Viewing Scenes Occluded by Smoke

Arturo Donate and Eraldo Ribeiro

Department of Computer Sciences
Florida Institute of Technology
Melbourne, FL 32905, USA
{adonate, eribeiro}@cs.fit.edu

Abstract. In this paper, we focus on the problem of reconstructing images of scenes occluded by thick smoke. We propose a simple and effective algorithm that creates a single clear image of the scene given only a video sequence as input. Our method is based on two key observations. First, an increase in smoke density induces a decrease in both image contrast and color saturation. Measuring the decay of the high-frequency content in each video frame provides an effective way of quantifying the amount of contrast reduction. Secondly, the dynamic nature of the smoke causes the scene to be partially visible at times. By dividing the video sequence into subregions, our method is able to select the subregion-frame containing the least amount of smoke occlusion over time. Current experiments on different data sets show very promising results.

1 Introduction

In this paper, we focus on the problem of image recovery in static scenes containing large amounts of thick smoke. Our goal is to generate a single clear image of the scene given only a video as input. We assume that no previous knowledge of the underlying scene is available, and both the camera and the observed scene are static. We base our study on two observations. First, an increase in smoke density induces a decrease in the contrast of images. Secondly, the dynamic nature of the smoke causes the scene to be partially visible at times in different parts of the image. Under these conditions, the smoke will considerably reduce the visibility of the scene. This scenario occurs in various real-world situations such as explosions, fire-related incidents, and armed conflicts.

We propose an algorithm that effectively reconstructs a clear image of the occluded scene. Our algorithm begins by separating video frames into subregions. Each subregion is analyzed in the frequency domain as well as the saturation channel of the HSV colorspace. We use a highpass filter to compare high-frequency information about each frame. Our experiments show that measuring the decay of high-frequency content as well as the reduction in chrominance of an image provides a successful approach to determining the level of smoke occlusion in each frame. For each subregion, the algorithm finds the frame containing the least amount of occlusion. The final reconstructed image is a mosaic of all the clear frames. This paper was inspired by the work of Efros *et al.* [1] on removing image distortion caused by water waves.

G. Bebis et al. (Eds.): ISCV 2006, LNCS 4292, pp. 1666–1675, 2006.

Fig. 1. Subset of frames from an experimental video sequence

The method is tested on different data sets created in the laboratory, each containing varying amounts of smoke as well as different number of image frames in order to present the robustness of the method. Current results show great success in generating clear images of the underlying scenes.

The remainder of the paper is organized as follows. Section 2 provides a review of the literature. Section 3 describes some of the effects of smoke occlusion in videos. An overview of the method is presented in Section 4. Experiments and results are shown in Section 5. Finally, Section 6 concludes the paper and presents directions for future research.

2 Related Work

The computer vision literature directly related to the problem of recovering a clear image of a smoke-occluded scene is somewhat limited. Most of the related works address the problems of smoke and fire detection in both digital [2] and satellite imagery [3,4], removal of weathering effects [5,6,7,8,9,10], and smoke synthesis [11]. The computer graphics community, on the other hand, has mostly addressed the problem of modeling and rendering realistic smoke.

Recently, there has been significant interest in computer vision methods for removing weathering effects such as fog and haze from images. An elegant modeling of the effects of weather conditions in images was proposed by Narasimhan and Nayar [5,6]. Their work is based on physical models that describe how light, colors, and contrast behave under certain weather conditions. In [5], Narasimhan and Nayar present two physics-based models that can be used for contrast restoration in images containing uniformly bad weather. The attenuation model describes how light weakens as it travels from a point on the scene to the observer, and the airlight model measures how the atmosphere can reflect environmental illuminations to the observer. These models provide a way of quantifying the decay in contrast of images with poor visibility conditions due to weather effects. In [6], Narasimhan and Nayar use the models introduced in [5] for removing weather effects such as fog and haze from a single image of a scene without

precise knowledge of the weather, and with minimal user input. Similarly, in [8] Narasimhan and Nayar present models for extracting the 3D structure of a scene occluded by bad weather. In the work presented in [10], Shwartz *et al.* approach the problem of blindly recovering the parameters needed for separating airlight from other measurements. Here, the authors describe a method that successfully recover the contrast of a hazy or foggy image with no user interaction and without having the sky in the image frame.

