Universal strategy for surveillance video defogging

Bin Xie
Fan Guo
Zixing Cai
Central South University
School of Information Science and Engineering
Changsha 410083, China
E-mail: xiebin@csu.edu.cn

Abstract. We present a new approach to remove haze from surveillance video sequences. This approach extracts the background image through the frame differential method, uses the dark channel prior to estimate the atmospheric light, and then calculates a universal transmission map on the intensity component of the background image through a process of multiscale retinex, parameter adjustment, bilateral filtering, and total variation denoising filtering. Finally, it renders haze-free video according to the haze image model. The main advantage of the proposed approach is its speed as this approach adopts a "universal strategy" that applies the same atmospheric light and a universal pseudo-transmission map to a series of video frames. Experiments demonstrate that our method produces visually pleasing defogging results and tends to preserve main details better than previous techniques. A comparative study and quantitative evaluation show the efficiency of the proposed method.
 [DOI: 10.1117/1.OE.51.10.101703]

Subject terms: video defogging; image haze model; universal strategy; transmission map.

Paper 111303SS received Oct. 19, 2011; revised manuscript received Dec. 7, 2011; accepted for publication Jan. 5, 2012; published online May 18, 2012.

1 Introduction

Most outdoor vision applications, such as surveillance camera systems, require robust detection of image features. In foggy weather conditions, the contrast and color of images are drastically altered or degraded. Hence, it is imperative to remove weather effects from video in order to make the vision system more reliable. So far, there have been many attempts to remove fog from a single image,[1–3] but only a few studies have focused on video sequences. These fog removal methods for video sequences mainly use a frame-by-frame strategy. For this kind of methods, the basic idea of most approaches lies in depthlike map computation by using multiple images under different weather conditions.[4,5] Although the results are satisfactory, the strict requirement that two reference images captured in different weather conditions are needed in the input limits application of these methods. Other approaches separate the foreground and background images and then combine two results to get the haze-free video sequences. For example, John and Wilscy[6] apply the wavelet fusion method on background and foreground pixels separately, which reduces the computation time. Xu et al.[7] removed fog from the foreground and background images by using a contrast-limited adaptive histogram equalization (CLAHE)-based method. However, these existing algorithms still take a lot of computation time.

In order to solve long computation problems, we consider adopting a "universal strategy" to the video defogging problem, which means applying one transmission map to restore as many foggy frames as possible. Our method is based on the observation that in surveillance situations, cameras usually are fixed high in the air, and the difference in depth between a foreground object and the background behind it is usually very small, which leads to very little change in the transmission map. Therefore, it is possible to estimate a universal transmission map through the background image only, and apply this universal map to a series of video frames without creating significant errors in the recovered video. Compared with the existing frame-by-frame methods, our method calculates the transmission map only once, so the fog removal speed is much faster.

The remaining of this paper is organized as follows: Sec. 2 describes our approach in detail. Section 2.1 describes the haze image model and the atmospheric light estimation, Sec. 2.2 discusses the transmission map extraction from surveillance video, and Sec. 2.3 describes the scene radiance restoration based on the haze image model. To show the effectiveness of the proposed method, the experimental results and performance evaluation is given in Sec. 3, and conclusions are drawn in Sec. 4.

2 Proposed Approach

Given a video that has haze, mist, or fog, we first extract the background image and then estimate the atmospheric light value using the dark channel prior proposed by He et al.[1] Then, the transmission map is calculated on the intensity component of the background image, through a process of multiscale retinex (MSR), parameter adjustment, bilateral filtering, and total variation (TV) denoising filtering. Finally, output haze-free video can be gained according to the haze image model. This process is also depicted in Fig. 1.

2.1 *Haze Image Model and Atmospheric Light Estimation*

The haze image model (also called the image degradation model), proposed by McCartney[8] in 1975, consist of a direct attenuation model and an air light model. The direct attenuation model describes the scene radiance and its decay in the medium, while air light results from previously scattered

0091-3286/2012/$25.00 © 2012 SPIE

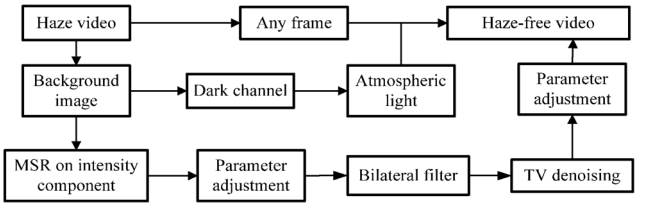


Fig. 1 Haze removal process for video.

light. The formation of a haze image model can be expressed as follows:

$$I(x) = J(x)t(x) + A[1 - t(x)], \tag{1}$$

where I is the observed intensity and the input haze image, J is the scene radiance and the restored haze-free image, A is the global atmospheric light, and t is the transmission map describing the portion of the light that is not scattered by the medium and reaches the camera. Theoretically, the goal of haze removal is to recover J, A, and t from I.

