Adaptive Background Defogging with Foreground Decremental Preconditioned Conjugate Gradient

Jacky S-C. Yuk and Kwan-Yee K. Wong

Dept. of Computer Science
The University of Hong Kong, Hong Kong

Abstract. The quality of outdoor surveillance videos are always degraded by bad weathers, such as fog, haze, and snowing. The degraded videos not only provide poor visualizations, but also increase the difficulty of vision-based analysis such as foreground/background segmentation. However, haze/fog removal has never been an easy task, and is often very time consuming. Most of the existing methods only consider a single image, and no temporal information of a video is used. In this paper, a novel adaptive background defogging method is presented. It is observed that most of the background regions between two consecutive video frames do not vary too much. Based on this observation, each video frame is firstly defogged by a background transmission map which is generated adaptively by the proposed foreground decremental preconditioned conjugate gradient (FDPCG). It is shown that foreground/background segmentation can be improved dramatically with such background-defogged video frames. With the help of a foreground map, the defogging of foreground regions is then completed by 1) foreground transmission estimation by fusion, and 2) transmission refinement by the proposed foreground incremental preconditioned conjugate gradient (FIPCG). Experimental results show that the proposed method can effectively improve the visualization quality of surveillance videos under heavy fog and snowing weather. Comparing with the state-of-the-art image defogging methods, the proposed method is much more efficient.

1 Introduction

Outdoor surveillance videos are always degraded by challenging bad weathers, e.g., haze, fog, raining, snowing, etc. Some degradations, like under the haze and fog weathers, are mainly due to light absorption and scattering by atmospheric particles. The light from the viewing objects is being partly absorbed before it reaches the camera. The farther the objects from the camera, the more the light is being absorbed. The degraded videos always have low contrast and bad color fidelity. These degraded videos not only produce poor visualizations, but also make further vision-based analysis, such as foreground/background segmentation, more difficult. There are desires to improve the visual qualities of surveillance videos under hazy or foggy weathers. The goal is not only for better

visualizations, but also improve the correctness of the further higher level video analysis.

Image haze/fog removal techniques have been researched for more than a decade. The defogging problem on a single image is under-constrained due to lack of depth information. Early researches required multiple images of the same scene under different exposures (e.g., under different weather conditions [1, 2], or different degree of polarization [3]) to recover a foggy scene. Although these methods can significantly improve the visual quality, manual works are always required to prepare suitable images under different conditions for defogging.

Later on, single image defogging [4–8] got great progress and success. Based on the assumption that non-foggy image patches usually have a high contrast, Tan [4] proposed to recover a foggy image by maximizing the local contrast. Tan's method produces nice defogging results, but his assumption may not be physically correct. Fattal [5] assumed the transmission and image shading were locally uncorrelated. He estimated the albedo values and inferred the medium transmission by MRF. Fattal's approach, however, may fail under heavy fog scenarios. He *et al.* [6, 9] proposed the state-of-the-art dark channel prior for estimating image transmissions which are refined by soft matting. Although single image defogging is now pretty mature, existing methods are seldom applied to defog video sequences. Most of the methods only target at defogging a single image, and no temporal information of the video sequences is considered. Without temporal information, each video frame has to be processed individually, and this makes the defogging procedure very time consuming (Tan's method [4] required 5 minutes to process a frame, while the methods of Fattal [5] and He *et al.* [6, 9] require about 20 to 30 seconds per frame).

Recently, Dong *et al.* [10] proposed to locate the foreground regions on foggy video frames by comparing the foreground and background transmission maps. Their method, however, requires manually selecting 2 foreground-free scenes under different weather conditions for calculating the background transmission map. Once the background transmission map has been calculated, there will not be any further update on the map, and therefore, the method is not able to tolerate any background change.

This paper proposes a novel adaptive fog removal method for foggy surveillance video scenes. Based on the observation that most of the background regions between consecutive video frames will not vary too much, a video frame is firstly defogged by a background transmission map (fig. 1) which is generated and updated adaptively by the proposed foreground decremental preconditioned conjugate gradient(FDPCG). FDPCG targets at reducing the influence of foreground regions during the estimation of the transmission map. The background-defogged frame is then processed by foreground/background segmentation to generate the foreground map. The foreground regions in a background-defogged frame could vary from nearly fog-free, when foreground and background are nearly at the same depth (fig. 1 (a)), to extremely dark, when the depth between foreground and background is large (fig. 1 (b)). Both cases can make the further foreground/background segmentations easier.