Other related problem is the automatic detection of smoke and fire in video sequences and satellite images. Treyin *et al.* [2] developed a real-time image-based smoke detection system that uses measurements on edge sharpness. The authors relate the effects of smoke's varying transparency to changes in image edge sharpness and saturation values of colors in the scene using a wavelet decomposition technique. The method requires the availability of a clean view of the observed scene. The problem of detecting smoke and fire in satellite imagery was addressed by [4,3]. Chung and Le [4] studied the possibility of automatic detection forest fire smoke plumes in satellite images. Farser *et al.* [3] use a neural network to analyze satellite imagery in order to classify the scenes into smoke, cloud, or clear background, in an attempt to automatically detect forest fires.

3 Effects of Smoke in Videos

In this section, we discuss some of the main effects of smoke in video sequences. Here, we focus on three main effects. First, we discuss how smoke occlusion causes variations in color saturation. Secondly, we study the effects of smoke on the image's frequency domain representation. Finally, smoke motion gives rise to image edges of weak intensity while weakening the response of static scene edges. Our method combines measurements of color saturation, frequency-domain energy, and edge maps to help select the best set of subregions of a frame sequence that can be stitched into a single clear view of the observed scene. In the analysis that follows, we assume that we have a video of a static scene occluded with a layer of moving smoke of varying thickness levels.

Smoke transparency and color. Smoke transparency will allow the underlying captured scene to be visible sporadically over time depending on the smoke thickness. Previous works have shown that variations in smoke transparency can be directly related to variations in both color contrast and color saturation. For example, Narasimhan and Nayar [5,6] provide color chrominance models that explain how these changes can be used to restore weathered images containing fog and haze. A decrease in chrominance values (i.e., hue and saturation components) has also been associated with the presence of smoke in images [2]. Figure 2 shows a plot of the total image saturation for a series of video frames with increasing levels of smoke occlusion. In the figure, the total saturation value decreases with the increase of smoke occlusion. The original video frames are also shown in the figure.

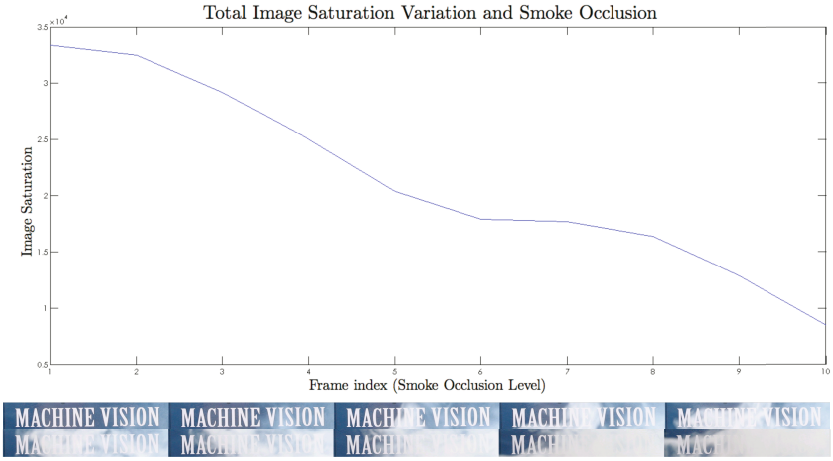


Fig. 2. Decreasing saturation due to smoke occlusion

Frequency domain analysis. The presence of the smoke in the observed scene causes a decrease in the color contrast of the video frames. The contrast reduction effect translates to a corresponding reduction in the image's high-frequency content [2,5,6,7,9]. The varying degrees of smoke thickness correspond to changes in both image transparency and contrast variation. This suggests that the analysis of the high-frequency content in the power spectrum of each frame in the video can provide a way to determine the level of smoke occlusion in each frame relative to the other frames of the video. The high-frequency power spectrum of an image can be computed by:

$$P'(u,v) = \|F(u,v)H(u,v)\|^2 \tag{1}$$

where F the Fourier transform of the image $I(x,y)$ defined by:

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x,v)e^{-2\pi j(ux+vy)}dxdy \tag{2}$$

and H is a high-pass filter. In this paper, we use a simple Butterworth high-pass filter [12] with cut-frequency D_0:

$$H(u,v) = \frac{1}{1+[D_0/D(u,v)]^2} \tag{3}$$

Finally, we can express the remaining high-frequencies in polar coordinates using the function $P(r,\theta)$, where P is the power spectrum function, and r and θ are the polar coordinate parameters [12]. One-dimensional histograms of the polar power spectrum can then be formed by:

$$S(r) = \sum_{\theta=0}^{\pi} S(r,\theta) \tag{4}$$

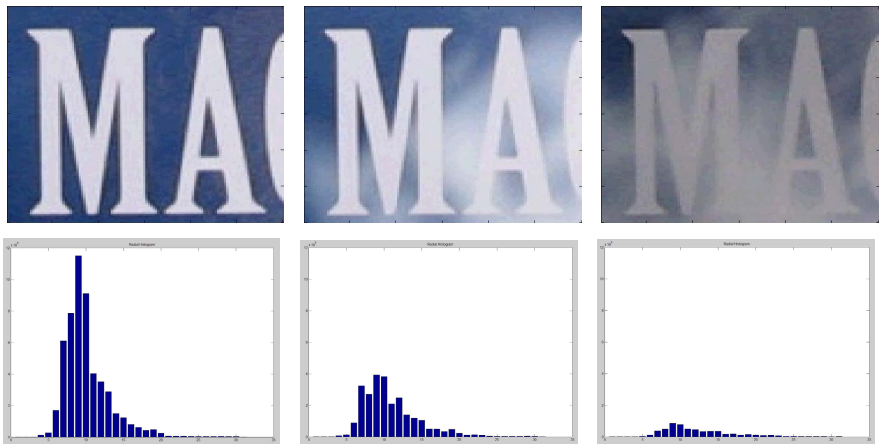


Fig. 3. Images with increasing levels of occlusion and their corresponding angular frequency histograms. The reduction in contrast causes a drop in the high-frequency content.

The total amount of high-frequency energy for each video frame is obtained by summing the bins of these histograms. Figure 3 illustrates this concept.

It is worth mentioning that measuring the decay of high-frequency content alone will not work well for quantifying the level of contrast reduction in images of different scenes. As pointed out by Field and Brady [13], a simple increase in the number of edges will result in an increase of high-frequency content, with no visual change in the degree of contrast in the images. This limitation may be overcome by using phase coherence measurements as described in [14]. A discussion on using the advantages of phase coherence is beyond the scope of this paper.

Edge accumulation map. The effects of chrominance and high-frequency variation represent measurements of the presence of smoke in an image. However, in the case of smoke occlusion, the high-frequency content attenuation is not caused by actual blur of edges. As a result, image edges will not suffer from localization uncertainty but they will appear weaker in some regions across the image. Additionally, for static scenes, some noisy edges will be temporally created by the gradient between the smoke cloud and a flat image background. These weak edges are likely to increase the image's high-frequency content. Here, we suggest the creation of an edge map obtained from the accumulation of edge evidence over the whole frame sequence. The information produced by this edge map will be used by our algorithm and it will be discussed later in this paper. An example of the accumulated edge map is shown in Figure 4.

Fig. 4. Accumulated edge map of static and dynamic "smokey" edges. Weaker edges are produced by the moving smoke.

4 Method Overview

In this section, we describe the details of our method. The main goal of our algorithm is to create a single clear image of an observed scene originally occluded by smoke given a video as input. This is accomplished by dividing all video frames into a number of non- overlapping subregions. These subregions will become "smaller videos" containing the same number of frames as the original video. The algorithm analyzes each subregion separately to select the single subregion-frame containing the least amount of smoke. In order to measure the relative amount of smoke in the image, we propose a function that combines chrominance information and frequency-domain measurements.

We commence by considering a video of N frames divided into K subregions. Our method of reconstructing the scene analyzes each subregion individually to select the frame containing the least amount of smoke occlusion. Let $q(\mathbf{x})$ be a function that measures the quality of a given image frame $\mathbf{x}$ such that an increase in the amount of smoke occlusion in $\mathbf{x}$ causes a reduction in the value of $q(\mathbf{x})$. We define $q(\mathbf{x})$ as follows:

$$q(\mathbf{x}) = \left\| \sum_{r} S(r) \right\|^{2p} \times \sum \text{saturation}(\mathbf{x}) \tag{5}$$

where

$$p = \begin{cases} 1, \text{edge}(\mathbf{x}) = 1 \\ 0, \text{otherwise} \end{cases}$$

Equation 5 has two main components. The first component corresponds to the total energy of the radial power spectrum representation of the image after the application of a high-pass filter as given by Equation 4. The second component corresponds to the total sum of the values in the saturation channel of the image. Here, we make two critical observations about the interplay between these two components. First, as the amount of smoke occlusion increases, the saturation values of the images decrease. While this property provides a simple approach for detecting smoke, it does not work well if the original colors in the scene have low saturation values. On the other hand, if the underlying scene contains no edge gradients, image frames with smoke will contain more high-frequency energy than clear frames (i.e., frames without smoke). This is due to the fact that