For a surveillance camera system, the primary interest is to enhance the visibility of the entire video (not any single image). For this purpose, the haze removal approach for video should adopt a different strategy than that for a single image. Here, we adopted a universal strategy to deal with the video defog problem; that is, to restore video according to the haze image model, only with the same atmospheric light value and a single transmission map. For universal usage consideration, the prior dark channel is used to estimate global atmospheric light in our previous dehazing algorithm.[9] We first picked the top 0.1% brightest pixels in the dark channel, and then selected the pixel with highest intensity in the first frame as the atmospheric light.

According to the haze image model, only if the value of atmospheric light A and the transmission map t are both known can we gain the final recovered result. Thus, in the following part, we present our pseudo-transmission map estimation approach in detail.

2.2 *Pseudo-Transmission Map Estimation*

The transmission map in the haze image model is obtained through the following three stages.

2.2.1 *Background extraction*

Our research considered an outdoor surveillance video camera capturing a scene (with moving objects) over an extended period of time. We wanted to process it in real time to obtain a haze-free video. The transmission map was only calculated for the background part of the video in order to apply the universal transmission map to more frames with tolerable errors. For simplicity, we defined the static part of the scene as the background part and the moving objects in the scene as the foreground part. There exist many methods for estimating the background image, like background modeling methods and simple averaging mechanisms. In this paper, the background image was obtained using a frame differential method.[10] For example, Fig. 2(a) and 2(b) shows the first and second frames of the input video sequence, the moving object is bounded by a box as shown in Fig. 2(c), and the estimated background is shown in Fig. 2(d).

2.2.2 *Initial transmission map extraction*

The transmission map is calculated based on the assumption that the intensity of an image reflects the amounts of photons received by every position of the image. The farther the distance between the scene points and camera, the fewer photons are received by the sensor, which leads to darker intensity in the image. Thus, the depth information reflected by the transmission map can be measured by intensity; in other words, the transmission map can be estimated by the intensity component of an image. So our objective is to obtain the transmission information from the intensity component of an image. For this purpose, multiscale retinex, and a bilateral filter are used on the intensity component of the background image.

Specifically, we first chose the transformed color space. For a RGB image, in order to separate its intensity component from its chroma component, the color space should be transformed to HIS, YUV, etc. Recently, the most widely used color space is HIS. However, the transformation between HIS and RGB takes trigonometric calculation, which costs a great deal of time. We chose the YCbCr space in our algorithm for the reason that the transformation between YCbCr and RGB takes only a simple algebraic operation, so it takes little calculation and its speed is relatively fast. Then, we use the MSR algorithm on the intensity component, the process can be expressed as follows:

$$R_m(x,y) = \sum_{n=1}^{N} w_n(\log Y(x,y) - \log[F_n(x,y) * Y(x,y)]), \tag{2}$$

where $R_m(x,y)$ is the output of the transformation to the intensity component image by using the MSR algorithm, N is the number of the scale (usually 3), w_n is the weight corresponding to each scale, $Y(x,y)$ is the distribution of the intensity image, $F_n(x,y)$ is the nth surround function (whose form is Gaussian), and this function should cover all kinds of scale. Experiments show that the scales should be chosen with small, medium, and large values for most

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

Fig. 2 Background extraction example. (a) Frame #1 of input foggy video; (b) Frame #2 of input foggy video; (c) detected moving object; (d) estimated background.

images. We choose 15, 80, and 250 for small, medium, and large scales, respectively, and simply set all w_n values to be 1/3. Next, we use the parameter C minus the value of each pixel to adjust the transmission map. The value of C is application-based. We find that 1.09 is suitable for most situations. However, MSR often causes the transmission map to contain redundant details that may be wrong. In order to refine the map, we use a bilateral filter[11] to make the transmission map blur while preserving the edges. For example, if the red bricks and the slots between them in Fig. 3(a) share the same depth values, there should be little difference between them in the transmission map. However, as shown in Fig. 3(b), slots between the bricks can be easily found in the transmission map, that might introduce unwanted errors in the restored image. The redundant details have been successfully removed by using a bilateral filter, as shown in Fig. 3(d), which makes the final results in Fig. 3(e) better than those shown in Fig. 3(c).

