UNIVERSITY OF BRISTOL

Anantrasirichai, P., Achim, A., & Bull, D. (2018). Atmospheric Turbulence Mitigation for Sequences with Moving Objects Using Recursive Image Fusion. In *2018 25th IEEE International Conference on Image Processing (ICIP 2018)* Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/ICIP.2018.8451755

Peer reviewed version

University of Bristol - Explore Bristol Research
General rights

ATMOSPHERIC TURBULENCE MITIGATION FOR SEQUENCES WITH MOVING OBJECTS USING RECURSIVE IMAGE FUSION

N. Anantrasirichai, Alin Achim, David Bull

Visual Information Laboratory, University of Bristol, UK

ABSTRACT

This paper describes a new method for mitigating the effects of atmospheric distortion on observed sequences that include large moving objects. In order to provide accurate detail from objects behind the distorting layer, we solve the space-variant distortion problem using recursive image fusion based on the Dual Tree Complex Wavelet Transform (DT-CWT). The moving objects are detected and tracked using the improved Gaussian mixture models (GMM) and Kalman filtering. New fusion rules are introduced which work on the magnitudes and angles of the DT-CWT coefficients independently to achieve a sharp image and to reduce atmospheric distortion, respectively. The subjective results show that the proposed method achieves better video quality than other existing methods with competitive speed.

Index Terms— image fusion, wavelet, atmospheric turbulence, object tracking, restoration

1. INTRODUCTION

Atmospheric turbulence effects in acquired imagery make it extremely difficult to interpret the information behind the distorted layer typically formed by temperature variations or aerosols. This occurs when an object, e.g. the ground or the air itself, is hotter than the surrounding air. In such cases, the air is heated and begins to form horizontal layers. Increasing the temperature difference leads 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 and causes the contents of the images and videos to appear shifted from their actual positions. The main problem is that these movements are random, spatially and temporally varying perturbations, making a model-based solution difficult, particularly for sequences with large moving objects.

High-speed cameras can be used with a short exposure time to freeze moving objects so as to minimise distortions associated with motion blur. However, a geometric distortion, which is the result of anisoplanatic tip/tilt wavefront errors, will still be present. The atmospheric turbulence can be viewed as being quasi-periodic; therefore, averaging a number of frames yields a geometric improvement in the image, but it remains blurred by an unknown point spread function (PSF) of the same size as the pixel motions due to the turbulence. Experiments in [1] revealed that assuming a Gaussian blur kernel for non blind deconvolution or a Bayesian blind deconvolution cannot efficiently remove turbulent distortions and showed insignificantly different subjective results. Other techniques exploit a subset of the data by selecting the best quality in the temporal direction. However, it is almost impossible to discard the regions that include moving objects, whilst still maintaining smooth motion in videos. Most methods detect long-distance target objects that are at sufficiently low fidelity to exhibit little or no detail, instead appearing as blurred silhouettes or blobs [2, 3]. A few methods concern large moving objects but they only detect them and do not correct distortion [4–7].

Our previous work, 'CLEAR' (Complex waveLEt fusion for Atmospheric tuRbulence), employs an image fusion technique in the wavelet domain, which has proved its capability in a variety of turbulent atmospheric environments [8, 9]. The method exploits existing information that is already possessed in the sequence. Moreover, denoising, sharpening and contrast enhancement, if necessary, can be performed simultaneously in the wavelet domain [10].

In this paper, we further develop the wavelet-based fusion method for distorted sequences that contain moving objects (CLEAR2). We apply motion-based tracking via a Kalman filter and model the background with Gaussian mixture models (GMM). These deal effectively with the uncertainty inherent in noisy data. The measurement of the object location is already integrated with a non-rigid registration process [11], employed to shift turbulence displacement. However, sometimes objects move faster than the ability of non-rigid registration. Hence, we also provide an object warping process for motion compensation. The sequence is restored in a recursive manner, widely used to minimise buffer size requirements, computational complexity [7] and the propagation of uncertainty [12].

