

Video Smoke Removal Based on Smoke Imaging Model and Space-Time Pixel Compensation

Shiori Yamaguchi^(✉), Keita Hirai^(✉), and Takahiko Horiuchi

Graduate School of Advanced Integration Science, Chiba University,
Yayoi-cho 1-33, Inage-ku, Chiba 263-8522, Japan
acda2711@chiba-u.jp, hirai@faculty.chiba-u.jp

Abstract. This paper presents a novel video smoke removal method based on a smoke imaging model and space-time pixel compensation. First, we develop an optical imaging model for natural scenes that contain smoke. Then, we remove the smoke in a video, frame-by-frame, based on the smoke imaging model and conventional dehazing approaches. Next, we align the smoke-removed frames using corresponding pixels. To obtain the corresponding pixels, we use SIFT and color features with distance constraints. Finally, to reproduce clear video appearance, we compensate pixel values by utilizing the space-time weightings of the corresponding pixels between the smoke-removed frames. Validation experiments show our method can provide effective smoke removal resulting in dynamic scenes.

Keywords: Smoke removal · Dehazing · Dark Channel Prior · Smoke imaging model · Pixel compensation

1 Introduction

Natural and artificial disasters often critically damage our lives. In such disaster situations, we have a critical need for quick lifesaving actions, disaster investigations, and post-disaster monitoring. In these situations, it is often difficult to enter disaster areas because of unstable footing and poisonous gases. Thus, the uses of machines, such as drones or small robots, is effective in dealing with such disasters. Drones, particularly, are useful investigation tools for disaster scenes [1, 2]. They make it possible to obtain a large amount of information by flying over the affected area. Rescue robots can take many forms for searching through rubble and water [3, 4]. In such machines, on-board compact cameras are employed for scene recognition and autonomous actions. However, the performance of these cameras and machine vision algorithms are degraded because of smoke and other gases in the disaster areas. Because fog and haze as well as smoke reduce scene visibility, many dehazing methods have been proposed [5–11]. Tan proposed a single-image dehazing method to enhance the contrast [5]. Fattal presented an image dehazing method based on a haze imaging model [6]. He et al. restored haze image visibility based on the above haze imaging model and Dark

Channel Prior algorithm [7] (see next section). Gibson and Nguyen evaluated He’s approaches by using principal component analysis and minimum volume ellipsoid approximations [8]. Fattal proposed a dehazing method using colorlines [11], which realized better clarity than his previous method [7]. Video dehazing methods for video are often realized by extending previous single-image dehazing techniques [12–14]. Tarel et al. presented a fast dehazing algorithm based on a median filter, and applied it for video dehazing in vehicle cameras [12]. Zhang et al. used spatial and temporal coherence based on a Markov random field (MRF) model for reducing spatial veiling and temporal flicker [13]. Kim et al. presented video dehazing based on block-based restoration [14].

However, there are two problems in applying conventional approaches to video smoke removal. One is the spatial non-uniformity of smoke density. The conventional dehazing techniques assume that uniform scene haze covers all parts of an image. Moreover, it is assumed that conventional haze imaging models depend only on scene distance and do not take into account non-uniform haze and density that do not depend on distance. Further, single-image dehazing approaches cannot sufficiently remove partially covered strong fog and smoke in each frame. Another problem is the inappropriate reuse of the haze imaging model. Even though a smoke imaging model is actually different from the haze imaging model, some conventional methods have applied the haze imaging model not only for dehazing, but also smoke removal. For proper image/video smoke removal, a smoke imaging model should be constructed in the same way as the haze imaging model.

In this study, we propose a smoke imaging model and smoke removal method for video sequences. In our approach, the video camera moves freely, and partially covered smoke areas are temporally shifting. The scene and smoke do not keep relative positions. First, we remove the smoke from each frame. Next, we calculate corresponding pixels between the frames. For this calculation, we use SIFT and color features with distance constraints. Then, we compensate each pixel color by space-time weighting of adjacent frames. This paper is organized as follows: We describe the haze imaging model and a conventional dehazing approach in Sect. 2. The model and approach are the basis of our proposed method. Then, we propose a smoke imaging model and our smoke removal method in Sect. 3. In Sect. 4, we show experimental results and discussions. Moreover, we compare our method with conventional methods. Finally, conclusions and future research are discussed in Sect. 5.

