Real-Time Video Dehazing Based on Spatio-Temporal MRF

Bolun Cai[1], Xiangmin Xu[1(✉)], and Dacheng Tao[2]

[1] South China University of Technology, Guangzhou, China
caibolun@gmail.com, xmxu@scut.edu.cn
[2] University of Technology Sydney, Ultimo, Australia
dacheng.tao@uts.edu.au

Abstract. Video dehazing has a wide range of real-time applications, but the challenges mainly come from spatio-temporal coherence and computational efficiency. In this paper, a spatio-temporal optimization framework for real-time video dehazing is proposed, which reduces blocking and flickering artifacts and achieves high-quality enhanced results. We build a Markov Random Field (MRF) with an Intensity Value Prior (IVP) to handle spatial consistency and temporal coherence. By maximizing the MRF likelihood function, the proposed framework estimates the haze concentration and preserves the information optimally. Moreover, to facilitate real-time applications, integral image technique is approximated to reduce the main computational burden. Experimental results demonstrate that the proposed framework is effectively to remove haze and flickering artifacts, and sufficiently fast for real-time applications.

Keywords: Video dehazing · Spatio-temporal MRF · Intensity value prior

1 Introduction

Haze is a traditional phenomenon where dust, smoke and other dry particles obscure the clear atmosphere, which makes the image/video contrast lost and/or vividness lost. The light scattering through the haze particles results in a loss of contrast in the photography process. Video dehazing has broader and broader prospects for real-time processing (e.g. automatic driving, video surveillance, automobile recorder). Since the haze concentration is spatio-temporal relevant, recovering the haze-free scene from hazy videos becomes a challenging problem.

X. Xu—This work is supported in part by the National Natural Science Founding of China (61171142, 61401163), Science and Technology Planning Project of Guangdong Province of China (2011A010801005, 2014B010111003, 2014B010111006), Guangzhou Key Lab of Body Data Science (201605030011) and Australian Research Council Projects (FT-130101457 and DP-140102164).

E. Chen et al. (Eds.): PCM 2016, Part II, LNCS 9917, pp. 315–325, 2016.
DOI: 10.1007/978-3-319-48896-7_31

Various image enhancement techniques [7,12] are proposed to deal with static image dehazing, which transform the color distribution without considering the haze concentration. Moreover, methods based on multi-image [11] or depth-information [8] are employed, but the additional information is hard to be acquired in real application scenes. Due to the use of strong priors, single image dehazing methods have made significant progresses recently. Dark channel prior [5] shows at least one color channel has some pixels with very low intensities in most of non-haze patches; Meng et al. [10] propose an effective regularization method to recover the haze-free image by exploring the inherent boundary constraint. Since the above algorithms only focus on static image dehazing, they may yield flickering artifacts due to the lack of temporal coherence when applied to video dehazing.

Little work has been done on video dehazing compared to extensive works on static image dehazing. Tarel et al. [13] segment a car-vision video into motorial objects and a planar road, then update the depth for haze removal based on a image dehazing scheme [12]. Therefore, this method is unable to apply in unrestraint conditions. To improve the spatio-temporal coherence, an optical flow method [16] is introduced to optimize the haze concentration map, which requires high computational complexity and is hard for real-time applications. Kim et al. [6] optimize contrast enhancement by minimizing a temporal coherence cost to reduce flickering artifacts. If the contrast is overstretched, some saturation values are truncated resulting in computationally intensive. In [8], authors combine depth and haze information to recover the clear scene. However, depth reconstruction depending on Structure-from-Motion (SfM) requires high complexity, and cannot get satisfying performance in the distance.

Extending image dehazing algorithm to video is not a trivial work. The challenges mainly come from the following aspects:

- Spatial consistency. There are two constraints of spatial consistency. The haze concentration is locally constant to overcome the estimation noise. In addition, the recovered video should be as natural as the original one to handle inner-frame discontinuity.
- Temporal coherence. Human visualization system is sensitive to temporal inconsistency. However, applying static image dehazing algorithm naively on frame-by-frame may break the temporal coherence, and yield a recovered video with severe flicking artifacts.
- Computational efficiency. The algorithm must be able to efficiently process the large number of pixels in video sequences. In particular, a practical real-time dehazing method should reach a speed of at least 15 frames per second.