The remainder of this paper is organised as follows. Section 2 describes related work and Section 3 presents the proposed method. The performance of the method is evaluated in Section 4. Finally, Section 5 presents the conclusions.
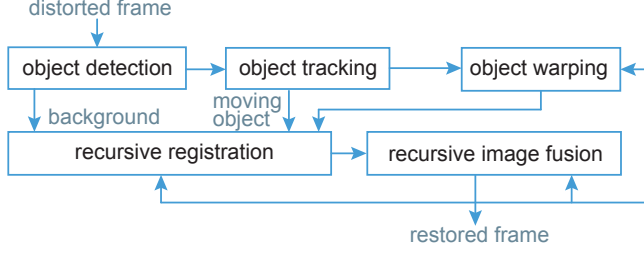
Fig. 1. Diagram of the proposed method.

2. RELATED WORK

The existing methods generally employ an image registration technique with deformation estimation [9, 13]. This process attempts to align objects temporally to solve for small movements of the camera and temporal variations due to atmospheric refraction. Then, these registered images are averaged or image fusion is employed in order to combine several aligned images [14]. A deblurring process is finally applied to this combined image [13]. Most methods in the literature however have been proposed for static scenes. Reducing atmospheric turbulence effects in a video requires an additional motion model when the objects in the scene are themselves moving.

To detect moving objects in the turbulent atmospheric medium, most methods reconstruct the static background first and employ thresholding techniques using motion vectors and/or intensity [4, 5]. Halder et. al. proposed an iterative approach to remove turbulent motion and the moving object is masked using simple thresholds on both motion and intensity [6]. Unfortunately, the method did not remove the distortion around the moving object. A low-rank matrix approach decomposing the distorted image into background, turbulence, and moving object is presented in [2]. An adaptive threshold technique applied to the background model using a temporal median filter in [3].

Three methods in the literature were introduced to mitigate turbulent distortion for large moving objects. The moving objects are detected using block matching techniques in [15, 16]. These are employed to separate the two types of motion with an assumption that the object movement is larger than the turbulent flow. The compensated moving areas are aligned in the 3D volume and the turbulent distortion on these areas is suppressed in the same way as the static background areas. In [16], the true motion is also estimated by smoothing the motion trajectories to remove small random movement caused by turbulence across a fixed number of successive frames. The authors in [17] developed '*dynamic local averaging*', which determines the number of frames to employ for averaging, to avoid any unwanted effects. This method however may not mitigate the distortion on the moving objects as mostly only one frame is employed.

3. PROPOSED SCHEME

The proposed method is depicted in Fig. 1 and the functionality of each block is explained below.

3.1. Object detection and tracking

For a new frame X_t at time t, the process starts with foreground (FG) and background (BG) separation. A background subtraction technique based on a Gaussian mixture model (GMM) is employed [18] . We improve the model by including probability density functions (pdf) of the motion estimated in the non-rigid registration process (Section 3.2). The weight, mean and variance of each distribution is updated in recursive manner. The BG mask M_t^B represents the region where the summation of distributions is larger than a threshold. Assuming the area of the BG is always larger than that of the FG, we set the threshold using the median value of all distributions. To track an object k from a total of K moving objects, the motion of the centroid of each FG mask $M_t^{F,k}$ is estimated using a Kalman filter. Briefly, Kalman filtering employs Bayesian inference and a joint probability distribution over the measured variables for each frame to estimate the locations of the observed objects. In this paper, the objects are assumed to be moving with constant velocity. For nonlinear systems, the extended Kalman filter (EKE) or unscented Kalman filter (UKF) can be employed [19].

3.2. Motion estimation through non-rigid registration

Registration of non-rigid bodies using the phase-shift properties of Dual-Tree Complex Wavelet Transform (DT-CWT) coefficients is employed, similar to our previous work [9]. Motion estimation is performed by firstly using coarser level complex coefficients to determine large motion components and then by employing finer level coefficients to refine the motion field. It should be noted that no more complexity is added to the framework for motion estimation used for object detection in Section 3.1.