The rest of the paper is organized as follows. Section 2 briefly describes the transmission model and the state-of-the-art dark channel prior defogging. Section 3 introduces the adaptive background defogging by the proposed FDPCG. Section 4 describes the foreground transmission estimation. Experimental results are presented in Section 5, followed by conclusions in Section 6.

(a) Background-defogging when the foreground and background are nearly at the same depth.

(b) Background-defogging when the depth between foreground and background is large.

Fig. 1. Examples of background defogging. The first column shows the original foggy video frames, followed by the corresponding background transmission maps and the background-defogged frames in column 2 and 3, respectively.

2 Dark Channel Prior Defogging

2.1 Transmission Model

In computer vision and graphics, a haze/fog image is widely formulated by the following transmission model [4–6, 9],

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

where $I(x)$ and $J(x)$ are the observed intensity and the fog-free scene radiance at pixel x, respectively. A is the global air light, and t is the transmission value which describes the portion of the light finally reaching the camera. The target of fog removal is to recover the fog-free scene radiance, J, according to the estimated t, and the observed I and A,

$$J(x) = \frac{I(x) - A}{t(x)} + A. \tag{2}$$

As suggested in [6, 9], $t(x)$ is lower bounded by $t_0 = 0.1$, so that the recovered image will not be too dim,

$$J(x) = \frac{I(x) - A}{\text{MAX}(t(x), t_0)} + A. \tag{3}$$

2.2 Dark Channel Prior

Dark channel prior was firstly proposed in [6]. It was shown to be able to effectively predict the transmission map based on the observation that in most of the non-foggy image patches, at least one color channel has very low intensity at some pixels. The dark channel, J^{dark}, at pixel x is defined as

$$J^{dark}(x) = \text{MIN}_{c \in \{r,g,b\}}(\text{MIN}_{\{y \in \Omega(x)\}}(J^c(y))), \tag{4}$$

where $J^c(y)$ is one of the RGB color channels at pixel y and $\Omega(x)$ is the local image patch centered at x.

The air light, A, is assumed to be a non-zero constant, and A^c, J^c and I^c of a particular color channel, c, are coplaner [6]. By applying the dark channels to the transmission model (1), the estimated transmission at pixel x, $\tilde{t}(x)$, can be derived as

$$\tilde{t}(x) = \frac{J^{dark}(x)}{A^{dark}}\tilde{t}(x) + 1 - \frac{I^{dark}(x)}{A^{dark}}. \tag{5}$$

Since $\tilde{t}(x)$ is ranging between [0, 1], and the non-foggy dark channel $J^{dark}(x)$ should have very low intensity, $J^{dark}(x) \to 0$, so $\tilde{t}(x)$ can then be directly calculated as

$$\tilde{t}(x) = 1 - \text{MIN}_c(\text{MIN}_{y \in \Omega(x)}\frac{I^c(y)}{A^c}). \tag{6}$$

2.3 Soft Matting

The transmission model (1) is similar to the alpha matting problem. This allows using soft matting [11] to refine the transmission map by treating the transmission map as an alpha map. This refinement can be done by minimizing the following cost function,

$$E(\mathbf{t}) = \mathbf{t}^T\mathbf{L}\mathbf{t} + \lambda(\mathbf{t} - \tilde{\mathbf{t}})^T(\mathbf{t} - \tilde{\mathbf{t}}). \tag{7}$$

where $\mathbf{t}$ and $\tilde{\mathbf{t}}$ are the refined and predicted transmission map, respectively. Both $\mathbf{t}$ and $\tilde{\mathbf{t}}$ are in l = width $\times$ height dimensions. λ is for regularization. $\mathbf{L}$ is a $l \times l$ dimensional Matting Laplacian matrix [11]. The (i,j)-th element of L is defined as

$$L_{i,j} = \sum_{k|(i,j)\in\omega_k} (\delta_{ij} - \frac{1}{|\omega_k|}(1 + (\mathbf{I}(i) - \mu_k)^T(\mathbf{\Sigma}_k + \frac{\varepsilon}{|\omega_k|}\mathbf{U}_3)^{-1}(\mathbf{I}(j) - \mu_k))), \tag{8}$$

where $\mathbf{I}(i)$ is the 3-dimensional RGB color at pixel i. δ_{ij} is the Kronecker delta. μ_k and $\mathbf{\Sigma}_k$ are the mean and covariance matrix of the colors in window ω_k. $\mathbf{U}_3$ is a 3x3 identity matrix. ε is for regularization, and $|\omega_k|$ is the number of pixel in ω_k.

