs. 15(c)(d)**. The horizontal axes are the assessment index values and vertical axes are the frame

Table 1. Quantitative evaluation.

	Contrast improvement			Color quality		Contrast improvement			Color quality	
	E	EME	ICI	CNI	CCI	E	EME	ICI	CNI	CCI
Method	**Fig. 1**					**Fig. 4**				
Input	3.36	24.84	1.01	0.59	0.81	4.82	18.16	3.39	0.66	1.13
He	4.47	26.68	8.45	0.52	2.24	4.96	23.95	8.75	0.58	1.19
Sheet	5.77	37.11	9.63	0.72	1.81	5.65	24.91	9.15	0.63	1.28
Our	6.48	40.62		0.77	2.53	7.08	27.96		0.83	1.43
Method	**Fig. 7**					**Fig. 8**				
Input	4.14	19.96	2.17	0.50	0.92	3.72	15.53	2.06	0.37	0.40
He	4.98	25.93	2.64	0.56	2.02	4.41	26.15	4.47	0.39	0.51
Sheet	5.71	30.37	3.07	0.56	1.80	5.21	29.84	5.15	0.39	0.60
Our	6.12	37.22		0.64	2.93	6.98	30.45		0.42	0.74
Method	**Fig. 9**					**Fig. 10**				
Input	3.22	7.49	4.37	0.38	0.39	3.68	15.16	3.76	0.37	0.27
He	5.09	12.54	4.52	0.45	0.47	4.25	15.10	4.79	0.37	0.28
Sheet	6.17	17.94	5.24	0.47	0.72	5.84	15.27	5.51	0.38	0.29
Our	6.75	18.58		0.50	1.97	6.97	15.39			
Method	**Fig. 11**					**Fig. 15**				
Input	2.89	11.10	4.68	0.46	1.72	2.59	7.71	3.96	0.39	0.40
He	3.07	16.58	5.46	0.58	2.04	3.75	16.72	4.19	0.42	0.43
Sheet	4.95	23.79	5.70	0.61	2.23	4.93	18.09	4.85	0.48	0.49
Our	6.63	24.84		0.66	2.31	6.54	24.46			

number index.

From **Figs. 15(a)(b)**, it can be seen that the EME and r of He's method and Sheet's method are smaller than that of our method. Since the higher the value of EME or r is, the better the enhancement effect will be. Thus, the conclusion that our method has better enhancement effect can be drawn. From **Figs. 15(c)(d)**, one can clearly see that the CNI of our results is clustered between 0.45 and 0.50, while the CNI of He's and Sheet's results are distributed between 0.42 and 0.45, 0.44 and 0.49, respectively. The higher the value of CNI is, the better the enhancement effect will be. Thus, our results have better color quality. Similar conclusion can be drawn by assessing the enhanced results using the CCI index.

Therefore, it can be deduced that the overall values of EME and CNI of our method and those of Sheet's method are very close, while the difference between the values of r and CCI of our method and Sheet's results are much greater. This indicates that, on the whole, our results are slightly better than Sheet's results in terms of perceived brightness and color naturalness, while the average visibility enhancement level and color carefulness of our results may much better than Sheet's results. Therefore, the conclusion that the proposed filter provides good performance can be confirmed.

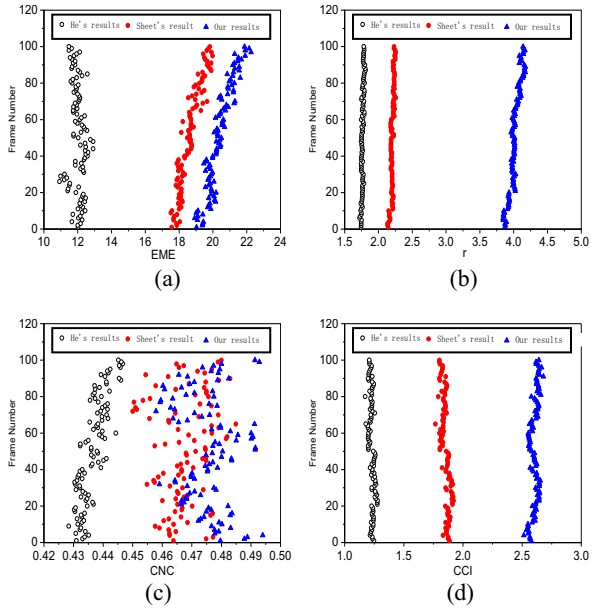


Fig. 15. Statistical results for a video clip with 100 frames. (a) EME. (b) r. (c) CNI. (d) CCI.