2 Dehazing Model and Conventional Approach

Figure 1 shows transmission of light in a natural scene containing haze. In general, the haze imaging model is given by the following equation:

$$\mathbf{I}(\mathbf{x}) = \mathbf{J}(\mathbf{x})t(\mathbf{x}) + \mathbf{A}(1 - t(\mathbf{x})), \quad (1)$$

where \mathbf{x} is pixel coordinates in camera image is pixel coordinates in camera image \mathbf{I} , \mathbf{J} is scene radiance, \mathbf{A} is global atmospheric color, and t is medium

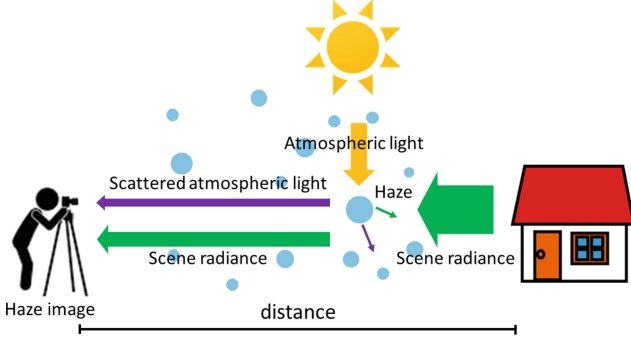


Fig. 1. Haze imaging model.

transmission of scene radiance. If an image does not contain scene haze, light from the scene objects reach the camera directly without any diffusions in the air. On the other hand, when haze is present in the air, the scene radiation is diffused by the haze prior to reaching the camera. In this situation, light scattered by the particles in the atmosphere also reach, as shown in Fig. 1. The transmission value t is defined by

$$t = \exp(-\beta \cdot d(\mathbf{x})), \quad (2)$$

where β is a diffusion coefficient, and d is the distance between objects and a camera. As shown in Eq. (2), haze is uniformly distributed in scenes, and depends only on distances. Input image I could be restored by estimating t , \mathbf{A} , and \mathbf{J} . He et al. found that at least one of the RGB values in a patch was very low (almost zero) when using an image under clear daylight [7]. This phenomenon, called Dark Channel Prior, is as follows:

$$J^{dark}(\mathbf{x}) = \min_{c \in r, g, b} (\min_{y \in \Omega(\mathbf{x})} J^c(\mathbf{y})) \simeq 0, \quad (3)$$

where Ω is a patch region of a pixel \mathbf{x} , and c is an RGB channel. Then, based on Eqs. (1) and (3), a transmission map is estimated:

$$\bar{t}(\mathbf{x}) = 1 - \omega \min_{c \in r, g, b} (\min_{y \in \Omega(\mathbf{x})} (\frac{I^c(\mathbf{y})}{A^c})), \quad (4)$$

where ω is a parameter for keeping some amount of haze for far-distant objects. In order to estimate atmospheric color \mathbf{A} , it is necessary to find a pixel of $t(\mathbf{x}) = 0$. Based on Eq. (2), the transmission value $t(\mathbf{x})$ will be 0 at the pixel of infinite distance $d(\mathbf{x}) \rightarrow \infty$. Assuming that the distance in the sky area will be infinite, they employ the brightest pixel in an input image as the sky area. The estimated transmission map \bar{t} generally contains block noise due to the patch-based processing. After the refinement of noisy transmission map \bar{t} by soft matting, scene radiance \mathbf{J} is estimated by

$$\mathbf{J}(\mathbf{x}) = \frac{\mathbf{I}(\mathbf{x}) - \mathbf{A}}{\max(t(\mathbf{x}), t_0)} + \mathbf{A}, \quad (5)$$

here t_0 is a lower limit transmission threshold for noise reduction.