3.3. Object warping

We know translation parameters for each moving object from Section 3.1. However, sometimes their movements are not simply pure translation and hence the motion compensation is not good enough to produce reasonable results using non-rigid registration. Therefore, we also introduce a warping process using multi-scale gradient matching [20]. This function is activated when the error from Section 3.2 exceeds a threshold. The 2×2 affine matrix $A_{t,t-1}$ and the 2×1 translation vector $T_{t,t-1}$ linked between two consecutive frames t and $t-1$ are computed using the highpass coefficients of the DT-CWT extracted within the $M_t^{F,k}$ and $M_{t-1}^{F,k}$, respectively. The moving object area O_t^k is constructed in recursive manner as shown in Eq. 1, where α is learning rate.

$$O_t^k = M_t^{F,k}\left((1-\alpha)(A_{t,t-1}^k O_{t-1}^k + T_{t,t-1}^k) + \alpha X_t\right) \quad (1)$$

3.4. Recursive registration

A recursive strategy is proposed for updating the reference R_t at time t. Subsequently, the current input frame X_t is non-rigidly registered to R_t, which happens only once per frame thereby significantly reducing the workload. The reference frame can simply be updated by adding a new frame in and subtracting the oldest frame out of the summation, but this system can develop error build-up over long time periods. Therefore, we create R_t with exponentially decaying weights as shown in Eq. 2, where α is the same parameter in Eq. 1. N_b and N_f are the numbers of previous frames to restore the BG and FG, respectively. Generally, $N_b > N_f$. We set $\alpha = 1/(N_b + 1)$, which is approximately equivalent to averaging the last $N_b + 1$ frames. The current frame X_t is registered to R_t using the method in Section 3.2.

$$R_t = M_t^B[(1-\alpha)R_{t-1} + \alpha X_t] + (1 - M_t^B)\sum_{k}^{K} O_t^k \quad (2)$$

3.5. Recursive image fusion

We denote $\{a^\theta, d^{\theta,l}\}_L = \Psi(X, L)$, where DT-CWT $\Psi(\bullet)$ decomposes an image X into lowpass subbands a^θ and highpass subbands $d^{\theta,l}$, $\theta \in \{\pm 15°, \pm 45°, \pm 75°\}$, $l \in \{1, ..., L\}$, and L is the total decomposition level. At time t, the distorted frame X_t is registered to R_t, resulting X_t^R. Then, we have $\{v_{R_t}^\theta, w_{R_t}^{\theta,l}\}_L = \Psi(R_t, L)$ and $\{v_t^\theta, w_t^{\theta,l}\}_L = \Psi(X_t^R, L)$. The mask m^l is the resized version of $(M_t^B M_{t-1}^B > 0)$ with the same size of $w^{\theta,l}$.

In the recursive image fusion, a_t^θ is constructed as described in Eq. 3, and $d_t^{\theta,l}$ is constructed following Eq. 4 and Eq. 5. The angle of $d_t^{\theta,l}$ is also merged using α to give exponentially decaying weight to those of previous frames. However, applying this idea to the absolute value of the coefficients will not be able to produce a sharp fused frame, since the high frequencies present in previous frames are diminished. Therefore we propose using a weight β with binary mask $Q_t^{\theta,l}$, which is set to 1 if the current wavelet magnitude is smaller than the median value of all coefficients in the same subband and level (med(z) is the median value of data z in Eq. 5). This ensures that strong structures, e.g. corners and lines, are sharp and the accumulated high frequencies presented in the homogeneous areas are suppressed to prevent undesired artefacts. Finally, the restored frame is produced as $Y_t = \Psi_L^{-1}(\{a_t^\theta, d_t^{\theta,l}\})$, where $\Psi_L^{-1}(\bullet)$ is an inverse DT-CWT.