3 Adaptive Background Defogging

3.1 Preconditioned Conjugate Gradient

The optimal $\mathbf{t}$ in (7) can be obtained by solving

$$(\mathbf{L} + \lambda\mathbf{U})\mathbf{t} = \lambda\tilde{\mathbf{t}}, \tag{9}$$

where $\mathbf{U}$ is $l \times l$ identity matrix, and $\lambda = 10^{-4}$ is a small constant so that $\mathbf{t}$ is softly constrained by $\tilde{\mathbf{t}}$.

Equation (9) can be solved by preconditioned conjugate gradient (PCG) method,

$$\mathbf{M}^{-1}(\mathbf{L} + \lambda\mathbf{U})\mathbf{t} = \mathbf{M}^{-1}(\lambda\tilde{\mathbf{t}}), \tag{10}$$

where $\mathbf{M}$ is $l \times l$ preconditioning matrix. The main purpose of the preconditioning matrix is to help the PCG converge faster. In this paper, $\mathbf{M}$ is chosen to be a diagonal Jacobi preconditioner [12]. The element, $m_{i,j}$, of the Jacobi preconditioner is defined as

$$m_{i,j} = \begin{cases} a_{i,i} & \text{if } i = j \\ 0 & i \neq j \end{cases}, \tag{11}$$

where $a_{i,i}$ is the i-th diagonal element of $(\mathbf{L} + \lambda\mathbf{U})$. Jacobi precontioner simply normalizes the i-th row of the matrix in (10) by its i-th coefficient value. The process of iterative PCG is presented in algorithm 1.

Algorithm 1 Preconditioned Conjugate Gradient (PCG). The initial transmissions $\mathbf{t}_0$ is initialized to $\mathbf{t}_{\text{init}}$, where each element of $\mathbf{t}_{\text{init}} = 0.2$.

Initialization:
$k = 0$, $\mathbf{t}_0 = \mathbf{t}_{\text{init}}$,
$\mathbf{r}_0 = \lambda\tilde{\mathbf{t}} - (\mathbf{L} + \lambda\mathbf{U})\mathbf{t}_0$, $\mathbf{z}_0 = \mathbf{M}^{-1}\mathbf{r}_0$, $\mathbf{d}_0 = \mathbf{z}_0$
$err_0 = \mathbf{z}_0^T\mathbf{r}_0$

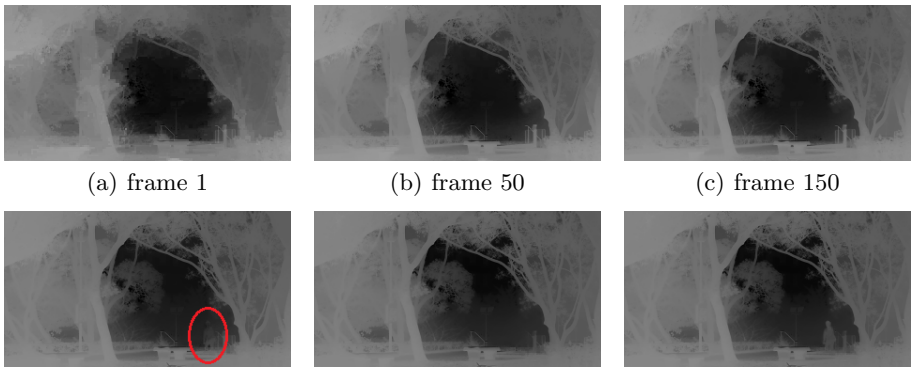
Iteration:
while $err_k > \epsilon$ and $k < K$ {

