Stabilizing Atmospheric Turbulence in a Long Range Surveillance using Dark Channel Prior

Abhijit R. Mange

Research Scholar

Government Engineering College Aurangabad

ABSTRACT

Various types of turbulences affect the visual quality during video surveillance. Turbulence such as fog or haze is responsible for image contrast. Removing this turbulence is a challenging problem in video surveillance. We used a method for stabilizing this atmospheric turbulence using dual tree complex wavelet fusion. Informative ROIs are selected using frame selection method from good quality frames. Morphological erosion is done afterwards. Image fusion is done using dual tree complex wavelet transform so as to fuse two images. In this paper, dark channel prior mechanism is proposed for haze removal. Finally image quality assessment is done. This proposed method is shown to outperform existing one and provide enhance visual quality in the range of surveillance scenarios.

General Terms

Dual Tree Complex Wavelet Transform (DT-CWT).

Keywords

Mitigation, Turbulence, Segmentation, Restoration.

1. INTRODUCTION

Many valuable data in a video can be degraded due to various types of atmospheric distortion. Distortion includes haze or fog which reduces contrast and turbulence due to temperature variation on aerosol. In summer, the ground becomes hotter than air above it; due to this, air is also heated and horizontal layer of hot air move upwards because of its light weight. This results to faster and greater micro scale change in air's refractive index. Due to this change in refractive index, interference pattern of light changes [1]. Videos captured in public areas are distorted because of atmospheric turbulence which results in blurring, shearing, wavering of images of objects in the video. An example of this effect is found in video captured from hot road or in desserts. This is particularly a problem close to the ground in hot environment in long range surveillance. Due to turbulence present in the air, information behind the distorted layer is difficult to interpret. Hence, there has been done a large research activity for reconstructing the useful information. However the perfect solution is found to be impossible since this problem is irreversible. Matrix vector multiplication expression for this problem is as in Eq. (1).

$$I_{obv} = DI_{idl} + \varepsilon \qquad (1)$$

Here I_{obv} and I_{idl} are vectors that contain the observed and ideal images, respectively. Matrix D stands for geometric distortion and blur, while ε stands for noise [2]. There are different approaches attempting to solve this problem by modeling it as a point spread function (PSF), in which D is considered as a convolution matrix, and then employing deconvolution with an iterative process to estimate I_{idl}. The PSF is generally unknown for the atmospheric distortion case so blind deconvolution is employed [2]. Removal of visible spatio-temporal distortion is not possible with single image. Hence all methods work on a set of images for construction of one enhanced image. Current multi frame method addressing this problem is illustrated in fig. 1.
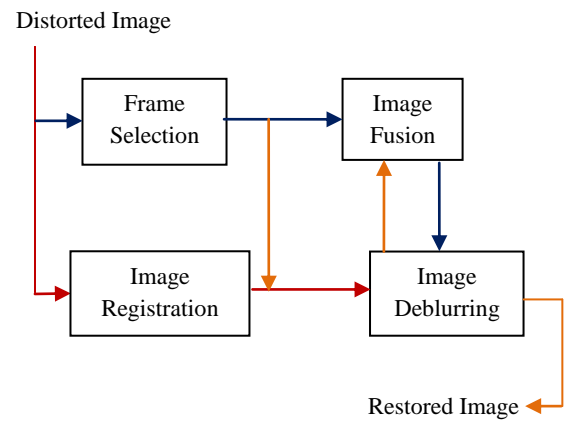

Fig. 1: Block diagram of image restoration for atmospheric turbulence

The other way (shown by blue line) is referred as 'lucky region' technique; this route employs image selection and image fusion. Informative ROIs having good quality frames are selected using image quality metric, which is normally applied in spatial frequency domain for extracting less burred and minimum distorted images. In image fusion block, selected images are then combined. Finally deblurring process is applied [2].

2. LITERATURE SURVEY

Proper mitigation of atmospheric turbulence is a challenging problem in many fields. When an astronomical object is observed from the surface of the earth, emitted or reflected light has to invariably pass through the atmosphere, which yields a blurry observed image. As blue is continually a change in time as well as spatially due to refraction-index fluctuations caused by atmospheric turbulence. Astronomers apply a "Lucky Region" technique that selects a few of the recorded frames that fulfill certain criteria. Averaging of frame is done to obtain a better image [3].