In this paper, we build a spatio-temporal MRF with IVP to optimize haze concentration estimation. This method effectively assures the spatial consistency and temporal coherence of video dehazing. In addition, integral image technique [14] is used for efficiently computing in $O(N)$ time to reduce the main computational burden. Typically, the only single CPU implementation achieves approximately 120 frames per second for real-time video with the size of 352×288.

2 Real-Time Video Dehazing

Currently, all of the static images dehazing algorithms can obtain truly good results on general outdoor images. However, when applied to each frame of a hazy video sequence independently, it may break spatio-temporal coherence and produce a recovered video with blocking and flickering artifacts. Moreover, its high computational complexity prohibits real-time applications. In this section, we propose a spatio-temporal optimization framework for real-time video dehazing, which is shown in Fig. 1.

2.1 Single Image Haze Removal

Single image haze removal is a classical image enhancement problem. According to empirical observations, existing methods propose various assumptions or prior (e.g. dark channel [5], maximum contrast [6] and hue disparity [1]) to estimate the haze concentration. Based on the atmospheric scattering model and the haze concentration, the haze-free image is recovered easily.

Atmospheric Scattering Model. To describe the formation of a hazy image, the atmospheric scattering model is proposed by McCartney [9]. The atmospheric scattering model can be formally written as

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

where $I(x)$ is the observed hazy image, $J(x)$ is the real scene to be recovered, $T(x)$ is the medium transmission, A is the global atmospheric light, and x indexes pixels in the image. The real scene $J(x)$ can be recovered after A and $T(x)$ are estimated. The atmosphere light A is constant in the whole image, so it is easy to estimate. The medium transmission map $T(x)$ describes the light portion that is not scattered and reaches the camera. Therefore, it is the key to estimate an accurate haze concentration map.

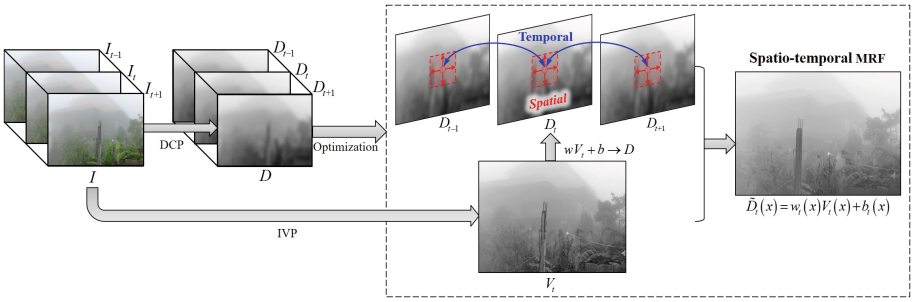
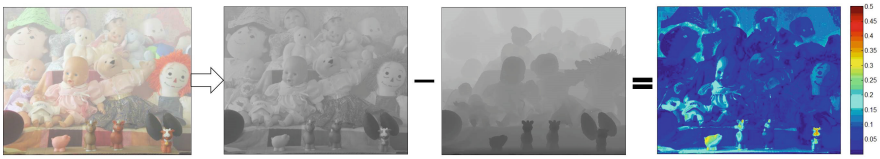


Fig. 1. Spatio-temporal MRF for video dehazing. DCP is used to estimate the haze concentration and an MRF is built based on IVP between inner-frame and inter-frame.



(a)Synthetic image (b)Intensity value (c)Concentration (d)Residual error

Fig. 2. Intensity Value Prior. The residual error is close to zero, and shows that the haze concentration is highly correlated with the intensity value.

Medium Transmission Estimation. Dark Channel Prior [5] (DCP) is discovered based on empirical statistics of experiments on outdoor haze-free images. In most of haze-free images, at least one color channel has some pixels whose intensity values are very low and even close to zero. The dark channel is defined as the minimum channel in RGB color space:

$$D(x) = \min_{c \in \{r,g,b\}} I^c(x), \tag{2}$$

where $I^c(x)$ is a RGB color channel of $I(x)$. The dark channel prior has a high correlation to the amount of haze in the image, and is used to estimate the medium transmission directly as $\tilde{T}(x) = 1 - \omega D(x)/A$, where a constant parameter ω is introduced to map dark channel value to the medium transmission. We fix it to 0.7 for all results reported in this paper.