$$\begin{aligned}
\alpha_k &= \frac{\mathbf{z}_k^T\mathbf{r}_k}{\mathbf{d}_k^T(\mathbf{L}+\lambda\mathbf{U})\mathbf{d}_k} \\
\mathbf{t}_{k+1} &= \mathbf{t}_k + \alpha_k\mathbf{d}_k \\
\mathbf{r}_{k+1} &= \mathbf{r}_k - \alpha_k(\mathbf{L} + \lambda\mathbf{U})\mathbf{d}_k \\
\mathbf{z}_{k+1} &= \mathbf{M}^{-1}\mathbf{r}_{k+1} \\
\beta_{k+1} &= \frac{\mathbf{z}_{k+1}^T\mathbf{r}_{k+1}}{\mathbf{z}_k^T\mathbf{r}_k} \\
\mathbf{d}_{k+1} &= \mathbf{z}_{k+1} + \beta_{k+1}\mathbf{d}_k \\
err_{k+1} &= \mathbf{z}_{k+1}^T\mathbf{r}_{k+1} \\
k &= k + 1
\end{aligned}$$

}

3.2 Adaptive Foreground Decremental Preconditioned Conjugate Gradient

The iterative PCG (alg. 1) will stop either when the error, err_k, is smaller than ϵ or the algorithm reaches maximum number of iterations, K. In our experiment, when K was not limited and ϵ was set to $\text{MAX}(err_0 \times 0.001, 10^{-7})$, the PCG algorithm converged after around 700 to 850 iterations for each frame.



(a) frame 1 (b) frame 50 (c) frame 150

(d) frame 700 (e) background transmission map of frame 700 (f) complete transmission map of frame 700

Fig. 2. (a)-(c) The iterative PCG converges over frames. More and more details at background regions are recovered in later frames. (d) Defects at the foreground regions (red circled) are caused by non-completed PCG. (e), (f) The corresponding FDPCG background transmission map and final transmission map, respectively.

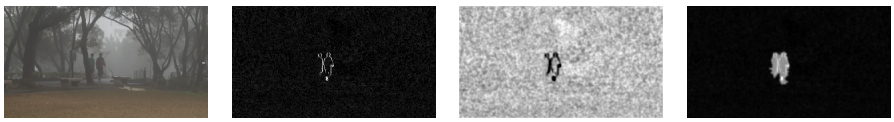
As the characteristics of surveillance video using a static camera, the background between consecutive frames does not change a lot. This suggests that the iterative part of PCG algorithm can be applied over frames. To achieve this, we slightly modify the initialization of $\mathbf{t}_0$ in PCG (alg. 1). We set $\mathbf{t}_0 = \mathbf{t}^{fn-1}$, where $\mathbf{t}^{fn-1}$ is the refined transmission map of the previous frame. The maximum iterations K is also limited to 50 in our experiments. This approach is able to adaptively refine the transmissions of the background regions from blocky (fig. 2(a)) to detail (fig. 2(c)). Every frame makes contributions to the transmission refinement, and therefore, the refined transmission maps can also tolerate continuous background environmental changes. However, adaptive PCG may also generate some rare defects at foreground regions as shown in figure 2(d). To overcome these defects, we propose a novel Foreground Decremental Preconditioned Conjugate Gradient (FDPCG). In addition to the Jacobi preconditioner, a foreground decremental preconditioning matrix $\mathbf{F_D}$ is introduced to reduce the influence of foreground pixels in PCG.

$$\mathbf{F_D}^{-1}\mathbf{M}^{-1}(\mathbf{L} + \lambda\mathbf{U})\mathbf{t}_{\text{bg}} = \mathbf{F_D}^{-1}\mathbf{M}^{-1}(\lambda\tilde{\mathbf{t}}). \tag{12}$$

$\mathbf{F_D}$ is also chosen to be a diagonal matrix, and is constructed based on the difference between two consecutive frames. Since $(\mathbf{L} + \lambda\mathbf{U})$ is a sparse matrix, the i-th diagonal element, $f_{i,i}^D$, of $\mathbf{F_D}$ will only affect the transmission results of the pixels in the neighborhood, $\mathbf{N}(i)$, of pixel i. Based on (8), $\mathbf{N}(i)$ is chosen to be a 5×5 windows centered at pixel i. The element of $\mathbf{F_D}$ is then defined as

$$f_{i,j}^D = \begin{cases} \left(\sum_{x \in \mathbf{N}(i)} G(x, \sigma_s) N(d, \sigma_d) \right)^{-1} & \text{if } i = j \\ 0 & i \neq j \end{cases}, \tag{13}$$

where $G_s(x, \sigma_s)$ is a spatial Gaussian function centered at pixel i. $N(d, \sigma_d)$ is a normal distribution function, and $d = d_r + d_g + d_b$ is the per pixel RGB-color difference between previous and current frames at pixel x. In the implementation, both $G(x, \sigma_s)$ and $N(d, \sigma_d)$ are pre-calculated for efficiency, and σ_s and σ_d are set to 1 and $0.1 \times d_{max}$, respectively. This formulation decreases the weights of the neighboring equations when the pixel difference increases. Figure 3(c) visualizes the diagonal element values of ${\mathbf{F_D}}^{-1}$.