The image atmospheric turbulence mitigation remains a hard problem especially in case of horizontal ground images. Blind deconvolution mechanism is employed for deconvolving the images. Algorithm estimates iteratively the deconvolved image and the kernel of blur. Image comprehension is also get affected by atmospheric distortion. Two combinations of an

elastic image registration algorithm and blind deconvolution are used in concordance with the physical model. These characteristics improve the visual identification performance rates and will be necessary for an automatic target recognition algorithm. Deconvolution is used for removing the effects of blurring. The blur is assumed to be global and known. Richardson-Lucy algorithm is used to deconvolve the image [4].

Yet, atmospheric turbulence mitigation is a challenging problem in video surveillance. Blind deconvolution methods works well in real atmospheric images but fail to remove high atmospheric turbulence. Due to PSF, it also suffers from spatial and temporal variation. Fusion is performed for mitigating atmospheric distortion caused by air turbulence. Image fusion is done at the feature level by a region-based scheme. An image is initially segmented in some way producing a set of regions. For determination of various features from which images are included in the fused image, various properties of these regions are calculated so that more intelligent semantic fusion rules based on actual features in the image, rather than on single or arbitrary groups of pixels can be considered [5]. The fusion is performed in the Dual Tree Complex Wavelet Transform (DT-CWT) domain as it provides near shift invariance and directional selectivity [6]. Additionally, the phase of a CWT coefficient is robust to noise and temporal intensity variations which provide an efficient tool to remove the distorting ripple. Before applying fusion, a set of selected images or ROIs must be aligned. An ROI alignment approach is used for distorted images. As randomly distorted images do not provide identical features, conventional methods cannot be used to find matching features. So, a morphological image processing technique is employed. Subsequently the ROI (or whole image) from only the informative frames measured by a novel quality matrix, based on sharpness, ROI size and intensity similarity is selected. Sigmoid function is used so that one factor should not dominate the other one. Then a non-rigid image registration is applied. We propose a new method for haze removal. Haze removal is done after the fusion by using dark channel prior mechanism [1].

3. PROPOSED MITIGATION SCHEME

The proposed process is shown in Fig. 3. Details of each step are described below.
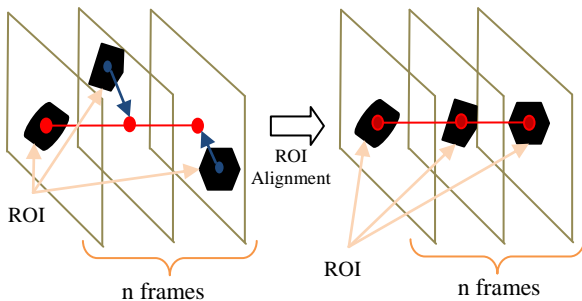
3.1 ROI Alignment

Fig. 2: ROI alignment technique

Shooting video in presence of atmospheric turbulence may cause the ROI in each frame to become misaligned. High variation in the image is observed, due to camera movements. The inter frame distance between the shaking objects of successive image may be so large that image registration cannot be applied directly. These variations significantly

impact on image quality more than the turbulence. For air turbulence problem, matching algorithm which use feature detection is not useful unfortunately. So a simple ROI alignment approach using morphological image processing is used. ROI alignment is the process of matching one image called template with another image. It chooses a subset of the original frames where the inter frame distance is less. The ROI is marked in the first frame. The histogram, generated from the selected ROI and the surrounding area, is utilized to find an Otsu threshold which is used to convert the image to a binary map [7]. The areas connected to the edge of the sub image are removed after applying an erosion process. Same process is repeated till the area near ROI is isolated. The same Otsu threshold with the same number of iterations is employed in other frames also. The centre position of each mask is then calculated. If more than one isolated area is present, the area neighboring in size and position to the ROI in the first frame is used. Finally ROI is aligned by using center of mask across set of frames. Finally, the centre of the mask in each frame is utilized so as to shift the ROI in the set of frames and it is align along a line (Fig. 2). Here, the frames with incorrectly detected ROIs will be detached in the frame selection process. Frames which are selected are significantly different from other frames.
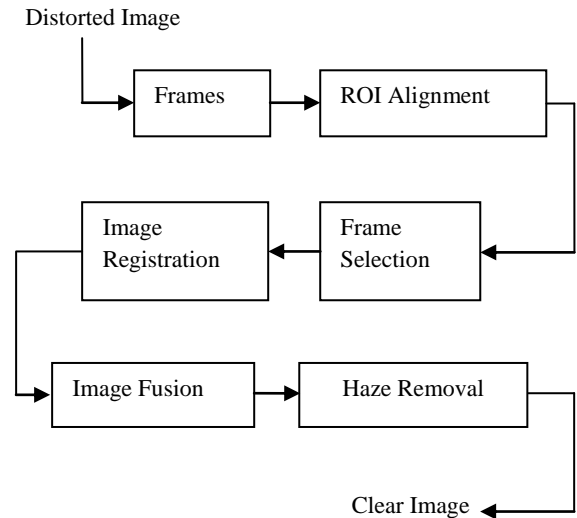