2.2 Spatio-Temporal MRF

To handle blocking and flickering artifacts, the haze concentration map should be refined by spatio-temporal coherence. Based on an intensity value prior, a spatio-temporal MRF is built to fine-tune the haze concentration map, as which the dark channel map $D(x)$ is regarded in this paper.

Intensity Value Prior. With the wide observation on hazy images, the intensity values of pixels in a hazy image vary sharply along with the change of the haze concentration. To show how the intensity value of pixels vary within a hazy image, Fig. 2 gives an example with an image synthetized by known haze concentration. It can be deduced from $A(1 - T(x))$ in the atmosphere scattering model, that the effect of the white or gray airlight on the observed values is related to the amount of haze. Thus, caused by the airlight, the intensity value is increased while haze concentration is enhanced.

Spatial Consistency. The pixel-level concentration estimation may fail to work in some particular situations. For instance, outlier pixel values in an image result in inaccurate estimation of the haze concentration. Based on the assumption that the haze concentration is locally constant, local filters (e.g. minimum filter [17], maximum filter [2] and medium filter [4]) are commonly to overcome this problem. However, blocking artifacts appear in the haze concentration map because

of these local filters. To handle the locally constant and inner-frame continuity, a spatial MRF is built based on IVP. In spatial neighborhood, the intensity value $V(x)$ is linear transformed to the haze concentration $D(x)$, and the transformation fields $W = \{w(x)\}_{x\in\forall}$ and $B = \{b(x)\}_{x\in\forall}$ are only correlated with the contextual information. The spatial likelihood function is

$$P_s(w,b) \propto \prod_{y\in\Omega(x)} \exp\left(-\frac{\|w(x)V(y)+b(x)-D(y)\|_2^2}{\sigma_s^2}\right), \tag{3}$$

where $\Omega(x)$ is a local patch centered at x with the size of $r \times r$, and σ_s is the spatial parameter.

Temporal Coherence. Flicking artifacts can be avoided by the relevant information between consecutive frames. The haze concentration changes due to camera and object motions. As an object approaches in the camera, the observed radiance gets closer to the original scene radiance. On the contrary, when an object moves away from the camera, the observed radiance becomes more similar to the atmospheric light. Thus, we can modify the haze concentration of a scene point adaptively according to its intensity value change. As shown in Fig. 3, the haze concentration of the neighbor frame can be transformed to the current frame's by IVP, which is similar to block-matching of optical flow estimation. As with the spatial consistency, a temporal MRF is used for temporal coherence, and at time t its likelihood function is defined by

$$P_\tau(w_t,b_t) \propto \prod_{\tau\in[-f,+f]} \exp\left(-\frac{\|w_t(x)V_t(x)+b_t(x)-D_{t+\tau}(x)\|_2^2}{\sigma_\tau^2}\right), \tag{4}$$

where f is the number of neighbor frames, and σ_τ is the temporal parameter.

Along the spatio-temporal dimension, we improve the spatial consistency and temporal coherence with an uniform likelihood function, which is rewritten as

$$P(w_t,b_t) = \prod_{\tau\in[-f,+f]} \prod_{y\in\Omega(x)} \exp\left(-\frac{\|w_t(x)V_t(y)+b_t(x)-D_{t+\tau}(y)\|_2^2}{\sigma_\tau^2}\right), \tag{5}$$

where σ_s is omitted because it is assumed as a constant in this paper.

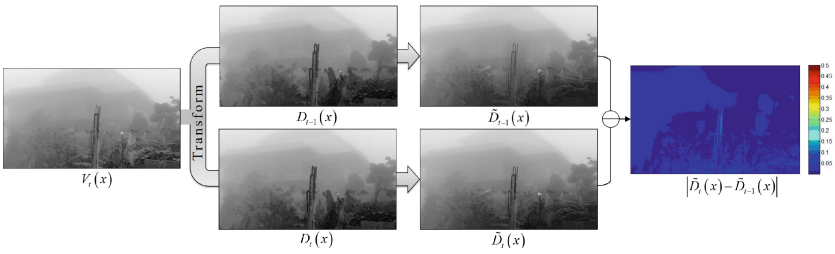


Fig. 3. Inter-frame correlation of the haze concentration. The intensity map $V_t(x)$ in current frame is transformed to the haze concentration map $D_{t,t-1}(x)$ in neighbor frames. The absolute error map between $\tilde{D}_t(x)$ and $\tilde{D}_{t-1}(x)$ is close to zero.