$$a_t^\theta = m^L[(1-\alpha)a_{t-1}^\theta + \alpha v_t^\theta] + (1 - m^L)v_{R_t}^\theta \quad (3)$$

$$\angle d_t^{\theta,l} = m^l \frac{(1-\alpha)d_{t-1}^{\theta,l} + \alpha w_t^{\theta,l}}{|(1-\alpha)d_{t-1}^{\theta,l} + \alpha w_t^{\theta,l}|} + (1 - m^l)\angle w_{R_t}^{\theta,l} \quad (4)$$

$$|d_t^{\theta,l}| = \max\left(\beta m^l |d_{t-1}^{\theta,l}|, |w_t^{\theta,l}|\right)$$

$$\beta = 1 - \alpha Q_t^{\theta,l}, \quad Q_t^{\theta,l} = \frac{|w_t^{\theta,l}|}{\text{med}(|w_t^{\theta,l}|)} < 1 \quad (5)$$

4. RESULTS AND DISCUSSION

The proposed method was evaluated with seven sequences, namely i) Car in Dubai ii) Two people at 1.5km, iii) People with tools, iv) Van driving in circles at 0.75km, v) Dodge in heat wave, vi) Train in strong heat, vii) Plane in airport. The first sequence was captured by our team (VI-Lab, University of Bristol). Sequences 2-4 were provided by DSTL, and the rest were acquired from YouTube. All these sequences are available on the VI-Lab website[1]. If not stated, the results were restored with $L = 4$, $N_b = 50$ and $N_f = 5$.

Firstly we examined the performance of the proposed recursive technique (CLEAR2) compared to transitional temporal sliding window used in CLEAR. The subjective results are shown in Fig. 2. Both methods produced videos that are clearly more stable than the original. The difference between the two registered results was hardly noticeable, whilst the processing time was reduced by 20-fold in this test.

Fig. 3 shows the effects of different N_b. The result in the middle row of this figure was generated using smaller N_b than that in the bottom row. The cropped xt plane at $y = 180$ of these results are presented in the right column. It is clear that the larger N_b offers higher contrast and smoother in temporal direction which is good for reconstructing static background. However, the unsuccessful warped object areas from the earlier frames may be accumulated and present as, for example, unclear edges in the area between the left man and the pole in the picture on the bottom row.

We compared our method with two existing methods: i) Embedded vision system (EVS) [7] and ii) Dynamic local averaging (DLA) adapted following [17]. Fig. 4 shows the results of '*Van driving in circles at 0.75km*' sequence and Table 1 show the average computational time of all seven test seqeunces. All methods were implemented in Matlab, CPU i7-3770S, 16GB RAM. The quality of the restored video from CLEAR2 is better, particularly when compared with EVS – obviously simple intensity and colour thresholding technique does not work for large moving objects. Our result is the sharpest and the ripple distortion is mitigated most. Some areas in the result of DLA still contain atmospheric distortion, whilst some areas appear to be over-sharpened. The performance of CLEAR2 in terms of computational time is significantly improved compared to the previous CLEAR as shown in Table 1. However, CLEAR2 is slower than EVS and DLA because of the non-rigid registration process. We tested CLEAR2 with non-rigid registration operating only on the coarse level (level L of DT-CWT), which reduced the computational time by half, whilst preserving the majority of the turbulence mitigation. Note that implementation in Field-programmable gate array (FPGA) or GPU should speed up the process, e.g. 15 times in [7] and 45 times in [17], respectively.