Notice that in our previous work,[9] a median filter is used to blur the transmission map. Although a median filter successfully blur the edges of the transmission map, it might cause halo effects near the depth discontinuities. In order to refine the detailed transmission map, we propose the bilateral filter that blurs the transmission map while preserving the edges. This filter can remove the influence of halo effects when estimating the transmission map.

2.2.3 *Universal pseudo-transmission map estimation*

For a surveillance camera system, the camera is usually fixed and often positioned high in the air, 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. Although the haze removal result obtained by using a bilateral filter to refine a transmission map is good for a single image, using the filter to calculate a transmission map for each frame is surely time-consuming. In order to calculate the pseudo-transmission map once so it can be used for all instances, this map should only contain an outlier of the scene. Therefore, the moving foreground object and image details can be regarded as noise. As the total variation based on Rudin-Osher-Fatemi (ROF) model has been proved to be quite efficient for denoising and regularizing images without smoothing the boundaries of the objects,[12] we use the total variation algorithm to calculate the pseudo-transmission map obtained by a bilateral filter. The ROF model regards the noiseless image u as the solution to a variational problem, to minimize the function

$$F(u) := \int_\Omega |\nabla u| + \frac{\lambda}{2}\int_\Omega |u-f|^2, \tag{3}$$

where Ω is the image domain and λ is a parameter need to be chosen. The first term is a regularization term, and the second is a data-fidelity term. Minimizing $F(u)$ has the effect of diminishing variation in u, while keeping u close to the noisy image f. We can estimate a universal pseudo-transmission map through the total variation algorithm, and apply this universal map to a series of video frames without creating significant errors in the restored frames. Figure 4(b) shows the initial transmission map calculated from the background, and Fig. 4(e) shows our final refined transmission map. We can see that the map does not induce significant artifacts into enhanced frame. Furthermore, the main advantage of this choice is its speed since the atmosphere light and the transmission map is estimated only once and is applied to as many frames as possible.

Figure 4 shows our video defogging results with universal pseudo-transmission map, with Fig. 4(a) and 4(d) being arbitrarily selected input foggy frames. From Fig. 4, we can see that though calculated through one previous initial transmission map [Fig. 4(b)], the universal pseudo- transmission map shown in Fig. 4(e) does not introduce significant artifacts into the restored frame [Fig. 4(f)]. Furthermore, the main advantage of this approach is its speed since the atmosphere light and the transmission map are estimated only using the background image, and are applied to all frames.

2.3 *Scene Radiance Restoration*

According to the haze image model, we can recover the scene radiance of each frame with the same transmission map obtained by the algorithm presented above. Because the direct attenuation term $J(x)t(x)$ can be very close to zero, so the transmission $t(x)$ is restricted to a lower bound t_0, which means that a small amount of haze is preserved in very dense haze regions. A typical value of t_0 is 0.1. Then the final scene radiance $J(x)$ can be recovered by

$$J(x) = \frac{I(x)-A}{\max[t(x), t_0]} + A. \tag{4}$$

From the first frame, the transmission map $t(x)$ and the atmospheric light A can be estimated, so we can apply Eq. (4) to each frame of the original video to obtain the final haze-free result.

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

Fig. 3 Refined transmission map with bilateral filter. (a) Foggy image; (b) transmission map obtained without filter; (c) restored haze-free result without filter; (d) transmission map obtained with filter; (e) restored haze-free result with filter.