3 Proposed Video Smoke Removal Method

In this study, we propose a novel video smoke removal method. The flow chart of our framework is shown in Fig. 2. As can be seen in this flowchart, the input is a video sequence of a smoke scene. The video smoke removal for the output is executed by compensating pixel colors based on space-time information. For realizing this purpose, first, we develop a smoke imaging model similar to a haze imaging model. Then, we apply a smoke removal method, frame by frame, based on the smoke imaging model and Dark Channel Prior [7] to calculate a smoke density map. In addition to a smoke density map, a detail layer is used for precise pixel selection. The detail layer is generated by applying the bilateral filter to an input frame. Then we align pixel positions between temporally-adjacent frames. Finally, we synthesize video frames using pixel selection maps based on smoke density maps and detail layers.

In the conventional method discussed in Sect. 2, there are several issues regarding video smoke removal. He et al.’s method assumed spatial uniformity of haze density. Their approach cannot sufficiently remove partially covered thick fog and smoke in each frame. Moreover, the haze imaging model was applied to smoke removal, in spite of the fact such a model was different from a smoke imaging model. Instead, we developed a smoke imaging model and a video smoke removal framework for addressing the above issues.

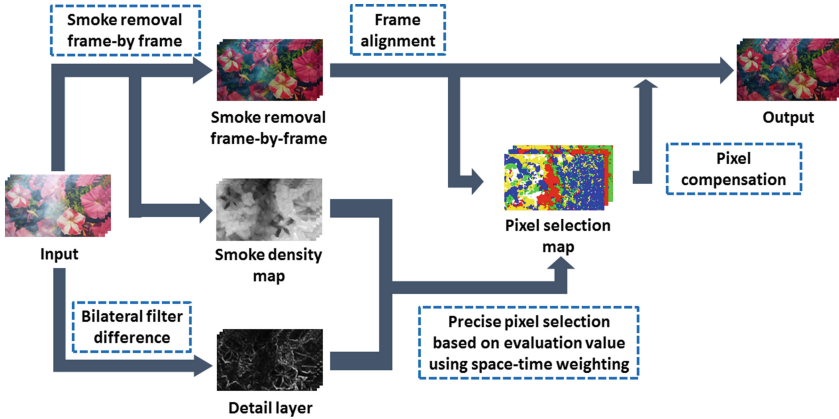


Fig. 2. Flowchart of the proposed algorithm.

3.1 Smoke Imaging Model

Figure 3 shows the imaging model for a scene containing haze and smoke. If input videos contain smoke, each frame can be represented by the sum of scene radiance, global atmospheric light, and light scattered by particles of smoke. Here, the smoke imaging model is given by

$$\mathbf{I}(\mathbf{x}) = (1 - \psi(\mathbf{x})) (\mathbf{J}(\mathbf{x})t(\mathbf{x}) + (1 - t(\mathbf{x})) \mathbf{A}) + \psi(\mathbf{x})\mathbf{S}, \quad (6)$$

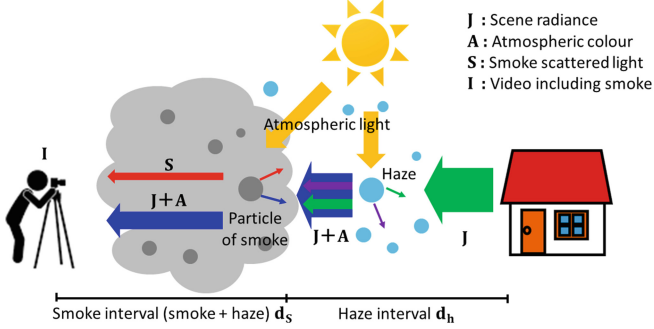


Fig. 3. Smoke imaging model. In a scene containing smoke, the three components (scene radiance, atmospheric light and light through smoke) reach the camera.

where \mathbf{x} , \mathbf{I} , \mathbf{J} , \mathbf{A} , and t are the same as Eq. (1). \mathbf{S} is smoke-scattered light color, and ψ is smoke density. This is a typical smoke imaging model containing both haze and smoke. Here, we assume that the smoke density ψ does not depend on distance d_s between objects and smoke. In addition, if the distance between objects and camera is sufficiently short, \mathbf{I} is not affected by the global atmospheric color \mathbf{A} due to scene haze. In other words, we can ignore the transmission ($t(\mathbf{x}) \approx 1$). In this situation, Eq. (6) can be rewritten as

$$\mathbf{I}(\mathbf{x}) = \mathbf{J}(\mathbf{x}) (1 - \psi(\mathbf{x})) + \psi(\mathbf{x})\mathbf{S}. \quad (7)$$