2.3 Maximum Likelihood Estimation

The log-likelihood function is more convenient to work with maximum likelihood estimation. Taking the derivative of a log-likelihood function and solving for the parameter, is often easier than the original likelihood function. Let temporal weights $\lambda_\tau = 1/\sigma_\tau^2$ $(s.t. \sum_\tau \lambda_\tau = 1)$ to express conveniently, and the log-likelihood function of Eq. 5 is given by:

$$L(w_t, b_t) = \sum_{\tau \in [-f,+f]} \sum_{y \in \Omega(x)} -\lambda_\tau \left\| w_t(x) V_t(y) + b_t(x) - D_{t+\tau}(y) \right\|_2^2 \quad (6)$$

To find the optimal random fields W and B, the maximum log-likelihood estimation is written as $(w_t, b_t) = \arg\max L(w_t, b_t)$. We maximize the probability by solving the linear system from $\partial L(w_t, b_t)/\partial w_t = 0$ and $\partial L(w_t, b_t)/\partial b = 0$, and generate the final haze concentration map by $\tilde{D}_t(x) = w_t(x) V_t(x) + b_t(x)$.

$$\begin{cases} w_t(x) = \dfrac{\sum_\tau \lambda_\tau \left(\mathcal{U}_\Omega \left[V_t(x) D_{t+\tau}(x)\right] - \mathcal{U}_\Omega \left[V_t(x)\right] \mathcal{U}_\Omega \left[D_{t+\tau}(x)\right]\right)}{\mathcal{U}_\Omega \left[V_t^2(x)\right] - \mathcal{U}_\Omega^2 \left[V_t(x)\right]} \\ b_t(x) = \sum_\tau \lambda_\tau \mathcal{U}_\Omega \left[D_{t+\tau}(x)\right] - w_t(x) \mathcal{U}_\Omega \left[V_t(x)\right] \end{cases} \quad (7)$$

Here, $\mathcal{U}[\cdot]$ is a mean filter defined as $\mathcal{U}[F(x)] = (1/|\Omega|) \sum_{y \in \Omega(x)} F(y)$, and $|\Omega|$ is the cardinality of the local neighborhood.

2.4 Complexity Reduction

A main advantage of the spatio-temporal MRF built in this paper is that it naturally has an $O(N)$ time non-approximate acceleration. The main computational burden is the mean filter $\mathcal{U}[\cdot]$ with the local neighborhood. Fortunately, the mean filter can be efficiently computed in $O(N)$ time using the integral image technique [14], which allows for fast computation of box type convolution filters. The entry of an integral image represents the sum of all pixels in the input image within a rectangular region formed by the origin and current position. Once the integral image has been computed, it takes three additions to calculate the sum of the intensities over any rectangular area. Hence, the calculation time of the mean function is independent of its size, and the maximum likelihood estimation in Sect. 2.3 is naturally $O(N)$ time.