(a) original frame (b) frame difference (c) ${\mathbf{F_D}}^{-1}$ (d) ${\mathbf{F_I}}^{-1}$

Fig. 3. (a) Original frame, and the visualizations of (b) frame difference, (c) ${\mathbf{F_D}}^{-1}$ in (12), and (d) ${\mathbf{F_I}}^{-1}$ in (18), respectively.

4 Foreground Transmission Recovery

4.1 Foreground/Background Segmentation

We need foreground/background segmentation algorithm for recovering the foreground transmissions. Applying traditional background modeling on foggy videos may not generate good foreground results, especially those texture-based background modeling. Instead, we apply background modeling on the background-defogged video frames. The foreground regions in the background-defogged frames could vary from nearly fog-free, when background and foreground are almost at the same depth, to extremely dark, when the depth difference between foreground and background is large. Any of these cases enhances the foreground to be more distinctive from the background, and therefore, improves the foreground/background segmentation results.

In this paper, a texture-based PLPM background modeling [13] was used for illustration. The main reason to choose PLPM is that texture-based background modeling is usually more tolerant to outdoor scenes, and PLPM can perform the foreground/background segmentation in a very efficient manner.

4.2 Foreground Transmission Estimation by Fusion

With the help of the foreground map, the transmissions of each foreground region is estimated by 1) temporal transmission prediction, and 2) environmental transmission prediction. These two predictions are then fused together for generating the final estimated foreground region transmissions.

The temporal transmission prediction, $p_t^{\mathbf{R}}$, predicts the transmissions of current foreground region, $\mathbf{R}$, from previous frame. $p_t^{\mathbf{R}}$ is defined as

$$p_t^{\mathbf{R}} = \frac{1}{N} \sum_{x \in \hat{\mathbf{R}}} \hat{t}(x), \tag{14}$$

where $\hat{\mathbf{R}}$ is the corresponding region of $\mathbf{R}$ in previous frame, N is the number of pixel in $\hat{\mathbf{R}}$, and $\hat{t}(x)$ is the resulting transmission value at pixel x of previous frame.

The environmental transmission prediction, $p_e^{\mathbf{R}}$, predicts the transmissions of foreground region, $\mathbf{R}$, from the current background transmission map, $\mathbf{t}_{\text{bg}}$ (12). Base on the observation that the foreground regions usually have larger transmission values than background (since foreground objects are usually closer to the camera) , $p_e^{\mathbf{R}}$ is defined as

$$p_e^{\mathbf{R}} = \mu_{t_{\text{bg}}} + \omega \sigma_{t_{\text{bg}}}, \tag{15}$$

where $\mu_{t_{\text{bg}}}$ and $\sigma_{t_{\text{bg}}}$ are the mean and standard deviation of transmission values of $\mathbf{t}_{\text{bg}}$, respectively. ω is a configurable parameter which was set to 1.5 in the experiments.

The final estimated foreground transmission value, $t_{\text{fg}}(x)$, at foreground pixel x is then fused as

$$t_{\text{fg}}(x) = \beta p_t^{\mathbf{R}} + (1 - \beta) p_e^{\mathbf{R}}, \tag{16}$$

where $\beta = e^{-\frac{1}{2}(\frac{d}{\sigma})^2}$. d is the per pixel RGB color difference between current and previous frames, and σ here is a control parameter which is set to $0.05 \times d_{max}$ in our experiments. When $d \to 0$, temporal transmission prediction, $p_t^{\mathbf{R}}$, is preferred. Otherwise, environmental transmission prediction , $p_e^{\mathbf{R}}$, is preferred.