Here, let be $\rho(\mathbf{x}) = 1 - \psi(\mathbf{x})$, Eq. (7) can be formed as

$$\mathbf{I}(\mathbf{x}) = \mathbf{J}(\mathbf{x}) \rho(\mathbf{x}) + \mathbf{S} (1 - \rho(\mathbf{x})). \quad (8)$$

When setting the smoke density $\psi(\mathbf{x}) = 0$, scene radiance is not affected by smoke. On the other hand, when $\psi(\mathbf{x}) = 1$, camera image \mathbf{I} is equal to smoke color \mathbf{S} . Comparing Eq. (1) with Eq. (8), the smoke imaging model and haze imaging model can be substantially given by an equivalent expression. Thus, we estimate ψ and \mathbf{S} from \mathbf{I} for recovering the scene radiance \mathbf{J} . We can solve Eq. (8) by the same manner described in Sect. 2. In this method, we apply the Dark Channel Prior algorithm [7] in each frame. After the frame-by-frame smoke removal process, smoke remains in several regions. Figure 4 shows an example of smoke removal. As can be seen in this example, the visibility of the smoke-removed frame is better than that of the input frame. However, regions with smoke still remain. Thus, in the next step, we address a method to recover better visibility by using temporally-adjacent frames (See Sect. 3.3).

3.2 Frame Alignment with Distance and Color Constraints

SIFT features are often used to detect corresponding points between frames. However, in frames containing smoke, it is difficult to achieve accurate alignment

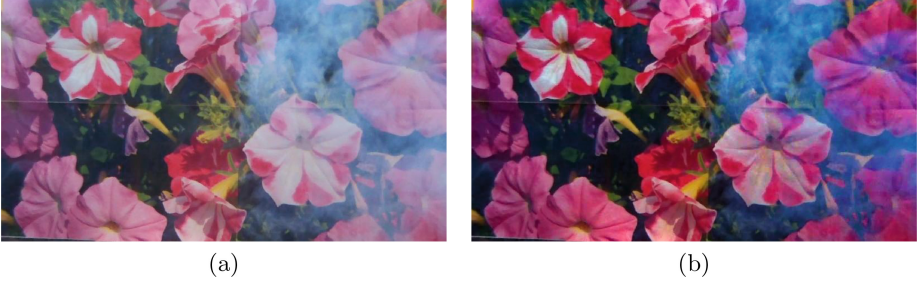


Fig. 4. Example of smoke removal; (a) input frame. (b) smoke-removed frame based on a smoke imaging model. This smoke removal is executed frame-by-frame (not using temporal information).

by using only SIFT features. Therefore, we add two constraints to SIFT for detecting robust corresponding points between smoke frames.

One constraint is to set a limitation on detection ranges. The amount of movement between frames can be assumed to be small. Thus, searching a feature point $k_{n'}$, which is corresponding to a feature point k_n is limited within the surrounding $h \times h$ pixels of the feature point k_n .

The other constraint is to use the color information of a patch. The corresponding points obtained by only SIFT features are few, because the pixel values affected by smoke are different in each frame. We use the RGB information of surrounding $l \times l$ pixels of a feature point. Then we employ the Euclidean distances of SIFT feature and color information for evaluating correspondences. The evaluation value E_{Align} is given by

$$E_{Align} = (1 - w)\varphi(\mathbf{v}^{k_n}, \mathbf{v}^{k_{n'}}) + w\varphi(\mathbf{p}^{k_n}, \mathbf{p}^{k_{n'}}), \quad (9)$$

$$\varphi(\mathbf{v}, \mathbf{v}') = \|\mathbf{v} - \mathbf{v}'\|_2, \quad (10)$$

where $\mathbf{v}^{k_n}, \mathbf{v}^{k_{n'}}$ are the SIFT features (128 dimensions) of points $k_n, k_{n'}$, respectively, in the frame n, n' . $\mathbf{p}^{k_n}, \mathbf{p}^{k_{n'}}$ are the RGB features ($3l^2$ dimensions), represented by the surrounding $l \times l$ pixels of feature points $k_n, k_{n'}$. w is a parameter containing the ratio of the Euclidean distance of SIFT features to color features. It is possible to obtain correct corresponding points by using smaller evaluation values $E_{Align} < th_{Align}$. Then, we calculate a homography matrix using RANSAC. When the number of the obtained corresponding points is too small, the homographic transformation cannot be correctly performed. In such situations, we do not use the frame for the pixel compensation. Figure 5 shows an example of corresponding point detection. As can be seen in Fig. 5, we obtain a correct homography matrix by using SIFT with the above two constraints.

