Complex Wavelet Based Video Defogging

Abstract
Restoring a scene distorted by atmospheric turbulence is a challenging problem in video surveillance. The effect, caused by random, spatially varying, perturbations, makes a model-based solution difficult and in most cases, impractical. In this paper, we propose a novel method for mitigating the effects of atmospheric distortion on observed images, particularly airborne turbulence which can severely degrade a region of interest (ROI). In order to extract accurate detail about objects behind the distorting layer, a simple and efficient frame selection method is proposed to select informative ROIs only from good quality frames. The ROI in each frame are then registered to further reduce offsets and distortions. We solve the space-varying distortion problem using region-level fusion based on the dual tree complex wavelet transform. Finally, contrast enhancement is applied. We further propose a learning-based metric specifically for image quality assessment in the presence of atmospheric distortion. This is capable of estimating quality in both full and no-reference scenarios. The proposed method is shown to significantly outperform existing methods, providing enhanced situational awareness in a range of surveillance scenarios.

1. Introduction
Various types of atmospheric distortion can influence the visual quality of video signals during acquisition. Typical distortions include fog or haze which reduce contrast, and atmospheric turbulence due to temperature variations or aerosols. In situations when the ground is hotter than the air above it, the air is heated and begins to form horizontal layers. When the temperature difference between the ground and the air increases, the thickness of each layer decreases and the air layers move upwards rapidly, leading to faster and greater micro-scale changes in the air’s refractive index. This effect is observed as a change in the interference pattern of the light refraction. In strong turbulence, not only scintillation, which produces small-scale intensity fluctuations in the scene and blurring effects are present in the video imagery, but also a shearing effect occurs and is perceived as different parts of
objects moving in different directions. Examples of this effect are found at locations such as hot roads and deserts, as well as in the proximity of hot man-made objects such as aircraft jet exhausts. This is particularly a problem close to the ground in hot environments and can combine with other detrimental effects in long range surveillance applications, where images can be acquired over distances up to 20 km. Turbulence effects in the acquired imagery make it extremely difficult to interpret information behind the distorted layer. Hence, there has been significant research activity attempting to faithfully reconstruct this useful information using various methods. In practice, the perfect solution is however impossible, since the problem is ill-posed, despite being simply expressed with a matrix–vector multiplication as in (1).

\[ I_{\text{obs}} = D I_{\text{idl}} + \epsilon. \]  

(1)

It is obvious that removal of the visible spatio-temporal distortions is not possible with a single image. Hence all methods utilise a set of images to construct one enhanced Image. Current multi-frame methods that address this problem are illustrated in Fig. 1, where most approaches employ all functions or a subset of them. The restoration process can be described by two main routes through the diagram. The first (green dashed line) employs an image.

Effective mitigation of atmospheric turbulence is thus an important yet challenging problem. Model-based solutions are impractical and blind deconvolution methods suffer from spatial and temporal variation due to PSF. Furthermore, conventional registration methods are ineffective for large distortion and are also time-consuming. Finally conventional fusion methods require a large number of frames in order to select lucky regions. In this paper we introduce a new approach that overcomes these problems. Fusion is performed in the Dual Tree Complex Wavelet Transform (DT-CWT) domain since this provides near shift-invariance and good directional selectivity. We also propose novel frame selection and ROI alignment methods for pre-processing region of interest (ROI) since this will frequently exhibit significant offsets and distortions between frames. Contrast enhancement is then used as the final step. Our proposed algorithm is tested with real distorted sequences as well as with simulated sequences. The latter case includes heat distortion generated from gas burners and hence ground truth information is available. We also investigate a quality metric that is suitable for measuring restored image quality for atmospherically distorted content where generically the ground truth is not available. Existing no-reference assessment (NR) methods are tested with our simulated sequences. The results however do not show high correlation with the objective results. Therefore we introduce a new NR measure based on machine learning. The remaining part of this paper is organised as follows. The proposed scheme for mitigating atmospheric distortion is described in detail in Section II. A test methodology for objective assessment is introduced in Section III. The performance
of the method is evaluated on a set of images and is compared with other techniques in Section IV. Finally, Section V presents the conclusions of the paper.

2. Conventional Methodology

Different with the Fourier transformation, the wavelet formation has the analytical ability of time domain and frequency domain. At present, it has become a powerful tool signal processing. Carrying on the multilevel wavelet composition to the two-dimensional function \( f(x, y)_L(R) \).

![Wavelet Decomposition](image)

Figure 2: Wavelet Decomposition

Whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to Fig.2.

![LUM filters operation as the (a) LUM smoother, (b)LUM sharpener, and (c) hybrid smoothing and Sharpening filter](image)

Figure 3. LUM filters operation as the (a) LUM smoother, (b)LUM sharpener, and (c) hybrid smoothing and Sharpening filter

3. Proposed Concept

We propose a new fusion method for reducing the effects of atmospheric turbulence as depicted First, before applying fusion, a subset of selected images or ROIs must be aligned. Here we introduce a new alignment approach for images. As randomly distorted images do not provide identical features, we cannot use conventional methods to find matching features. Instead, we apply a morphological image processing technique, namely erosion, to the ROI (or whole mage) based only on the most informative frames. These are selected using a quality metric based on sharpness, intensity molarity and ROI size. Then, non-rigid image registration is applied.