Algorithm 1. Smoke-occluded scene reconstruction

Given N-frame video sequence:

```
Divide video into K subregions
for Every subregion do
    Build an edgemap
    for Every frame in the subregion do
        Calculate high-frequency from polar histograms
        Calculate the sum of the values in the saturation channel
        Calculate a quality measurement from frequency and saturation measurements
    end for
    Extract frame with greatest quality measurement
end for
Generate reconstructed image I from retrieved subregions
```

partial smoke coverage will form a gradient between the boundary of the smoke and the flat background of a video frame. As a result, measuring smoke occlusion using high-frequency content only can produce erroneous results when analyzing image subregions with no edge gradients on the original image. We address this problem by adding the exponential term p in the first component of Equation 5. This term determines whether the high-frequency measurements should be taken into consideration or not. We set p to one if the analyzed subregion contains any edges in the accumulated edge map as described previously in this paper, and zero otherwise. The idea of this term is to allow the influence of high frequency measurements only when the subregion being analyzed contain static edges. If no edges are present, p will be set to zero and the quality of the frame will depend solely on color saturation. With $q(\mathbf{x})$ defined, the problem of finding the clearest frame in a subregion can be posed as a subregion-frame selection algorithm such that:

$$i' = \arg\max_i q(\mathbf{s}_i) \tag{6}$$

where $\mathbf{s}$ is the subregion containing the N frames ($\mathbf{s}_i \in \mathbf{s}$), and i' is the index corresponding to the frame in $\mathbf{s}$ with the least amount of smoke occlusion. Therefore, the problem now becomes a selection of subregion with maximum quality given by the measurement $q(\mathbf{x})$. The main steps of this method are listed in Algorithm 1. This process is applied to each subregion of the video. Once the "smokeless" subregion selection process is completed, a mosaic is formed from all the selected frames to create the final resulting image of the observed scene. The algorithm does not require any previous knowledge of the underlying scene.

5 Experiments and Results

In this section, we present experimental results on two different data sets recorded in our laboratory. Our objective is to show the effectiveness of the proposed method in selecting the cleanest frame- regions in the videos.

(a) original frame sequence

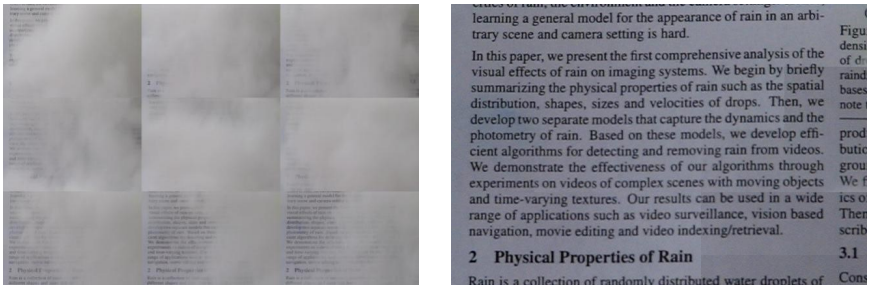
(b) mean (c) median (d) our method

Fig. 5. Subset of frames showing occlusion due to smoke with reconstructions using (a) a simple temporal average pixel value method, (b) the median over time, and (c) our method

In the experiments, we analyze three 100-frame video sequences containing varying amounts of smoke occlusion. Some frames contain a small amount of smoke, while other frames contain very thick layers of smoke that completely obscure the observed scene. The first dataset consists of a video of a machine vision book with an overall lighter quantity of smoke occlusion. A selection of the frames in this dataset is shown in Figure 5(a). One of the simplest ways to approach this problem is to calculate the mean (or median) of each pixel over time. This is a very simplistic approach that will result in poorly reconstructed images, especially in the case of video sequences containing scene occlusions for the majority of the frames due to heavy clouds of smoke (as seen in Figure 5). We present the results for the mean and median calculations only as an example.

The second dataset is a video of a printed section of text from the paper in [7]. The frames in this dataset contain smoke in larger amounts and of larger density than the first dataset. Here, the scene was under occlusion for approximately 95% of the video, and each section of the scene was only visible for a few frames over the entire sequence. The purpose of this experiment was to show that increasing the number of occluded frames has no ill effect on the algorithm, as long as each subregion of the scene is clearly visible in at least one frame of the video. Additionally, these video frames contain mostly frequency information (i.e., saturation values over the frames were relatively low).