3.3 Pixel Compensation with Space-Time Weighting

After frame alignment, we compensate pixel values by space-time weightings of corresponding pixels in smoke-removed frames. For using a precise pixel, we

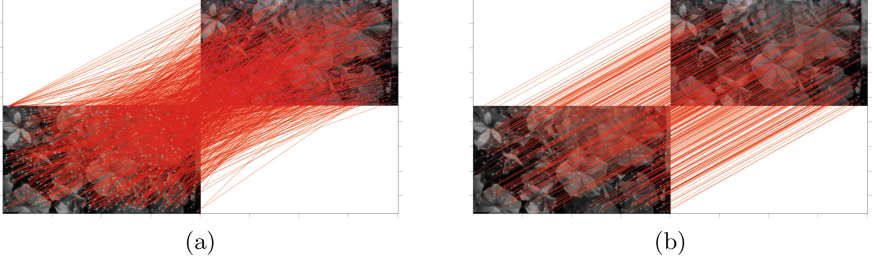


Fig. 5. Corresponding point detection of adjacent two frames; (a) SIFT features only. (b) SIFT features with distance and color constraints.

apply smoke density maps $\psi(\mathbf{x})$ as same as $t(\mathbf{x})$ in He et al.’s method [7]. In addition to smoke density maps, we use detail layers to evaluate the decrease of component detail caused by smoke. This is because smoke reduces details, as well as color saturations of a scene. We generate a detail layer by calculating the difference between the input frame and bilateral filtered frame as follows:

$$Y_D = Y - Y_B, \quad (11)$$

where Y_D , Y , and Y_B are a detail layer, an input frame and filtered frame, respectively. Then, we compensate pixel values based on the combinations of the smoke density maps and detail layers. The evaluation value E is given by

$$E(\mathbf{x}) = \lambda\rho(\mathbf{x}) + (1 - \lambda)Y_D(\mathbf{x}), \quad (12)$$

where λ is a parameter to control the weighting between a smoke density map and detail layer.

Pixel correspondence reliability is affected by spatial and temporal distances. Thus, we add spatial and temporal weights for compensating precise pixel values. The weighted evaluation value E_{weight} is given by

$$E_{weight}(\mathbf{x}, n, n') = G_t(n, n') \cdot G_s \cdot E(\mathbf{x}), \quad (13)$$

where $G_t(n, n')$ is the temporal Gaussian weight given by

$$G_t(n, n') = \frac{1}{2\pi\sigma_t^2} \exp\left(-\frac{|n' - n|}{2\sigma_t^2}\right), \quad (14)$$

and $E_s(\mathbf{x})$ is $E(\mathbf{x})$ in Eq. (12) with the spatial Gaussian weight given by

$$G_s \cdot E(\mathbf{x}) = \sum_{\mathbf{y} \in \Omega(\mathbf{x})} \lambda \cdot \rho(\mathbf{y}) \cdot g(\mathbf{y}, \sigma_s) + (1 - \lambda) \cdot Y_D \cdot g(\mathbf{y}, \sigma_s), \quad (15)$$

$$g(\mathbf{y}, \sigma_s) = \frac{1}{2\pi\sigma_s^2} \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|_2^2}{2\sigma_s^2}\right), \quad (16)$$