Fig. 3: Block diagram of the proposed method

3.2 Frame Selection

This method do not process all the frames in the sequence for image restoration, only selected frames are employed using frame selection method. The low quality frames (e.g. the very blurred ones) are not taken into consideration as they would possibly deteriorate the fused result. Three factors which are sharpness G_n, intensity similarity S_n and detected ROI size A_n, are used for selection of set of images.

3.2.1 Sharpness

Sharpness is arguably the most important image quality factor because it determines the amount of detail an imaging system can reproduce. Sharpness G_n is used to determine the amount of details which image can convey. Here, the sharpness parameter G_n is calculated from the summation of intensity gradient of the high pass coefficient magnitudes.

3.2.2 Intensity Similarity

Intensity similarity S_n is used to remove outliers. This works under the presumption that most frames in the sequence contain broadly similar areas. There is a high probability that

the frames with significantly different contents than the others are very much distorted. To calculate S_n, the average frame of the whole sequence is used as a reference for computing the mean square error (MSE) for frame n. Then MSE^{-1} represent the similarity of each frame. This approach is not robust to illumination changes.

3.2.3 Detected ROI Size

Larger ROIs are expected to contain more valuable information so it is important to calculate the size of ROI containing more details about image. Detected ROI size A_n gives the total number of pixels in the ROI.

The cost function C_n [2] for frame n is computed using Eq. 2.

$$C_n = \frac{w_G G_n}{\lambda_G + |G_n|} + \frac{w_S S_n}{\lambda_S + |S_n|} + \frac{w_A A_n}{\lambda_A + |A_n|} \tag{2}$$

Where w_k and λ_k are the weight and slope control of the factor $k \in \{G, S, A\}$, respectively [2]. The sigmoid function is used to avoid one factor dominating the others, e.g. a blocking artifact may cause extensively high values of sharpness, yet this frame should not be included in the selected data set. The λ_k is set to equal the mean of factor k. So at the mean value, the cost value of such factor is 0.5 [2]. The cost C_n is ranked from high to low.

3.3 Image Registration

The phase shift properties of the DT-CWT are used for registration of non-rigid bodies [8]. Phase based multidimensional volume registration is robust to noise and temporal intensity variations. Coarser level complex coefficients and finer level coefficients are done iteratively for motion evaluation. Coarser level complex coefficients are used to determine large motion components whereas finer level coefficients are used to process the motion field [1].

3.4 Image Fusion

Image fusion technique is used for combining information from two or more images of the scene, but either captured from different view-points at different times. The fused image is a single image containing all important features from different images thereby being more comprehensive and more appropriate for further human operators or further machine processing works. The region based image fusion technique using complex wavelets is being used for atmospheric turbulence mitigation. This method first reconstructs each image into the DT-CWT domain. The DT-CWT is used in image fusion for combing useful information from a number of source images as it has some special properties such as shift invariance, orientation selectivity and multiscale properties[2] [9].

In region based fusion method, N different images are segmented. Segmentation map S_n of each image is down sampled by 2 so as to give a decimated segmentation map $S_n^{\theta,l}, n \in N$ of level l and sub-band θ of DT-CWT representation, where $\theta \in (1, ..., 6)$. If a list of all region, T_n, of image n is $R_n = \{r_{n,1} r_{n,2}, ..., r_n\}$, a multi-resolution priority map P_n is generated as $P_n = \{p_{n,r_{n_1}}, p_{n,r_{n_2}}, ..., p_{n,r_{n,T_n}}\}$ [2]. For each region in each image n. Based on the priority and fusion rule φ, region are either chosen or discarded and give wavelet coefficients of the fused image. Mask M [2] is generated where $M_t = \varphi(p_{1,t}, p_{2,t}, ..., p_{N,t})$ [2].