5. Conclusions

This paper proposes a new lighting enhancement framework using the proposed strategy and temporal-spatial filtering. Compared with other state-of-the-art algorithms, the proposed method has better contrast enhancement performance and noise reduction capability. However, it tends to degrade the quality of the transmission map and blur some scene textures. This result indicates that the transmission map generated by the proposed method is not very accurate. This inaccuracy may reduce the effect of the proposed method. For instance, when dealing with the frames of an object without a clear shape, such as trees far from the camera, the enhancement results may sometimes be high colored. Nevertheless, the proposed method may largely improve the overall quality of low-light video frames by enhancing the main detail, contrast, and color. Therefore, the proposed algorithm

would be a promising enhancement technique that can be useful in several computer vision and pattern recognition applications.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (61502537, 61573380, 91220301), Hunan Planned Projects for Key Scientific Research Funds (No. 2015WK3006), and the Postdoctoral Science Foundation of Central South University (No. 126648).

References:

[1] S. Arigela and V. K. Asari, "Self-tunable transformation function for enhancement of high contrast color images," J. of Electronic Imaging, Vol.22, No.2, pp. 023010-1-023010-22, 2013.

[2] H. Ngo, L. Tao, M. Zhang, A. Livingston, and V. Asari, "A visibility improvement system for low vision drivers by nonlinear enhancement of fused visible and infrared video," Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 25-32, 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," Proc. of the IEEE Int. Conf. on Computer Vision, pp. 1-8, 2007.

[4] P. Chatterjee, N. Joshi, S. B. Kang, and Y. Matsushita, "Noise Suppression in Low-light Images through Joint Denoising and Demosaicing," Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 321-328, 2011.

[5] W. S. Yin, X. B. Lin, and Y. Sun, "A Novel Framework for Low-light Colour Image Enhancement and Denoising," Proc. of the 3rd Int. Conf. on Awareness Science and Technology, pp. 20-23, 2011.

[6] K. M. He, J. Sun, and X. O. Tang, "Guided image filtering," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.35, No.6, pp. 1397-1409, 2013.

[7] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multi-scale Retinex for bridging the gap between color images and the human observation of scenes," IEEE Trans. on Image Processing: Special Issue on Color Processing, Vol.6, No.7, pp. 965-976, 1997.

[8] Q. Wang and R. K. Ward, "Fast image/video contrast enhancement based on weighted threshold histogram equalization," IEEE Trans. on Consumer Electronics. Vol.53, No.2, pp. 757-764, 2007.

[9] D. Sheet, H. Garud, A. Suveer, M. Mahadevappa, and J. Chatterjee, "Brightness preserving dynamic fuzzy histogram equalization," IEEE Trans. on Image Processing, Vol.56, No.4, pp. 2475-2480, 2010.

[10] K. A. Panetta, E. J. Wharton, and S. S. Agaian, "Human Visual System-Based Image Enhancement and Logarithmic Contrast Measure," IEEE Trans. on Systems, Man, and Cybernetics-Part B (Cybernetics), Vol.38, No.1, pp. 174-188, 2008.

[11] E. Wharton, K. Panetta, and S. Agaian, "Adaptive Multi-Histogram Equalization using Human Vision Thresholding," Proc. of the SPIE – The Int. Society for Optical Engineering, Vol.6497, pp. 64970G-1-11, 2007.

[12] E. Wharton, K. Panetta, and S. Agaian, "Human Visual System Based Multi-Histogram Equalization for Non-Uniform Illumination and Shoadow Correction," Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 1-729-1-732, 2007.

[13] N. Hautiere, J. P. Tarel, J. Lavenant, and D. Aubert, "Automatic fog detection and estimation of visibility distance through use of an onboard camera," Machine Vision and Applications, Vol.17, No.1, pp. 8-20, 2006.

[14] K. M. He, J. Sun, and X. O. Tang, "Single image haze removal using dark channel prior," IEEE Trans. on Pattern Analysis and Machine Intelligence. Vol.33, No.12, pp. 2341-2353, 2011.

[15] J. Darbon, A. Cunha, T. F. Chan, S. Osher, and G.J. Jensen, "Fast nonlocal filtering applied to electron cryomicroscopy," Proc. of 5th IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro, pp. 1331-1334, 2008.

[16] A. Loza, H. Bhaskar, M. Al-Mualla, and D. Bull, "Fast algorithm for restoration of foggy images," Proc. of IEEE 20th Int. Conf. on Electronics, Circuits, and Systems, pp. 735-738, 2013.

[17] C. Harris and M. J. Satephens, "A combined corner and edge detector," Fourth Alvey Vision Conf., pp. 147-152, 1988.

[18] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp. 511-518, 2001.

[19] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.25, No.12, pp. 1615-1618, 2003.

[20] S. S. Agaian, B. Silver, and K. A. Panetta, "Transform Coefficient Histogram-Based Image Enhancement Algorithms Using Contrast Entropy," IEEE Trans. on Image Processing, Vol.16, No.3, pp. 741-7558, 2007.

[21] E. Whartona, S. Agaianb, and K. Panetta, "Comparative Study of Logarithmic Enhancement Algorithms with Performance Measure," Proc. SPIE Electronic Imaging, Vol.6064, pp. 606412-1-606412-12, 2006.

[22] K. A. Panetta, E. J. Wharton, and S. S. Agaian, "Human Visual System-Based Image Enhancement and Logarithmic Contrast Measure," IEEE Trans. on Systems, Man, and Cybernetics-Part B (Cybernetics), Vol.38, No.1, pp.174-188, 2008.