3.1 ROI Alignment

Capturing video in the presence of atmospheric turbulence, especially when using high magnification lenses, may cause the ROI in each frame to become misaligned. The displacement between the distorted objects in the successive frames may be too large for conventional image registration, using non-rigid deformation, to cope with. Equally, matching using feature detection is
not suitable since strong gradients within each frame are randomly distorted spatially. Hence, an approach using morphological image processing is proposed.

![Figure 4. Average frame of Number Plate sequence by using (a) 200 original frames, (b) 50 frames most similar (lowest MSE) to (a), and (c) 200 frame after applying ROI alignment approach.](image)

The ROI (or ROIs) is manually marked in the first frame. Then the histogram, generated from the selected ROI and the surrounding area, is employed to find an Otsu threshold, which is used to convert the image to a binary map. An erosion process is then applied and the areas connected to the edge of the sub-image are removed. This step is performed iteratively until the area near the ROI is isolated. The same Otsu threshold with the same number of iterations is employed in other frames. The centre position of each mask is then computed. If there is more than one isolated area, the area closest in size and location to the ROI in the first frame is used. Finally, the centre of the mask in each frame is utilized to shift the ROI and align it across the set of frames.

### 3.2 Frame Selection

In CLEAR, not all frames in the sequence are used to restore the image since the low quality frames (e.g. the very blurred ones) would possibly degrade the fused result. A subset of images are carefully selected using three factors: sharpness, intensity similarity and detected ROI size. Sharpness is one of the most important image quality factors since it determines the amount of detail an image can convey. Here, the sharpness parameter $G$

1) **Sharpness**: $G_n$ is computed from the summation of the high pass coefficient magnitudes. Intensity gradients can also be used as the result is insignificantly different from high pass coefficients. is employed to remove outliers. This operates under the assumption that most frames in the sequence contain fairly similar areas. Frames with significantly different content to others are likely to be greatly distorted. To compute $S$

2) **Intensity Similarity**: $S_{nn}$, the average frame of the whole sequence is used as a reference for calculating the mean square error (MSE) for frame $n$. Then MSE-1 represents the similarity of each frame. It should be noted that this approach is not robust to illumination changes. $A_n$ is the total number of pixels contained in the ROI. This is used because, from observation, larger ROIs are likely to contain more useful information.

3) **Detected ROI Size**: The cost function $C_{nn}$ for frame $n$ is computed using $n$ Note that the frames with incorrectly detected ROIs will be removed in the frame selection process (Section II-B). These frames are generally significantly different from others.

$$C_n = \frac{\omega_G G_n}{\lambda_G + |G_n|} + \frac{\omega_S S_n}{\lambda_S + |S_n|} + \frac{\omega_A A_n}{\lambda_A + |A_n|}$$ (2)
3.3 Image Registration
Registration of non-rigid bodies using the phase-shift properties of the DT-CWT, as proposed in [25], is employed. This algorithm is based on phase-based multidimensional volume registration, which is robust to noise and temporal intensity variations. Motion estimation is performed iteratively, firstly by using coarser level complex coefficients to determine large motion components and then by employing finer level coefficients to refine the motion field. It shows an improvement in temporal direction (z) of the Number Plate sequence after applying the proposed ROI alignment and image registration.

3.4 Image Fusion
Due to its shift invariance, orientation selectivity and multi-scale properties, the DT-CWT is widely used in image fusion where useful information from a number of source images are selected and combined into a new image. We employ a region-based scheme in the DT-CWT domain.

![Diagram of Region-based image fusion process in CLEAR.](image)

Figure 5: Region-based image fusion process in CLEAR.

![Images of fusion results.](image)

Figure 6. (a) Region-based fusion result without the mask. (b) Mask (c) Enhanced result with the mask. θ, l

The low pass DT-CWT coefficients of the fused image are simply constructed from the average of the low pass values of all registered images, while the high pass (detail) coefficients are selected according to the priority P indicating the importance of each region. Here we employ the average magnitude of high pass coefficients in each region, since wavelet coefficients having large absolute values contain information on the salient features of an image such as lines and texture. To produce sharper results compared to the results in , we operate on each sub-band separately. The priority P of region r, n in image n is computed with the high pass coefficients is the size of such area used for normalization. The fusion rule, f, selects the region with maximum priority to construct the fused image. The air-turbulence scenario differs from other image-fusion problems as the segmentation boundaries which separate inhomogeneous regions vary significantly from frame to frame (due to turbulence distortion). To provide the sharpest and most temporally consistent boundaries for each region, we use the maximum of DT-CWT coefficient magnitudes over all frames instead of selecting...
only one region based on. To each boundary map $B \theta, l$ (constructed from the multi scale watershed segmentation approach for each sub-band $\theta$ at level $l$), the dilation operation with a size of 1 pixels applied. A 2D averaging filter is then applied to $B$ to prevent discontinuity after combining neighbouring areas.