4.3 Foreground Transmission Refinement

The resultant transmission map, $\mathbf{t}_r$, is then constructed by combining foreground and background transmission maps,

$$t_r(x) = \begin{cases} t_{\text{fg}}(x) & \text{if pixel } x \text{ is on foreground.} \\ t_{\text{bg}}(x) & \text{otherwise} \end{cases} \tag{17}$$

We further refine the transmission map by foreground incremental preconditioned conjugate gradient (FIPCG). FIPCG is similar to FDPCG but targeting to increase the foreground effect during PCG,

$$\mathbf{F_I}^{-1}\mathbf{M}^{-1}(\mathbf{L} + \lambda\mathbf{U})\mathbf{t} = \mathbf{F_I}^{-1}\mathbf{M}^{-1}(\lambda\tilde{\mathbf{t}}), \tag{18}$$

in which, the final transmission map, $\mathbf{t}$, is initialized to $\mathbf{t}_r$ in the PCG (alg. 1). $\mathbf{M}$ and $\tilde{\mathbf{t}}$ are Jacobi preconditioner [12] and the dark channel transmission map (6), respectively. Similar to $\mathbf{F_D}$ (13), $\mathbf{F_I}$ is also chosen to be a diagonal matrix, and its elements are defined as

$$f^I_{i,j} = \begin{cases} \left(\sum_{x \in \mathbf{N}(i)} G(x, \sigma_s)(\delta(x) + \frac{1}{N(d,\sigma_d)+\varepsilon}) \right)^{-1} & \text{if } i = j, \\ 0 & i \neq j \end{cases}, \tag{19}$$

where $G(x, \sigma_s)$ is the same spatial Guassisn function in (13). $\delta(x)$ is a delta function, $\delta(x) = 1$ when pixel x is on foreground region, and $\varepsilon = 0.01$ for regularization. $\mathbf{F_I}$ increases the importance of pixels with large frame difference and/or pixels at foreground regions in PCG. Figure 3(d) visualizes the pixel map of $\mathbf{F_I}$.