[23] K. Panetta, Y. C. Zhou, S. Agaian, and H. W. Jia, "Nonlinear Unsharp Masking for Mammogram Enhancement," IEEE Trans. on Information Technology in Biomedicine, Vol.15, No.6, pp. 918-928, 2011.

[24] N. Hautiere, J. P. Tarel, D. Aubert, and E. Dumont, "Blind contrast enhancement assessment by gradient ratioing at visible edges," Image Analysis and Stereology, Vol.27, No.2, pp. 87-95, 2008.

[25] R. A. Houstoun and J. F. Shearer, "Fechner's Law," Nature, Vol.125, pp. 891-892, 1930.

[26] S. Yendrikhovskij, F. Blommaert, and H. D. Ridder, "Perceptual optimal color reproduction," Proc. of SPIE., pp. 274-281, 1998.

[27] K. Q. Huang, Q. Wang, and Z. Y. Wu, "Color image enhancment and evaluation algorithm based on human visual system," Proc. of IEEE ICASSP, Vol.3, pp.iii-721-iii-724, 2007.

[28] K. Q. Huang, Q. Wang, and A. Y. Wu, "Natural color image enhancement and evaluation algorithm based on human visual system," Computer Vision and Image Understanding, Vol.103, No.1, pp. 52-63, 2006.

Name:
Fan Guo

Affiliation:
Lecturer, School of Information Science and Engineering, Central South University

Address:
Changsha, Hunan 410083, China

Brief Biographical History:

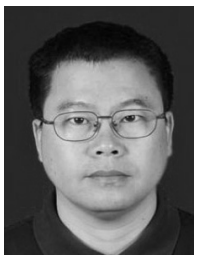
2005 Received B.S. degree from Central South University
2008 Received M.S. degree from Central South University
2012 Received Ph.D. degree from Central South University

Main Works:

• "Adaptive Estimation of Depth Map for Two-Dimensional to Three-Dimensional Stereoscopic Conversion," Optical Review, Vol.21, No.1, pp. 60-73, 2014.

Membership in Academic Societies:

• Member, the Institute of Electrical and Electronics Engineers (IEEE)
• Member, China Computer Federation (CCF)
• Member, Chinese Association for Artificial Intelligence (CAAI)



Name:
Jin Tang

Affiliation:
Professor, School of Information Science and Engineering, Central South University

Address:
Changsha, Hunan 410083, China

Brief Biographical History:
1987 Received B.S. degree from Peking University
1990 Received M.S. degree from Peking University
2002 Received Ph.D. degree from Central South University
2003-2004 Postdoctoral Researcher, National University of Defense technology

Main Works:
• "2.5D Multi-View Gait Recognition Based on Point Cloud Registration," Sensors, Vol.14, No.4, pp. 6124-6143, 2014.

Membership in Academic Societies:
• Member, Chinese Association for Artificial Intelligence (CAAI)

Name:
Hui Peng

Affiliation:
Professor, School of Information Science and Engineering, Central South University

Address:
Changsha, Hunan 410083, China

Brief Biographical History:
1983 Received B.S. degree from Central South University
1986 Received M.S. degree from Central South University
2003 Received Ph.D. degree from the Graduate University for Advanced Studies, Japan

Main Works:
• "Nonlinear predictive control using neural nets-based local linearization ARX model – Stability and industrial application," IEEE Trans. on Control Systems Technology, Vol.15, No.1, pp. 130-143, 2007.
• "A modeling approach to financial time series based on market microstructure model with jumps," Applied Soft Computing, Vol.29, pp. 40-51, 2015.
• "Multivariable RBF-ARX model-based robust MPC approach and application to thermal power plant," Applied Mathematical Modeling, Vol.35, No.7, pp. 3541-3551, 2011.

Membership in Academic Societies:
• Member, the Institute of Electrical and Electronics Engineers (IEEE)

Name:
Beiji Zou

Affiliation:
Professor, School of Information Science and Engineering, Central South University

Address:
Changsha, Hunan 410083, China

Brief Biographical History:
1978-1982 B.S. degree from Zhejiang University
1982-1984 M.S. degree from Tsinghua University
1997-2001 Ph.D. degree from Hunan University
2001-2003 Postdoctoral Researcher, Tsinghua University

Main Works:
• "Enhanced hexagonal-based search using direction-oriented inner search for motion estimation," IEEE Trans. on Circuits and Systems for Video Technology, Vol.20, No.1, pp. 156-160, 2010.
• "A novel particle filter with implicit dynamic model for irregular motion tracking," Machine Vision and Applications, Vol.24, No.7, pp. 1487-1499, 2013.
• "Motion recognition for 3D human motion capture data using support vector machines with rejection determination," Multimedia Tools and Applications, Vol.70, No.2, pp. 1333-13625, 2014.

Membership in Academic Societies:
• Member, the Institute of Electrical and Electronics Engineers (IEEE)
• Member, China Computer Federation (CCF)