The size of wavelet coefficient is same as that of the mask size in fused image. The region with maximum priority to

determine which image each of the coefficients representing region is chosen by algorithm. If $S_i \neq S_j$, a segmented map, S_F, is created such that $S_F = S_1 \cup S_2 \cup ... S_N$. Thus where two region $r_{i,p}$ and $r_{j,q}$ from image i and j overlap, both will get split in two region. Fused image is achieved by performing the inverse transform on the fused wavelet coefficients [2].

In segmentation, to divide the image into R regions, version of the combined morphological spectral unsupervised image segmentation and multiscale watershed segmentation [10] is used. The average of the lowpass values of all images is used for production of the lowpass DT-CWT coefficients of the fused image. Activity measurement representing the significance of that region is categorized according to highpass coefficient. Operating on each band separately is necessary to produce proper sharper result. The priority P of region $r_n^\theta \in R$ in image n is computed with the detail coefficients $d_n^{\theta,l}(x, y)$ of level l and sub-band θ as shown in Eq. 3, where $|r_n^\theta|$ is the size of such region used for normalization [2]. Priority map is employed for consequent construction of fused image.

$$P(r_n^\theta) = \frac{1}{|r_n^\theta|} \sum_{\forall l,(x,y) \in r_n^\theta}^{n} |d_n^{\theta,l}(x, y)| \tag{3}$$

In air turbulence scenario, segmentation boundaries which separate inhomogeneous region vary significantly frame to frame. This air turbulence scenario is very much different from other image fusion problems. The ultimate aim is to get the sharpest and most temporally consistent boundaries. The maximum of DT-CWT coefficient magnitudes from all frames are used at the boundary of each region instead of using $P(r_n^\theta)$. The phase of the complex wavelet coefficients corresponds to the type of dominant directional features in its support regions. Therefore, it plays a vital role for reducing the distortion due to rippling. Hence, the DT-CWT coefficients $d_n^{\theta,l}(x, y)$, of the fused image are adjusted with a unit vector representing the average phase from all frames, N, used in the fusion process (Eq. 4). Pixels deviate from their actual location in turbulent motion so the average phase can be used with approximately zero mean and with a quasi-periodic motion [1] [2].

$$d^{\theta,l}(x, y) = \frac{\sum_n^N d_n^{\theta,l}(x,y)}{\left|\sum_n^N d_n^{\theta,l}(x,y)\right|} |d_n^{\theta,l}(x, y)| \tag{4}$$

Shrinkage function is applied to reduce noise. This function derived as Maximum A Posteriori (MAP) estimators. To enhance high pass coefficient magnitudes, a gain Ag > 1 can be applied only when sharpening is required. The disadvantage of this technique is that it might cause the noise to become clearly visible again. Due to this reason it is suggested to only enhance the high pass coefficient when they form regions of large magnitude. Now, the binary map $M^{\theta,l}$ for each sub band is created. The $M^{\theta,l} = 1$ if ⌐, where ⌐ is a predefined threshold. The isolated pixels are subsequently detached from $M^{\theta,l}$. Eq. 5 depicts the modified high pass coeffients [1] [2].

$$d^{\theta,l} = A_g^{\theta,l} M^{\theta,l} A_s^{\theta,l} d^{\theta,l} \tag{5}$$
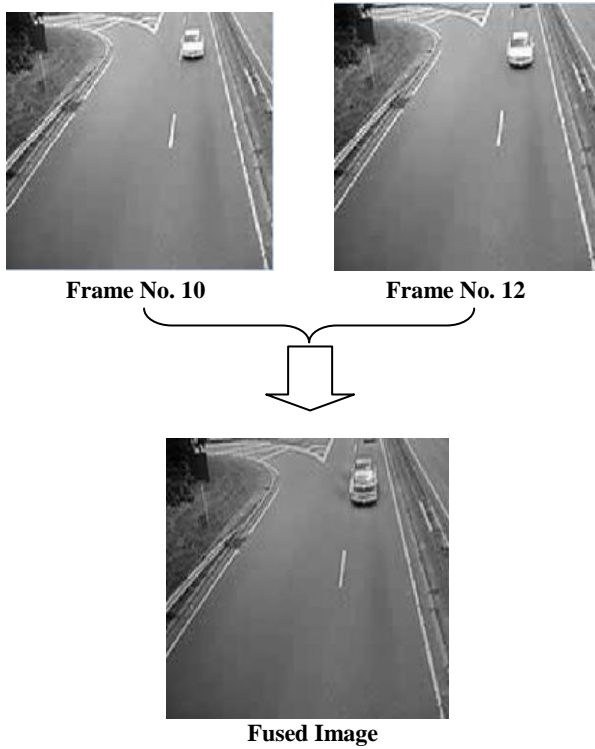
Fig. 4: Image fusion process

Fig. 4 shows the fused image obtained from two images. Here frame no. 10 and frame no. 12 are fused to get the fused image.