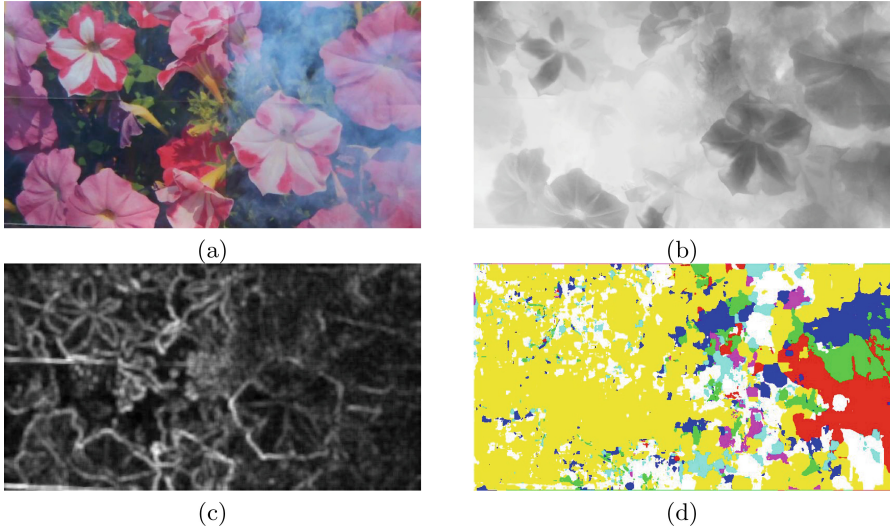


Fig. 6. Example of a pixel selection map; (a) input frame, (b) smoke density map, (c) detail layer, (d) pixel selection map (red: $n - 3$, green: $n - 2$, blue: $n - 1$, yellow: n , white: $n + 1$, cyan: $n + 2$ and magenta: $n + 3$ frame, respectively). (Color figure online)

where Ω is a patch of pixel \mathbf{x} , and σ_s, σ_t are parameters that control the space and time weightings. By selecting the pixel with the maximum evaluating value, we replace a pixel value of a current frame with one from temporally adjacent frames. We store selected frame numbers, which have precise pixel values in a pixel selection map. Figure 6 shows a pixel selection map using a smoke density map and detail layer. As described above, the pixel selection map is actually generated by using smoke density maps and detail layers of temporally adjacent frames. Finally, we synthesize smoke-removed frames via the pixel selecting maps.

4 Results and Discussion

In this study, we captured videos containing smoke by using a Drone camera (Parrot’s Bebob Drone). Smoke in a scene was generated using commercial fireworks. The drone is freely flown in the scene with smoke. When we executed the proposed method, the videos were resized from original 1920×1080 to down-sampled 800×450 pixels, in order to shorten the processing time. Parameters were set as shown in Table 1.

Figure 7 shows an example of our experimental results. In this figure, we used seven adjacent frames in the synthesis. As shown in Fig. 7(a), an input frame is fully covered by smoke. In particular, we cannot see a part of the tree on the right. As shown in Figs. 7(b) and (c), a smoke density map and a detail layer enable

Table 1. Parameter setting.

Smoke removal of frame-by-frame		
ω	Parameter in Eq. (4)	0.95
t_0	Transmission lower limit	0.1
Frame alignment with constraints		
h	Limit of detection range	61
l	Patch size	5
w	Weight of corresponding point	0.5
th_{Align}	Evaluation of corresponding point	1.0
Frame select and pixel compensation		
λ	Parameter for space-time weighting	0.7
σ_t	Parameter of temporal weighting	10
σ_s	Parameter of spatial weighting	0.64