3.5 Post Processing

1) Contrast Enhancement: In many cases, atmospherically degraded images also suffer from poor contrast due to severe haze or fog. In such cases, pre- or post-processing is needed to improve image quality. Numerous techniques have been proposed for haze reduction using single images. Here we employ a simple and fast method using contrast limited adaptive histogram equalization (CLAHE). The method enhances intensity locally, so it is suitable for applications which consider the ROI and its information content.

2) Other Possible Enhancements: Generally the embedded parameter $A g$ in our approach produces sharp results; however, in cases which are out-of-focus or which lack a “lucky region”, post-processing may be required to further sharpen the images. A number of sharpening methods exist. However, if the constituent images are very poor, it is almost impossible to obtain a sharp result. Moreover, it may exhibit halo effect due to over sharpening.

4. Quality Assessment Of Proposed Nr Method For Atmospheric Distortion

The methods described in Section III-A do not work well with atmospheric distortion, since they are usually based on prior knowledge of the distortion characteristics and none are derived from spatially varying distortions. In this paper, we therefore introduce a new blind image quality assessment metric specifically for this scenario. We employ support vector regression (SVR) to model and predict image quality scores using the features listed in Table I. There are three groups of features.

1) Individual Scale: The magnitude of high pass coefficients relate to details and sharpness of the image, while the phase can be linked to edge information. We therefore employ the mean and variance of both values to compute the feature vectors at each scale level. We decompose the image into 3 levels using the DT-CWT.

2) Inter-Scale: Weighted mean and variance at level $l$ are computed using the magnitudes of the next coarse level to calculate a weight.

First, we examine which NR methods (Section III-A) are suitable for the turbulence case. Measurement values are compared with the chosen FR methods (PSNR, MS-SSIM, VSNR and PIM). Then the selected metrics are used to assess the results of our proposed atmospheric turbulence mitigation and to compare with existing methods.

Figure 7. Distorted sequences. (a) Simulated datasets (A1–A8) generated from gas burners. (b) Real datasets without ground truth (B1–B6).
The objective results shown in Table IV support the subjective results. The proposed fusion approach achieves better JP2K, AQI and QSVR scores for all distorted sequences, apart from Number Plate where the AQI value of SVOLA is slightly better (probably insignificant: 0.003 ~ 0.092%) than CLEAR. However, referring to the subjective result of the Number Plate using the proposed approach clearly reveals more readable numbers. It should be noted that this sequence is highly distorted and the number on the plate is impossible to read in any single frame. It should be further noted that the quality values calculated from the whole frame are slightly different from those in, which relate solely to the ROI. The subjective results clearly show that the CLEAR algorithm removes atmospheric distortion more efficiently than the other approaches. Shan’s BD is inefficient for the sequences degraded by spatially-varying blur since the PSF is assumed to be similar for the entire image, while our method processes sets of homogeneous regions separately. Shan’s method also takes four times longer to process than CLEAR mainly due to PSF estimation. SVOLA subdivides an image into overlapped regions and estimates the PSF separately; as a result, it provides better results compared to Shan’s method. However, the computation time is even longer an the results are not as sharp as the proposed method. In addition, prior knowledge of PSF size is required for both previous methods.

3) Real Datasets with Ground Truth: Three sequences, Chimney, Books and Building have been made available with their ground truth by Hirsch et.al. Also, the results from their approach are available. The results are also compared to another atmospheric turbulence removal approach from Zhu The subjective and objective results for these sequences are shown in Figure 9 and Table V, respectively. The PSNR, MS-SSIM, VSNR and PIM values reveal that CLEAR outperforms SVOLA and Zhu’s approach.
Figure 9: Mirage sequence (B6). (a) Magnified number plates of original frame, SVOLA, CLEAR, and undistorted frames captured at closer distance. (b) Original image. (c) Result of SVOLA. (d) Result of CLEAR. Again it should be noted that BD and SVOLA results use CLEAR’s pre-processing.

5 Conclusion

This paper has introduced a new method for mitigating atmospheric distortion in long-range surveillance imaging. Significant improvements in image quality are achieved using region-based fusion in the DT-CWT domain. This is combined with a new alignment method and cost function for frame selection to pre-process the distorted sequence. The process is completed with local contrast enhancement to reduce haze interference. CLEAR offers class-leading performance for off-line extraction of enhanced static imagery and has the potential to achieve high performance for on-line mitigation for full motion video—this is topic of ongoing research. Experiments with real data show superior performance compared with existing methods. Using simulated data, full reference metrics clearly show the superiority of this method. We have also introduced a new metric, QSVR, based on support vector regression for blindly assessing image quality.

6. References