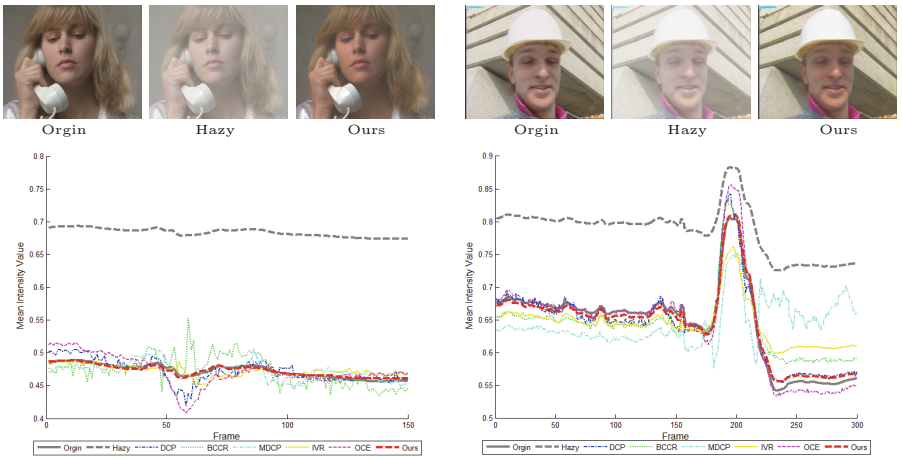
3 Experiments

We analyze the validity of the proposed framework and compare it with the state-of-art image/video dehazing methods, including DCP [5], BCCR [10], MDCP [4], IVR [13], OCE [6]. Based on the transmission estimated and the atmospheric scattering model, a haze-free video can be recovered by (1). The atmospheric light A is estimated as the brightest color [3] in an image: $A = \max_{x \in \forall} \left(\min_{y \in \Omega(x)} V(y)\right)$. At the t-th frame, the airlight is updated by $A_t = \rho A + (1 - \rho) A_{t-1}$, where $\rho = 0.1$ is a learning parameter. The other parameters mentioned in Sect. 2.2 are specified as follows: the number of neighbor frames f is set to 1, and the temporal weights λ_τ is set to a Hanning window.

3.1 Temporal Coherence Analysis

Temporal coherence is the main challenge compared to static image dehazing. However, the evaluation of temporal coherence is difficult on real videos since no reference is available. To show the proposed framework can suppresses flickering artifacts well, we compare the mean intensity value (MIV) between consecutive frames on five hazy videos, which are synthesized from non-haze videos[1] with flat haze $T(x) = 0.6$.

Figure 4 plots MIV between consecutive frames in *Suzie* and *Foreman*. When the static dehazing algorithms (including DCP [5], BCCR [10], MDCP [4]) are independently applied to each frame, the MIV curves experience relatively large fluctuations as compared with the original sequences, especially between 50–75 frames in Fig. 4(a) and 175–225 frames in Fig. 4(b). We also quantify the flickering artifacts based on the correlation analysis of MIV between the dehazing result and the original video, shown in Table 1. In contrast, our video dehazing method alleviates the fluctuations and reduces the flickering artifacts efficiently.



(a) Suzie (b) Foreman

Fig. 4. Comparsion of the mean intensity value in consecutive frames.

3.2 Quantitative Results on Synthetic Videos

To verify the dehazing effectiveness, the proposed framework is tested on hazy videos synthesized from stereo videos [15] with a known depth map[2], and it is compared with 5 representative methods. Among the competitors, MDCP [4], IVR [13] and OCE [6] are the most recent state-of-the-art video dehazing

[1] http://trace.eas.asu.edu/yuv/.
[2] http://www.cad.zju.edu.cn/home/gfzhang/projects/videodepth/data/.

Table 1. The correlation coefficients of MIV between dehazing and original videos

	DCP [5]	BCCR [10]	MDCP [4]	IVR [13]	OCE [6]	Ours
Suzie	0.783	0.612	0.641	0.584	0.649	**0.976**
Foreman	0.980	0.920	0.015	0.949	0.994	**0.995**
Container	0.929	0.703	0.927	0.998	**1.000**	0.955
Hall	0.784	0.429	0.444	0.824	0.845	**0.991**
Silent	0.853	0.898	0.936	0.892	0.770	**0.990**
Avg.	0.866	0.712	0.592	0.849	0.851	**0.982**

methods; DCP [5] and BCCR [10] are classical static image dehazing methods which are used as comparison baselines. The hazy video is generated based on (1), where we assume pure white atmospheric airlight $A = 1$.

To quantitatively assess these methods, we calculate Mean Square Error (MSE) between the original non-haze video and dehazing result. A low MSE represents that the dehazed result is satisfying while a high MSE means that the dehazing effect is not acceptable. In Table 2, our method is compared with 5 state-of-the-art methods on three synthetic video. Our method achieves the lowest MSEs outperforming the others.