The third dataset is a video containing the largest amount of smoke occlusion. This time, the camera was placed further back in order to cover a larger area of the scene. The camera position allowed for capture of larger textured and untextured regions in the same scene. Results are shown in Figure 7. In this dataset, there is only a single frame that does not contain smoke. The idea behind this experiment is to show the effectiveness of the method to find a clear



(a) (b)

Fig. 6. Subset of video frames from paper experiment. (a) Sample of original video frames. (b) Final reconstructed image.

(a) (b)

Fig. 7. Dataset containing nearly-complete occlusion. (a) Original video frames. (b) Final reconstructed image.

frame of the scene regardless of the amount of occlusion in the other frames. Although the complexity of this dataset is greater than the previous ones, the algorithm is equally successful in generating an accurate reconstruction.

6 Conclusions

In this paper, we focus on the problem of image recovery in static scenes containing large amounts of smoke occlusion. Our main goal was to develop an algorithm to generate a single clear image of the scene given only a video as input, with no previous knowledge about the underlying scene.

Our analysis shows that smoke occlusion causes a reduction in the frequency content of images, as well as a reduction of chrominance values. Our proposed algorithm attempts to quantify the relative amount of occlusion in a given frame by measuring reductions in chrominance as well as high-frequency content. As a result, our algorithm successfully reconstructs clear images of occluded scenes using only a video sequence as input. The experiments presented here show the

feasibility of our approach. It is important to note that for a given subregion, if the input video does not contain a single clear view, then the algorithm will fail to generate a clear view of that subregion in the final reconstruction since it does not attempt to actually remove the smoke present in the scene.

We are currently working on an extension to the proposed method to work with a moving cameras. This is a much more complex problem as feature-based image alignment and mosaicking methods might are likely to fail due to possible erroneous feature detection on the moving smoke regions.

References

1. Efros, A., Isler, V., Shi, J., Visontai, M.: Seeing through water. In Saul, L.K., Weiss, Y., Bottou, L., eds.: Advances in Neural Information Processing Systems 17. MIT Press, Cambridge, MA (2004) 393–400
2. Treyin, B.U., Dedeoglu, Y., Cetin, A.E.: Wavelet based real-time smoke detection in video. In: European Signal Processing Conference. (2005)
3. Li, Z., Khananian, A., Fraser, R.H., Cihlar, J.: Automatic detection of fire smoke using artificial neural networks and threshold approaches applied to avhrr imagery. IEEE Transactions on Geoscience and Remote Sensing **39** (2001) 1859–1870
4. Chung, Y.S., Le, H.V.: Detection of forest-fire smoke plumes by satellite imagery. Atmos. Environ. **18** (1984) 2143–2151
5. Narasimhan, S.G., Nayar, S.K.: Contrast restoration of weather degraded images. IEEE PAMI **25** (2003) 713–724
6. Narasimhan, S.G., Nayar, S.K.: Interactive deweathering of an image using physical models. In: IEEE Workshop on Color and Photometric Methods in Computer Vision, In Conjunction with ICCV. (2003)
7. Garg, K., Nayar, S.K.: Detection and removal of rain from videos. In: International Conference on Computer Vision and Pattern Recognition. (2004) 528–535
8. Nayar, S.K., Narasimhan, S.G.: Vision in bad weather. In: International Conference on Computer Vision. (1999)
9. Schechner, Y.Y., Narasimhan, S.G., Nayar, S.K.: Polarization-based vision through haze. Applied Optics **42** (2003) 511–525
10. Shwartz, S., Namer, E., Schechner, Y.Y.: Blind haze separation. In: International Conference on Computer Vision and Pattern Recognition. (2006)
11. Fedkiw, R., Stam, J., Jensen, H.W.: Visual simulation of smoke. In: SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, New York, NY, USA, ACM Press (2001) 15–22
12. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Addison-Wesley, Reading, MA (1992)
13. Field, D., Brady, N.: Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. Vision Research **37** (1997) 3367–3383
14. Wang, Z., Simoncelli, E.P.: Local phase coherence and the perception of blur. In Thrun, S., Saul, L., Schölkopf, B., eds.: Advances in Neural Information Processing Systems 16. MIT Press, Cambridge, MA (2004)