3.5 Haze Removal

A problem that often accompanies turbulence is haze or fog. Dark channel prior mechanism is employed for haze removal. Dark channnel prior is based on the statics of outdoor haze free images, we find most of the local region of the haze free image that contain some type of pixcels that are called as dark pixcels whose intensity becomes very low in atleast any one of the color channel. Using this prior, we can directly approximate the thickness of the haze and recover a high quality haze-free image [11]. In haze free images the intensity of these dark pixcels in that channel is mainly contributted by the attmospheric light, therefore these dark pixcels provide an accurate estimation of haze transmition.

There are three factors which are responsible for low intensities in the dark channel. These factors are a) shadows, b) colorful objects or surfaces, c) dark objects or surfaces. The dark channel prior is based on the following observation on haze-free outdoor images: at least one color channel has very low intensity at some pixels, in most of the non-sky patches. It means the low intensity in such a patch should have a very minimum value. J, a haze free image is defined in Eq. 6.

$$J^{dark}(x) = \min_{c \in \{r,g,b\}} \left(\min_{y \in \Omega(x)} (J^c(y)) \right) \tag{6}$$

Where $\Omega(x)$ is a local patch centered at x and J^c is a color channel of J. Except for the sky region, the intensity of J^{dark} is minimum and tends to be zero, if J is a haze-free outdoor image. J^{dark} is the dark channel of J, and the above statistical observation or knowledge is the dark channel prior[8].

4. RESULT

Imaging systems may introduce distortion or artifacts in the signal, so quality assessment is an important issue when assessing the performance of individual systems of differentiation between different solutions to the some methods. Two approaches to quality assessment exist; subjective and objective. Subjective assessment captures the visual perception of human subjects directly through human trials and provides a benchmark for objective analysis. Objective assessments are mathematical models that approximate the results of subjective quality assessment, but are based on criteria and metrics that can be measured readily and automatically evaluated without the need for human trials.

In the past, a great deal of efforts had been made for developing a new objective image/video quality metrics that incorporate perceptual quality measures by considering Human Visual System (HVS) characteristics. Most of the proposed image quality assessment approaches require the original image as a reference. When reference is not available, the quality assessment becomes challenging.

When the reference images or video (undistorted images) are available, the simplest and most widely used full-reference quality metric is mean squared error (MSE) and peak signal-to-noise ratio (PSNR). MSE is calculated by averaging the squared intensity differences of distorted and reference image pixels. The higher values of MSE imply the lower image quality and the lower values of MSE correspond to better image quality.

A related quantity metric is peak signal-to-noise ratio PSNR which is defined in Eq. 7.

$$PSNR = 10 \log\left[\frac{P^2}{MSE}\right] \tag{7}$$

Where P denotes the maximum possible value for any pixel. For an example 8 bit monochrome images, P = 255. The PSNR is measured in units of decibels (dB). Higher value of PSNR implies to better quality of the image and lower the value of PSNR correspond to lower quality image as value of PSNR value is directly proportional to image quality. The distortion-based metrics are straightforward to implement; however, they are mathematical difference models rather than perceptually-based metrics.

Table 1: Comparison between existing and proposed system

Video	Method	MSE	PSNR
Road Traffic	Existing	252.21	24.15
	Proposed	0.0009902	78.2076
Tunnel Road Accident	Existing	374.93	22.43
	Proposed	0.00096942	78.29997
Room Video	Existing	168.57	25.90
	Proposed	0.0009711	78.2922