Table 2. The dehazing results of MSE on the synthetic videos

	DCP [5]	BCCR [10]	MDCP [4]	IVR [13]	OCE [6]	Ours
Flower	0.0228	0.0240	0.0257	0.0479	0.0174	**0.0034**
Lawn	0.0198	0.0176	0.4902	**0.0141**	0.0408	0.0166
Road	0.0141	0.0191	0.0108	0.0364	0.0274	**0.0092**
Avg.	0.0189	0.0202	0.1756	0.0328	0.0285	**0.0097**

3.3 Qualitative Results on Real-World Videos

In addition, we also evaluate the performance of the proposed framework on the real-world videos collected in related works. Figure 5 shows the results on four representative sequences with different methods[3]. The contrast maximizing methods (BCCR [10], IVR [13], OCE [6]) are able to achieve impressive results, but they tend to produce over-saturated and spatial inconsistency (for example, the mountain in *Bali* and the halo of the sky in *Playground*). In Fig. 5(c) and (d), it is observed that the static image dehazing methods (DCP [5] and BCCR [10]) yield severe flickering artifacts (such as, the road region in *Cross* and the sky region in *Hazeroad*). Although, the OCE [6] method uses overlapped block

[3] More comparisons can be found at http://caibolun.github.io/st-mrf/.

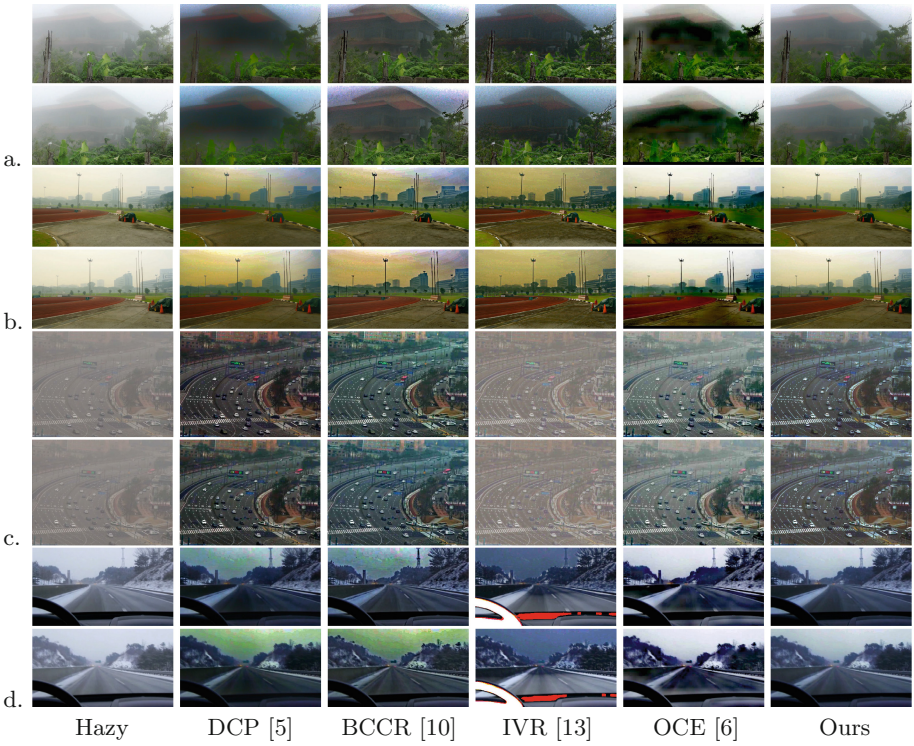


Fig. 5. Qualitative comparison of different methods on real-world hazy videos, including (a) Bali, (b) Playground, (c) Cross, (d) Hazeroad.

filter to reduce blocking artifacts, there are still a small number of blocking artifacts in the results. Compared with the other methods, our results avoid image over-saturation and keep spatio-temporal coherence.

3.4 Real-Time Analysis

We evaluate the computational complexity of the proposed framework on hazy videos with different sizes of general video standards. The experiments are run on a PC with Intel i7 3770 CPU (3.4 GHz), and we report the average speed (in fps) comparison with DCP [5], BCCR [10], MDCP [4], IVR [13], and OCE [6]. According to Table 3, our method is significantly faster than others and achieves efficient processing even when the given hazy video is large. Typically, our framework achieves the processing speed of about 120 fps on Common Intermediate Format (CIF, 352×288), which is close to quadraple that of the real-time criterion. Thus, the proposed framework leaves a substantial amount of time for other processing, and is transplanted into embedded system easily.