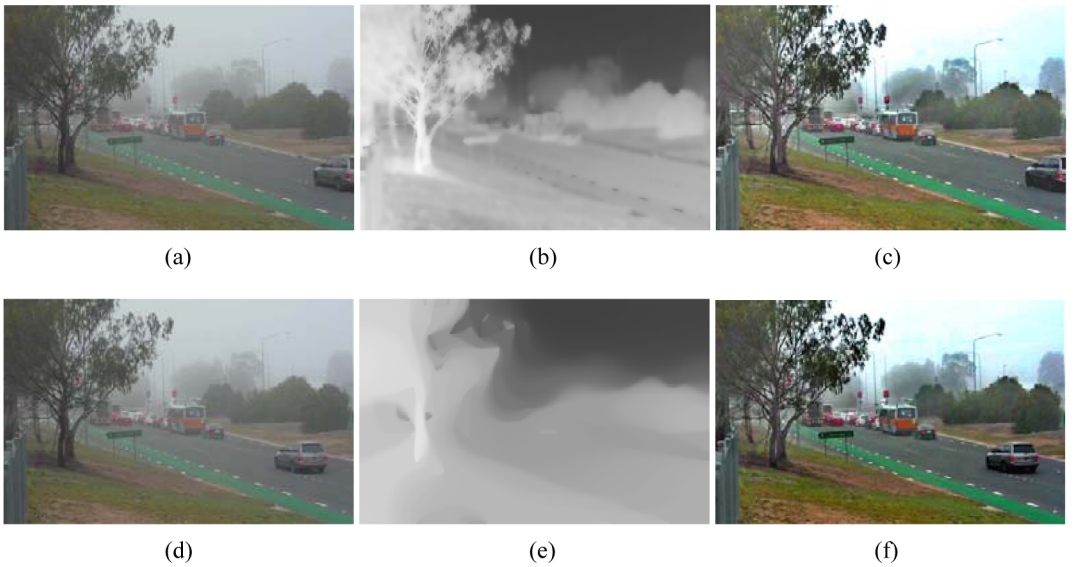


Fig. 4 Our defogging result for video clip #1. (a) and (d) are arbitrarily selected input foggy frames; (b) is the initial transmission map calculated from the background; (e) is the refined transmission map estimated by using total variation denoising algorithm to (b); (c) and (f) are defogging frames restored by transmission map (b) and (e).

3 Experimental Results and Performance Evaluation

3.1 *Video Defogging Results*

Here, we present a set of results using videos that were captured in foggy weather conditions. We first extracted the background image using the frame differential method, and then the universal transmission map can be obtained through total variation denoising. Figure 4 shows the procedure of our approach. As can be seen in the figure, the initial transmission map obtained by using bilateral filtering on the intensity component of the background image contains too much "noise" for video processing, such as the moving foreground object and image details [Fig. 4(b)]. Thus, we used a total variation denoising filter to get our universal video transmission map [Fig. 4(e)]. Finally, we applied the obtained atmospheric light and our final transmission map to each frame of the original video.

Some other video defogging results can be seen in Fig. 5, with which all prove the effectiveness of our universal strategy. Moreover, this strategy reduces both the calculation time and the memory requirement; thus, it is feasible for real-time applications.

3.2 *Performance Evaluation*

So far, there is a lack of widely accepted methodology to assess the performances of the defog algorithm, or to compare them with one another. Unlike image quality assessment or image restoration issues, it is not easy (sometimes even impossible) to have a fully referenced image. In order to compare our method with other state-of-the-art video defogging techniques, three criteria were considered: (i) qualitative comparison, (ii) quantitative evaluation, and (iii) computation time.

3.2.1 *Qualitative comparison*

Figure 6 shows a comparison between results obtained by John and Wilscy[6] and our algorithms, where particular parts (in the red rectangular) of three consecutive frames are enlarged on the right. It can be seen that the main problems for John and Wilscy's approach are (1) an unavoidable block effect due to the foreground subimages and (2) a moving object detection error due to optimal threshold determination. Notice that our algorithm does not produce such problems, while maintaining an even better defogging result compared with John and Wilscy's method.

3.2.2 *Quantitative evaluation*

For quantitative evaluation, the definition of meteorological visibility distance proposed by CIE[13] was considered here. In order to be consistent with the definition, the set of edges that have a local contrast above 5% was computed to obtain the visible edges under foggy weather. To implement the definition, Kohler's segmentation method that based on the Logarithmic Image Processing (LIP) model[14] was used here. The visible edges in the video frames before and after haze removal were selected by a 5% contrast thresholding. The computation results of visible edges for the method in John and Wilscy's[6] and our method are given in Fig. 7. From Fig. 7(e) and 7(f), we can see that both methods acquired more visible edges than before. Since the performance evaluation was based on the meteorological visibility distance, the result coincided with the visual evaluation of human eyes.