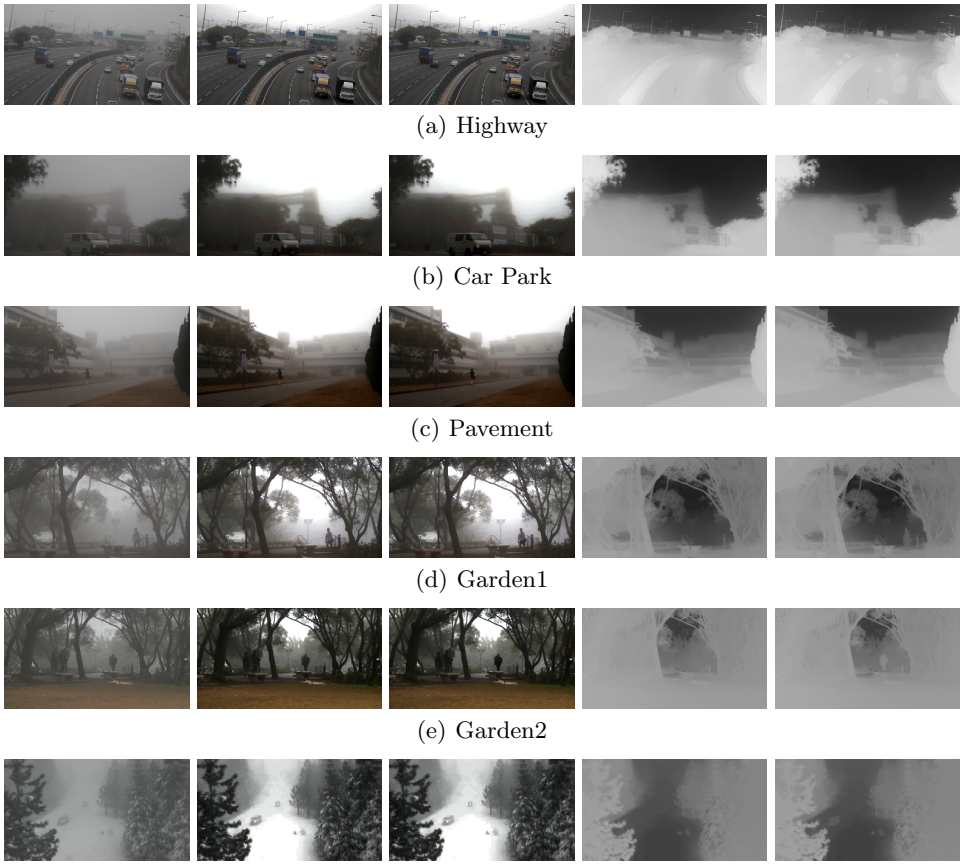
5 Experimental Results

Six challenging real-life surveillance video sequences [1] (see fig. 4) were used to evaluate the proposed methods. Five of the videos are foggy scenes, including highway, car park, and garden scenarios. The remaining one is a heavily snowing scene. The experiments were performed on a computer with an Intel Core 2 CPU 6300 @ 1.86GHz. In our implementation, without any optimization, the proposed background defogging requires about 1.5 to 2 seconds for a 320 × 180 frame, and an additional 1 second for PLPM background modeling and foreground defogging. The detailed time measurement is listed in table 1. Comparing to the state-of-the-art dark channel prior defogging modules [6, 9], which requires about 25 to 30 seconds to completely defog a frame, the proposed method is much more efficient, and has a high potential to be optimized on GPU in order to fulfill the real-time requirement.

	He *et. al.* [6, 9]		Proposed Method			
	avg. PCG iterations	avg. time(sec)	avg. FDPCG iterations	avg. FIPCG iterations	1st frame time (sec)	avg. time(sec)
highway	791	28.24	17	18	3.11	2.77
car park	805	28.73	16	19	3.03	2.71
pavement	779	27.87	16	20	3.03	2.72
garden1	710	25.50	16	9	3.05	2.27
garden2	773	27.70	19	24	3.05	3.15
snowing	809	35.35	32	30	3.64	4.44

Table 1. The number of PCG converging iterations and processing time per frame.

[1] The testing video sequences are available for download at http://www.cs.hku.hk/~scyuk/downloads.htm.



(a) Highway

(b) Car Park

(c) Pavement

(d) Garden1

(e) Garden2

(f) Snowing

Fig. 4. Testing video sequences: (a) highway, (b) car park, (c) pavement, (d) garden1, (e) garden2, and (f) snowing. The 1st column shows the original frames. The 2nd and 3rd columns are the background and final defogged frames, respectively, and the 4th and 5th columns show the FDPCG background transmission maps and final transmission maps, respectively.

	highway	car park	pavement	garden1	garden2	snowing	avg.
orig. foggy	0.82	0.90	0.05	0.61	0.71	0.38	0.58
final defogged	0.82	0.92	0.83	0.79	0.87	0.75	0.83
bg. defogged	0.81	0.93	0.82	0.83	0.86	0.75	0.83

Table 2. *F-Score* foreground/background segmentation measurement of PLPM [13] running on original foggy frames, final defogged frames and background defogged frames, respectively.

Figure 4 shows the defogging results by the proposed method. The proposed FDPCG is shown to be able to effectively remove the foreground effects when calculating the background transmission maps. As discussed in the earlier sections, when the foreground objects are nearly at the same depth as the background such as the highway and snowing sequences, the resultant quality of background-defogged video is already very good. On the other hand, if the depths of the foreground objects are largely different from the background, the foreground objects on the background-defogged frame could be extremely dark. In both cases or the cases in between, the foreground regions became more distinguishable from the background, and this improves the foreground/background segmentation results.

Figure 5 shows the foreground/background segmentation results, and table 2 lists the $F\text{-}Score$ [13] of the results. The $F\text{-}Score$ is defined as $\frac{2TP}{2TP+FP+FN}$, where TP, FP and FN are true positive, false positive and false negative, respectively. Results show that the PLPM background modeling [13] running on the background-defogged videos is almost the same, or even better than (car park and garden1) running on the completely defogged frame. In most of the scenarios, the completely defogged videos as well as the background-defogged videos got better PLPM results then the original foggy videos.

Figure 6 shows the comparisons between the proposed method and the state-of-the-art dark channel prior [6, 9]. The results of [6, 9] are supposed to be the best results that the proposed method can achieve as the results of [6, 9] (fig. 6 (b)) converged completely for each frame. As shown in figure 6 (a), the proposed method performed almost the same as He *et al*. [6, 9] in garden1 sequence. For the car park sequence, the proposed method can only recover the transmissions on background regions in details, but is not able to recover the transmission details on the foreground objects. This is because the proposed method did not perform enough iterations for converging the transmissions on foreground regions. The complete converges on the foreground regions could be very time consuming. Such detailed transmissions [6, 9] on foreground regions, however, may not be necessary. Instead, it is reasonable to assume that the whole area of each foreground object should be at nearly the same depth from the shooting camera. Therefore, the transmissions within each object was assumed to be nearly the same. Results also show that the defogged results of the proposed method are not degraded much comparing with He *et al*. [6, 9].