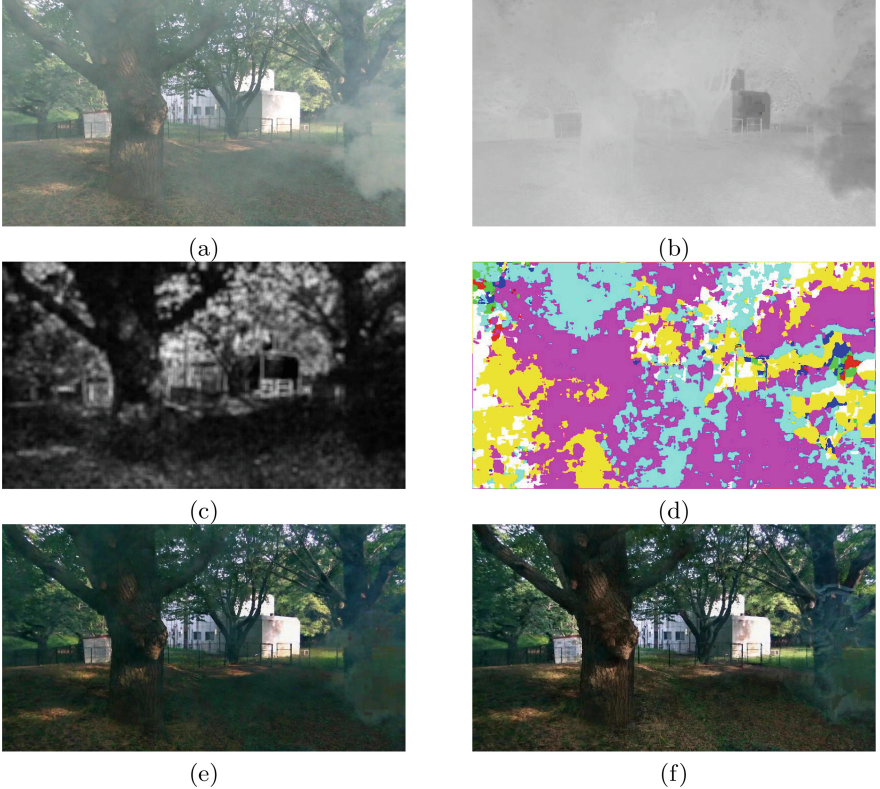


Fig. 7. Our input and result example; (a) input frame, (b) smoke density map, (c) detail layer, (d) pixel selection map (red: $n - 3$, green: $n - 2$, blue: $n - 1$, yellow: n , white: $n + 1$, cyan: $n + 2$ and magenta: $n + 3$ frame, respectively), (e) frame-by-frame smoke removal, (f) our final result. (Color figure online)

precise smoke detection. Then, a pixel selection map in Fig. 7(d) was generated based on smoke density maps and detail layers of temporally-adjacent frames. As shown in Fig. 7(d), we can see that pixel colors can be restored from temporally-adjacent frames. The result of smoke removal, frame-by-frame, in Fig. 7(e) has better visibility than that in Fig. 7(a). However, Fig. 7(e) still presents a dull appearance. On the other hand, as shown in Fig. 7(f), our method can restore a clear appearance in the tree on the left and fallen leaves on the ground. A part of the tree on the right was not fully restored because scene radiance information is almost lost in this dense smoke region.

Further, we recorded videos with and without smoke for comparing the ground truth with the smoke-removed results. The videos were recorded using a camera with constant motion and a panel in front of the camera. Figure 8 is a

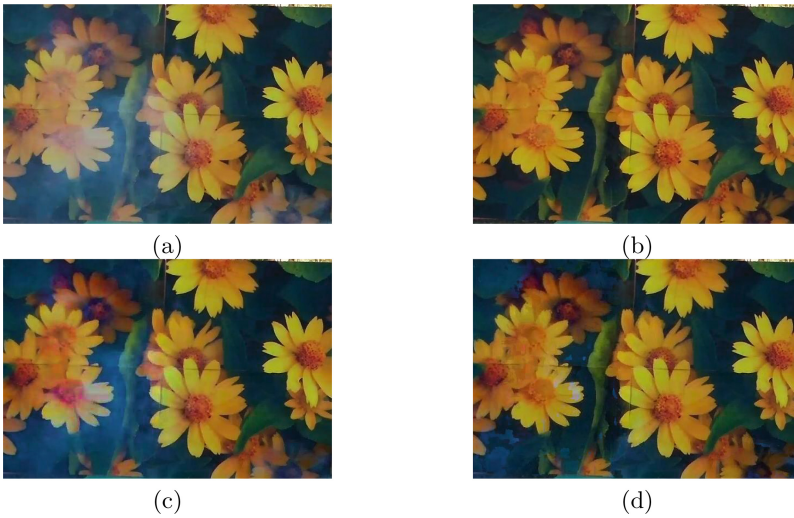


Fig. 8. Comparison of each method; (a) input frame containing the smoke, (b) ground truth, (c) He et al. [7], (d) our method.