We adopted the blind contrast enhancement assessment method[15] to evaluate the performance of a video defogging algorithm. Here, we transformed the color level image to the gray level image and used three indicators e, $\bar{r}$, and σ to compare two gray level images: the input image and the restored image. The value of e evaluates the ability of the method to restore edges that were not visible in the original image but were visible in the defogging image. Then, the ratio $\bar{r}$ of the average gradients after and before restoration was computed. This indicator $\bar{r}$ estimates the average visibility enhancement obtained by the restoration algorithm takes into account both invisible and visible edges. Finally, we computed the number

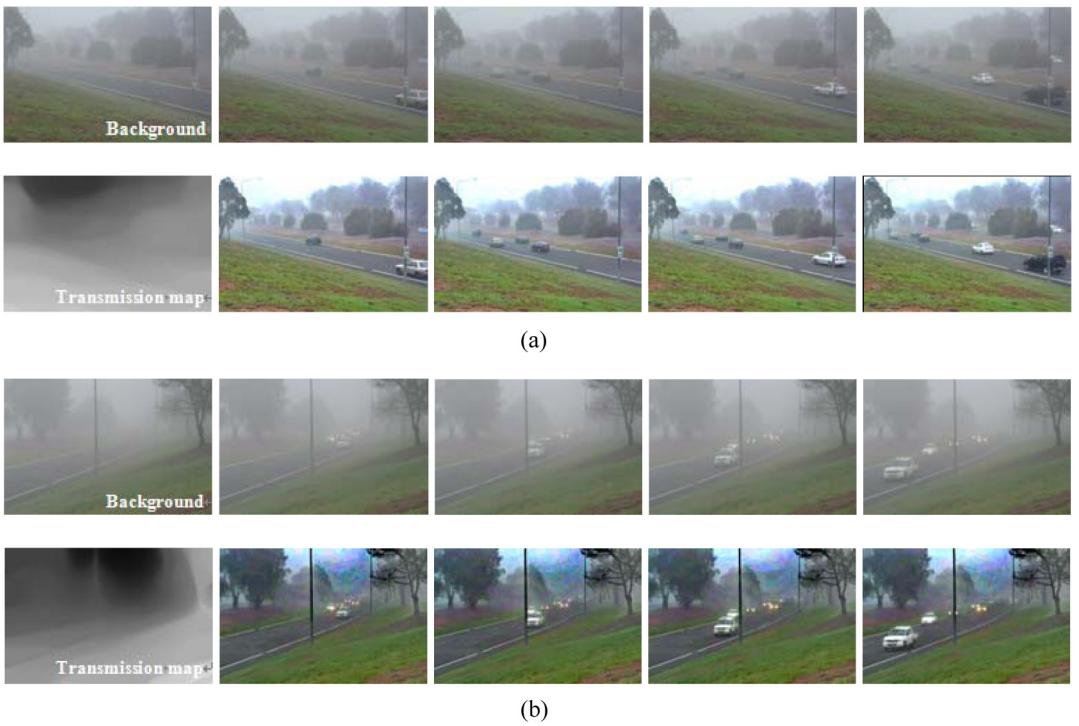


Fig. 5 Some other video defogging results. (a) Video clip #2 (a total of 80 frames, of which frames 20, 40, 60 and 80 are shown); (b) Video clip #3 (a total of 100 frames, of which frames 20, 50, 70 and 90 are shown).

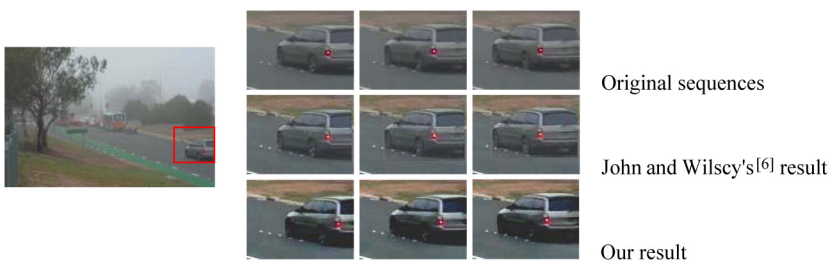


Fig. 6 Local comparison of three consecutive frames.