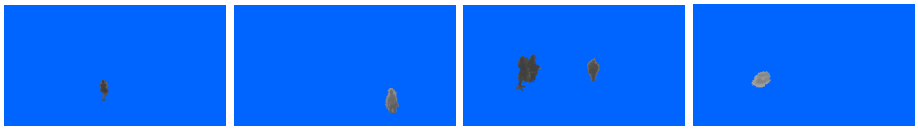
6 Conclusions

This paper proposes a novel Foreground Decremental Preconditioned Conjugate Gradient (FDPCG) for adaptive background defogging of surveillance videos. Each background-defogged frame is then processed by foreground/background segmentation algorithm, and the transmissions on foreground regions are recovered by the proposed fusion technique. Afterward, the final transmissions of each frame are refined by Foreground Incremental Preconditioned Conjugate Gradient (FIPCG). Unlike the previous state-of-the-art algorithms [4–6], which completely defog an image without using any temporal information, the pro-

(a) Ground Truth

(b) PLPM[13] running on original foggy frames

(c) PLPM[13] running on background-defogged frames



(d) PLPM[13] running on final defogged frames

Fig. 5. Foreground/background segmentation results: (a) ground truth, and PLPM[13] running on (b) original foggy frames, (c) background-defogged frames, and (d) final defogged frames.

(a) Defogging results of the proposed method.

(b) Defogging results of He *et al.* [6, 9]

Fig. 6. Defogging results compare with He *et al.* [6, 9]].

posed method defogs the video scenes adaptively. Hence, the proposed method is able to tolerate any background change in the scenes. Experimental results show that the proposed method can produce high quality defogged videos. The foreground/background segmentation results based on the background-defogged frames are also improved dramatically. Comparing to the previous state-of-the-art defogging techniques [4–6], the proposed method is much more efficient, and therefore, retains a high capability to be implemented and optimized on GPU which can fulfill the real-time purpose.

References

1. Narasimhan, S., Nayar, S.: Chromatic framework for vision in bad weather. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Hilton Head, SC, USA (2000) 598–605
2. Nayar, S., Narasimhan, S.: Vision in bad weather. In: Proc. International Conference on Computer Vision (ICCV), Kerkyra, Corfu, Greece (1999) 820–827
3. Shwartz, S., Namer, E., Schechner, Y.: Blind haze separation. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), New York, NY, USA (2006) 1984–1991
4. Tan, R.: Visibility in bad weather from a single image. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Anchorage, Alaska, USA (2008)
5. Fattal, R.: Single image dehazing. In: Proc. SIGGRAPH, Los Angeles, California, USA (2008)
6. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Miami, Florida, USA (2009) 1956–1963
7. Ancuti, C., Ancuti, C., Bekaert, P.: Effective single image dehazing by fusion. In: Proc. IEEE Conf. International Conference on Image Processing (ICIP), Hong Kong, China (2010)
8. Ancuti, C., Ancuti, C., Hermans, C., Bekaert, P.: A fast semi-inverse approach to detect and remove the haze from a single image. In: In Proceedings of the 10th Asian Conferencte on Computer Vision (ACCV), Queenstown, New Zealand (2010)
9. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) **33** (2011) 2341–2353
10. Dong, W., Jia, Z., Shao, J., Li, Z., Liu, F., Zhao, J., Peng, P.Y.: Adaptive object detection and visibility improvement in foggy image. Journal of Multimedia **6** (2011)
11. Levin, A., Lischinski, D., Weiss, Y.: A closed form solution to natural image matting. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), New York, NY, USA (2006) 61–68
12. Smith, I., Wong, S.: Pcg methods in transient fe analysis. part i: First order problems. International Journal for Numerical Methods in Engineering **28** (1989) 1557–1566
13. Yuk, J.S.C., Wong, K.Y.K.: An efficient pattern-less background modeling based on scale invariant local states. In: In Proc. IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS), Klagenfurt, Austria (2011) 285–290