Performance of the existing and proposed system is evaluated in table 1. Three different videos, road traffic, tunnel road accident and room video are used for testing purpose. MSE and PSNR values of these videos are calculated for existing and proposed system and are compared. Proposed system significantly enhance the PSNR value and the MSE value get decremented, hence proposed system have enhanced the visual quality in a long range video surveillance.
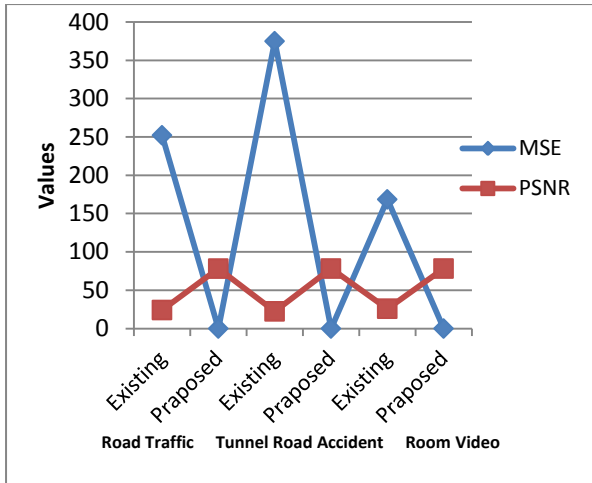
Fig. 5: Graph comparing proposed and existing system

Fig. 5 gives the graphical view of table 1. Here MSE and PSNR values of existing and proposed system for three different videos are shown.

5. CONCLUSION

This paper has introduced a novel method for mitigating atmospheric distortion in long range surveillance imaging system. We have used a simple ROI alignment method and a cost function for frame selection of the distorted sequence. The images are fused in series of steps using DT-CWT. We propose dark channel prior technique for haze removal. The haze removal technique that we have proposed shows improved performance when compared to the existing techniques used. This technique helps us to significantly enhance the visual quality in a long range video surveillance.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Abhijit R. Mange, Chitra M. Gaikwad, "Stabilizing Atmospheric Turbulence in a Video Surveillance using Dual Tree Complex Wavelet Fusion" International Journal of Computer Applications (0975 - 8887), Volume 88 – No. 5, February 2014.

[2] Nantheera Anantrasirichai, Member, IEEE, Alin Achim, Senior Member, IEEE, Nick G. Kingsbury, Fellow, IEEE, and David R. Bull, Fellow, IEEE, "Atmospheric Turbulence Mitigation Using Complex Wavelet-Based Fusion" IEEE Transactions On Image Processing, Vol. 22, No. 6, June 2013.

[3] S. Harmeling, M. Hirsch, S. Sra, and B. Scholkopf, "Online blind image deconvolution for astronomy," in Proc. IEEE Conf. Comp. Photogr., Apr. 2009, pp. 1–7.

[4] Jerome Gilles, Tristan Dagobert, Carlo De Franchis, "Atmospheric turbulence restoration by diffeomorphic image registration and blind deconvolution".

[5] J. J. Lewis, R. J. O'Callaghan, S. G. Nikolov, D. R. Bull, C. N. Canagarajah, and Essa Basaeed, "Region-based image fusion using complex wavelets," in Proc. 7th Inter. Conf. Info. Fusion, 2004, pp. 555–562.

[6] I.W. Selesnick, R.G. Baraniuk, and N.G. Kingsbury, "The dual-tree complex wavelet transform," Signal Processing Magazine, IEEE, vol. 22, no. 6, pp. 123 – 151, nov. 2000.

[7] N. Otsu, "A threshold selection method from gray-level histograms," Systems, Man and Cybernetics, IEEE Transactions on, vol. 9, no. 1, pp. 62 –66, jan. 1979.

[8] H. Chen and N. Kingsbury, "Efficient registration of nonrigid 3-d bodies," Image Processing, IEEE Transactions on, vol. 21, no. 1, pp. 262–272, jan. 2012.

[9] T. Wan, N. Canagarajah, and A. Achim, "Segmentation driven image fusion based on alpha-stable modeling of wavelet coefficients," IEEE Trans. Multimedia, vol. 11, no. 4, pp. 624–633, Jun. 2009.

[10] R.J. O'Callaghan and D.R. Bull, "Combined morphological-spectral unsupervised image segmentation," Image Processing, IEEE Transactions on, vol. 14, no. 1, pp. 49 –62, jan. 2005.

[11] Kaiming He ; Dept. of Inf. Eng., Chinese Univ. Of Hong Kong, Hong Kong, China ; Jian Sun ; Xiaoou Tang, "Single image haze removal using dark channel prior", Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on 20-25 June 2009.