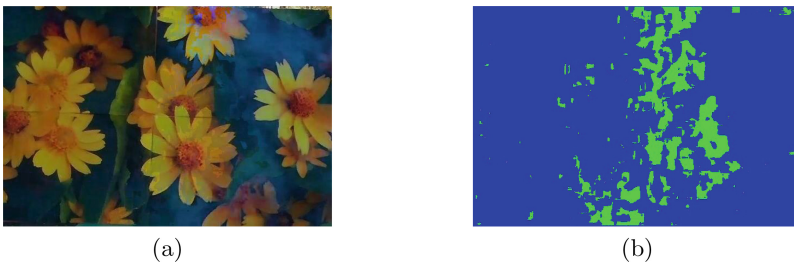


Fig. 9. Failure case of the proposed method; (a) result frame with smoke remaining, (b) pixel selection map using our algorithm.

comparison of the conventional dehazing methods and our result. In this figure, we used five adjacent frames in the synthesis. As can be seen in Fig. 8(c), smoke in the lower left corner was not completely removed by He et al.'s method [7]. Moreover, smoke in the other region was similarly not removed well by their method. Figure 8(d), using our pro-posed method, achieves removal of almost all the smoke. Figure 9 shows a smoke-removed frame and its frame selection map in a failure case. The smoke region remains in this result. This is because the number of detected corresponding points is too small. In this case, only two frames were used for pixel compensation.

5 Conclusion

In this paper, we have proposed an algorithm to remove smoke in a video by combining multiple frames. We described optical phenomena for natural scenes contain smoke. Then, we developed a smoke imaging model. Moreover, we applied dehazing methods in each frame, detected the corresponding point using SIFT with two constraints, and aligned frames. Finally, we selected the clearest pixels without smoke using the smoke density map and detail layer for synthesizing the smoke-removal frame. In our experiment, some smoke still remained in the video frame, because of the wrong correspondence of feature points between frames. We should improve the matching technique by brightness adjustment and additional image information.

References

1. Flyability. Introducing Gimball, the collision-tolerant drone, May 2016. <http://www.flyability.com/product/>
2. Search and Rescue with UAVS, May 2016. <https://www.microdrones.com/en/applications/areas-of-application/search-and-rescue/>
3. Popular Science. Watch google's humanoid robot learn the world is a harsh place, May 2016. <http://www.popsci.com/watch-googles-humanoid-robot-learn-world-is-harsh-place>
4. IEEE Spectrum. DARPA's Rescue-Robot Showdown, May 2016. <http://spectrum.ieee.org/robotics/humanoids/darpas-rescuerobot-showdown>
5. Tan, R.T.: Visibility in bad weather from a single image. In: Computer Vision and Pattern Recognition, pp. 1–8. IEEE Press, Alaska (2008)
6. Fattal, R.: Single image dehazing. ACM. Trans. Graph. **27**(3), 1–9 (2008). (Proceeding SIGGRAPH 2008)
7. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. In: Computer Vision and Pattern Recognition, pp. 1956–1963. IEEE Press, Miami (2009)
8. Gibson, K.B., Nguyen, T.Q.: On the effectiveness of the dark channel prior for single image dehazing by approximating with minimum volume ellipsoids. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1253–1256. IEEE Press, Prague (2011)

9. Yang, S., Zhu, Q., Wang, J., Wu, D., Xie, Y.: An improved single image haze removal algorithm based on dark channel prior and histogram specification. In: 6th International Conference on Model Transformation, pp. 279–292, Budapest (2013)
10. Tan, Z., Bai, X., Wang, B., Higashi, A.: Fast single-image defogging. *FUJITSU Sci. Tech. J.* **50**(1), 60–65 (2014)
11. Fattal, R.: Dehazing using color-lines. *ACM. Trans. Graph.* **34**(1), 1–14 (2014). (Proceeding SIGGRAPH 2015)
12. Tarel, J.P., Hautiere, N.: Fast visibility restoration from a single color or gray level image. In: 20th IEEE International Conference on Computer Vision, pp. 2201–2208. IEEE Press, Kyoto (2009)
13. Zhang, J., Li, L., Zhang, Y., Yang, G., Cao, X., Sun, J.: Video dehazing with spatial and temporal coherence. *Vis. Comput.* **27**(6), 749–757 (2011)
14. Kim, J., Jang, W., Sim, J., Kim, C.: Optimized contrast enhancement for real-time image and video dehazing. *J. Vis. Commun. Image R.* **24**(3), 410–426 (2013)