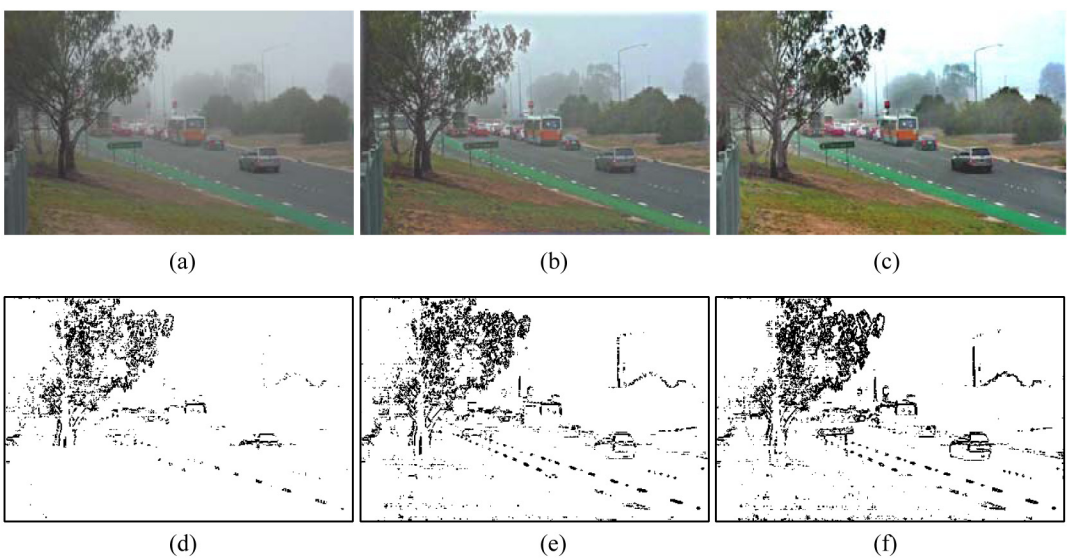


Fig. 7 Comparison of visible edges. (a) The arbitrarily selected frame from input foggy video; (b) haze removal result by John and Wilscy's method; (c) haze removal result by our method; (d)–(f) are corresponding visible edges of (a)–(c).

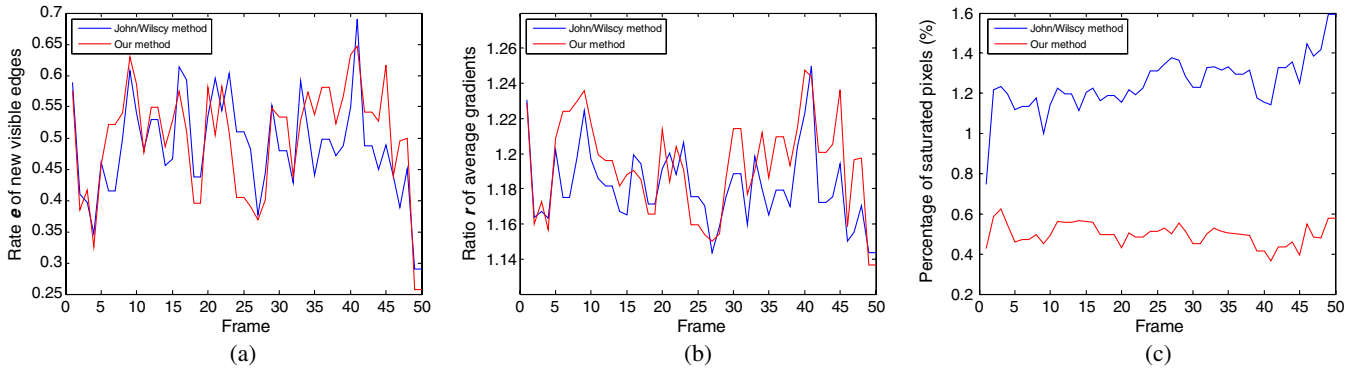


Fig. 8 Comparison results for John and Wilscy's and our methods on 50 frames of video #1; (a) rate e of new visible edges; (b) ratio $\bar{r}$ of average gradients at the visible edges; (c) percentage of pixels that become saturated after restoration.