Table 3. Comparision of the processing speeds in terms of frames per second (fps)

	DCP [5]	BCCR [10]	MDCP [4]	IVR [13]	OCE [6]	Ours
CIF (352 × 288)	1.485	1.322	7.343	1.205	97.076	**116.371**
VGA (640 × 480)	0.566	0.467	2.430	0.171	30.539	**36.609**
D1 (704 × 576)	0.414	0.358	1.830	0.102	22.930	**27.493**
XGA (1024 × 768)	0.216	0.197	0.842	0.028	12.106	**14.515**

4 Conclusion

In this work, we propose a real-time video dehazing framework based on spatio-temporal MRF. We introduce the notion of spatial consistency and temporal coherence to yield a dehazed video without blocking and flickering artifacts. Moreover, the integral image technique is applied to reduce the computational complexity significantly. Experimental results demonstrate that the proposed algorithm can efficiently recover a hazy video at low computational complexity. However, DCP is unable to estimate the haze concentration in high accuracy. Moreover, this spatio-temporal framework can be extended for other real-time video processing. We leave these problems for future research.

References

1. Ancuti, C.O., Ancuti, C., Hermans, C., Bekaert, P.: A fast semi-inverse approach to detect and remove the haze from a single image. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010. LNCS, vol. 6493, pp. 501–514. Springer, Heidelberg (2011). doi:10.1007/978-3-642-19309-5_39
2. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: an end-to-end system for single image haze removal. arXiv preprint arXiv:1601.07661 (2016)
3. Chiang, J.Y., Chen, Y.C.: Underwater image enhancement by wavelength compensation and dehazing. IEEE Trans. Image Process. **21**(4), 1756–1769 (2012)
4. Gibson, K., Vo, D., Nguyen, T.: An investigation in dehazing compressed images and video. In: OCEANS 2010, pp. 1–8 (2010)
5. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Trans. Pattern Anal. Mach. Intell. **33**(12), 2341–2353 (2011)
6. Kim, J.H., Jang, W.D., Sim, J.Y., Kim, C.S.: Optimized contrast enhancement for real-time image and video dehazing. J. Vis. Commun. Image Represent. **24**(3), 410–425 (2013)
7. Kim, T.K., Paik, J.K., Kang, B.S.: Contrast enhancement system using spatially adaptive histogram equalization with temporal filtering. IEEE Trans. Consum. Electron. **44**(1), 82–87 (1998)
8. Li, Z., Tan, P., Tan, R.T., Zou, D., Zhou, S.Z., Cheong, L.F.: Simultaneous video defogging and stereo reconstruction. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4988–4997 (2015)
9. McCartney, E.J.: Optics of the Atmosphere: Scattering by Molecules and Particles. Wiley, New York (1976)

10. Meng, G., Wang, Y., Duan, J., Xiang, S., Pan, C.: Efficient image dehazing with boundary constraint and contextual regularization. In: IEEE International Conference on Computer Vision (ICCV), pp. 617–624 (2013)
11. Narasimhan, S.G., Nayar, S.K.: Contrast restoration of weather degraded images. IEEE Trans. Pattern Anal. Mach. Intell. **25**(6), 713–724 (2003)
12. Tarel, J.P., Hautiere, N.: Fast visibility restoration from a single color or gray level image. In: IEEE International Conference on Computer Vision, pp. 2201–2208 (2009)
13. Tarel, J.P., Hautiere, N., Cord, A., Gruyer, D., Halmaoui, H.: Improved visibility of road scene images under heterogeneous fog. In: 2010 IEEE conference on Intelligent Vehicles Symposium (IV), pp. 478–485. IEEE (2010)
14. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. I–511 (2001)
15. Zhang, G., Jia, J., Wong, T.T., Bao, H.: Consistent depth maps recovery from a video sequence. IEEE Trans. Pattern Anal. Mach. Intell. **31**(6), 974–988 (2009)
16. Zhang, J., Li, L., Zhang, Y., Yang, G., Cao, X., Sun, J.: Video dehazing with spatial and temporal coherence. Vis. Comput. **27**(6–8), 749–757 (2011)
17. Zhu, Q., Mai, J., Shao, L.: A fast single image haze removal algorithm using color attenuation prior. IEEE Trans. Image Process. **24**(11), 3522–3533 (2015)