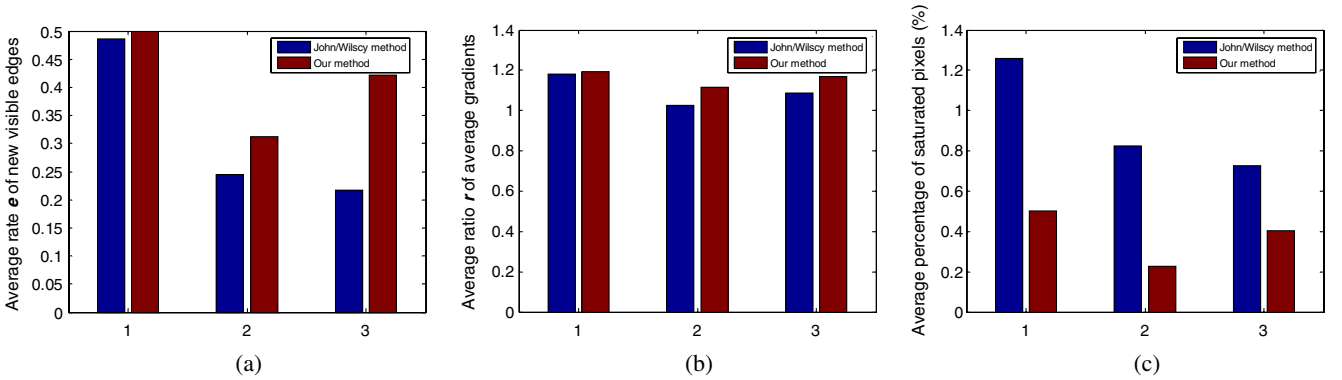


Fig. 9 Comparison results for John and Wilscy's and our methods on three video clips; (a) average rate e of new visible edges; (b) average ratio $\bar{r}$ of average gradients at the visible edges; (c) average percentage of pixels that become saturated after restoration.

of pixels that were saturated (black or white) after applying the defogging algorithm, but not before. We normalized this value by the size of the image to get the σ index, indicating the percentage of pixels that becomes completely black or white after restoration.

These indicators e, $\bar{r}$, and σ were evaluated for John and Wilscy's[6] and our approach on video #1 with a total of 50 frames (see Fig. 8). In order to compare the two approaches better, we also computed the average value of these indictors on three video clips (see Fig. 9). In each approach, the aim of defogging was to increase the contrast without losing detailed visual information. Hence, good results are described by high values of e and $\bar{r}$, and low values of σ. From Figs. 8 and 9, we can see that our method can generate comparable results with John and Wilscy's method. Moreover, depending on the video frames, our method generally has little more visible edges, the two methods are comparable in average gradients, and our method has a much smaller percentage of pixels that become saturated after restoration. Therefore in terms of blind contrast enhancement assessment, our method outperforms John and Wilscy's method. This confirms our visual observations on all video clips.

Table 1 Comparison of computation time (in seconds).

Sequence	Size (pixel × pixel × frames)	John and Wilscy's method[6]	Our method
Video #1	480 × 270 × 50	96.7382	13.5630
Video #2	480 × 270 × 80	180.4530	21.8590
Video #3	480 × 270 × 100	237.6250	27.3750

3.2.3 *Computation times*

The time required for video defogging depends on the size of the frames in foggy videos and the complexity of the scenes. The foreground detection and the wavelet fusion of John and Wilscy's method takes a lot of time, while our method is a transformation of the haze image model, and its complexity is a linear function of the number of frame pixels only, making the computation much faster than John and Wilscy's method, as shown in Table 1. All the experiments are implemented on a PC with 3-GHz Intel Pentium dual-core processor.

4 Conclusions

This paper proposes a fast video defogging approach using the haze image model with the same atmospheric light and a universal transmission map. The proposed method is applicable in a situation where the camera is fixed and positioned high above the ground. Our approach produces visually pleasing defogging results and tends to preserve the main details better than previous techniques. Compared with state-of-the-art algorithms, our approach has two main

advantages: (1) no other reference image is needed and (2) its speed is relatively faster while maintaining comparable performance. A comparative study and quantitative evaluation demonstrate the efficiency of our method. Our approach could be further improved by improving the atmospheric light estimation method and by employing some advanced haze image models.

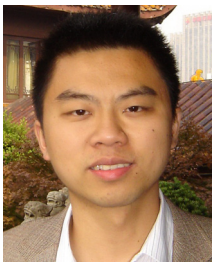
We have shown that we can generate real-time and good haze removal results by applying the same atmospheric light and transmission map of the haze image model to a series of video frames. However, our results show a tendency to degrade the quality of the transmission map, resulting in blurring effects in some scene edges. This indicates that although our method generates a universal transmission map, it is not an accurate transmission map. This disqualifies our approach from dealing effectively with frames of an object without a clear shape, such as trees far from the camera, and the haze removal result is sometimes high-colored. Nevertheless, we can improve the overall quality of a foggy video by enhancing the main details.

Acknowledgments

This work was partially supported by the State Key Program of the National Natural Science Foundation of China (No. 90820302), the Science and Technology Program of Changsha (No. K1101001-11), the Freedom Explore Program of Central South University (No. 2011QNZT036), and the Postdoctoral Science Foundation of Central South University.

References

1. K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," in *Proc. IEEE Comp. Soc. Conf. Comp. Vision Pattern Rec. (CVPR)*, pp. 1956–1963, IEEE Computer Society, Piscataway, New Jersey (2009).
2. R. Tan, "Visibility in bad weather from a single image," in *Proc. IEEE Comp. Soc. Conf. Comp. Vision Pattern Rec. (CVPR)*, pp. 1–8, IEEE Computer Society, Piscataway, New Jersey (2008).
3. R. Fattal, "Single image dehazing," *ACM Trans. Graphics* **27**(3), 72–79 (2008).
4. S. G. Narasimhan and S. K. Nayar, "Contrast restoration of weather degraded images," *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(6), 713–724 (2003).
5. G. Chen, H. Zhou, and J. Yan, "A novel method for moving object detection in foggy day," *Proc. 8th ACIS Intl. Conf. Software Eng. Artificial Intell., Networking, and Parallel/Distributed Comp.*, pp. 53–58, IEEE Computer Society, Piscataway, New Jersey (2007).
6. J. John and M. Wilscy, "Enhancement of weather degraded video sequences using wavelet fusion," *Proc. Cyber. Intell. Syst., 2008. CIS 2008. 7th IEEE Intl. Conf.* 1–6, IEEE Computer Society, Piscataway, New Jersey (2008).
7. Z. Xu, X. Liu, and X. Chen, "Fog removal from video sequences using contrast limited adaptive histogram equalization," *Proc. Comp. Intell. Software Eng. Intl. Conf.* pp. 1–4, IEEE Computer Society, Piscataway, New Jersey (2009).
8. E. J. McCartney, *Optics of Atmosphere: Scattering by Molecules and Particles*, John Wiley and Sons, New York, pp. 23–32 (1976).
9. B. Xie, F. Guo, and Z. Cai, "Improved single image dehazing using dark channel prior and multi-scale Retinex," *Proc. Intl. Conf. Intell. Sys. Design Eng. App.*, pp. 848–851, IEEE Computer Society, Piscataway, New Jersey (2010).
10. I. Haritaoglu, D. Harwood, and L. S. Davis, "W4: Real-time surveillance of people and their activities," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(8), 809–830 (2000).
11. C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," *Proc. IEEE Intl. Conf. Comp. Vision*, pp. 839–846, IEEE Computer Society, Piscataway, New Jersey (1998).
12. A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vision* **20**(1), 89–97 (2004).
13. International Commission on Illumination, International Lighting Vocabulary, Number 17.4, CIE, 4th ed., 978-3-900734-07-7 (1987).
14. R. Kohler, "A segmentation system based on thresholding," *Comp. Graph. Image Proc.* **15**(4), 319–338 (1981).
15. N. Hautiere et al., "Blind contrast enhancement assessment by gradient ratioing at visible edges," *Image Anal. Stereol. J.* **27**(2), 87–95 (2008).



Bin Xie received a BS degree in electronic information engineering and a PhD in information and communication engineering from Zhejiang University, Hangzhou, China, in 2003, and 2008, respectively. Currently, he is a lecturer with the School of Information Science and Engineering, Central South University. His main research interests include image processing, pattern recognition, and robotics.

Fan Guo received a BSc in computer science in 2005 and an MSc in computer science in 2008 from Central South University, China. She is a PhD candidate at the School of Information Science and Engineering, Central South University, China. Her research interests include image processing and virtual reality.

Zixing Cai received the diploma degree from the Department of Electrical Engineering, Jiao Tong University, Xi'an, China, in 1962. He has been teaching and doing research at the School of Information Science and Engineering, Central South University, Changsha, China, since 1962. He has authored/coauthored over 900 papers and 30 books/textbooks. His current research interests include intelligent systems, artificial intelligence, intelligent computation, and robotics. Professor Cai has received over 30 state, province, and university awards in science, technology, and teaching. One of his most recent honors is the State Eminent Professor Prize of China.