---

[1]eis.bris.ac.uk/∼eexna/download.html

Fig. 2. Results of '*People with tools*' sequence (Left), using a registration method with a sliding window of $N_b = 50$ (Middle) and a recursive technique (Right). In each group, the left image shows the frame at $t = 150$ and the right image shows the yt planes at column $x = 190$ of this frame in the first 200 frames.
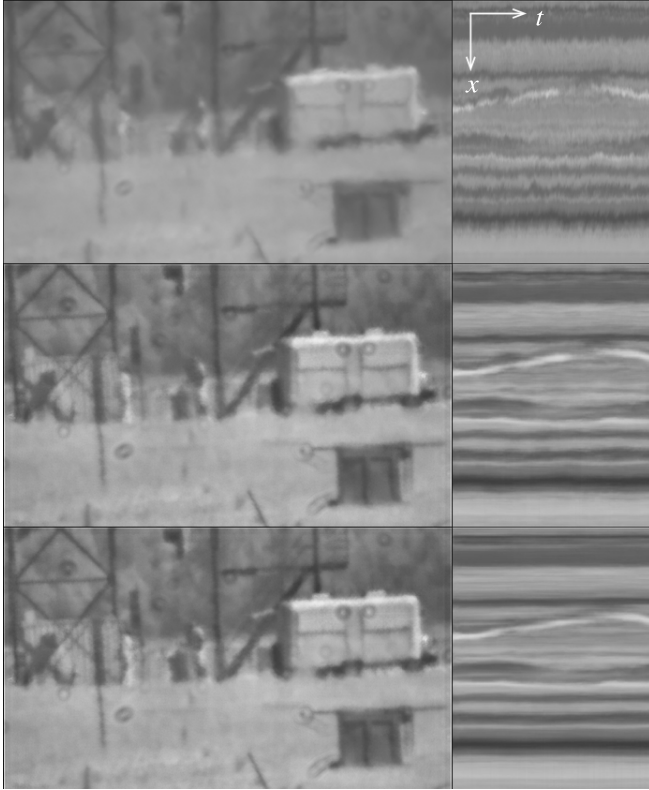

Fig. 3. Results of '*Two people at 1.5km*' sequence (top) enhanced with (middle) $N_b = 20$ and (bottom) $N_b = 50$. Left column: frame $t = 100$. Right column: the xt plane at $y = 180, x \in [1, 440], t \in [730, 100]$.

Fig. 4. Frame $t = 19$ of '*Van driving in circles at 0.75km*' sequence. (top-left) distorted frame. (top-right) CLEAR2 (coarse registration). (bottom-left) DLA adapted from [17]. (bottom-right) EVS [7].

Table 1. Computational performance comparison (sec/frame)

method	resolutions		
	224×320	576×704	720×1280
EVS [7]	0.25	1.59	3.18
DLA adapted from [17]	1.58	6.25	10.67
CLEAR (5 references)	5.32	18.24	25.97
CLEAR (20 references)	16.32	55.39	78.23
CLEAR2 full registration	1.64	8.69	13.05
CLEAR2 coarse registration	0.65	3.87	7.01

5. CONCLUSIONS

This paper has introduced a new method for mitigating atmospheric distortion in long range surveillance imaging. The improvement of visibility of moving objects in observed sequences is achieved using recursive image fusion in the DT-CWT domain. The moving objects are detected and tracked using modified GMM and Kalman filtering. Both background and moving objects are restored by adding the current frame to the previous result with exponentially decaying weight. With recursive registration and fusion, our CLEAR2 technique improves computational performance over the previous CLEAR. We also introduce a coarse-registration option which achieves comparable speed to competing methods with significantly better subjective quality.

6. REFERENCES

[1] Yu Mao, Jérôme Gilles, and Selim Esedoglu, "Non rigid geometric distortions correction - application to atmospheric turbulence stabilization," *Inverse Problems & Imaging*, vol. 6, no. 3, pp. 531–546, 2012.

[2] O. Oreifej, X. Li, and M. Shah, "Simultaneous video stabilization and moving object detection in turbulence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 2, pp. 450–462, Feb 2013.

[3] Eli Chen, Oren Haik, and Yitzhak Yitzhaky, "Detecting and tracking moving objects in long-distance imaging through turbulent medium," *Applied Optics*, vol. 53, pp. 1181–1190, 2014.

[4] S. Gepshtein, A. Shtainman, B. Fishbain, and L. P. Yaroslavsky, "Restoration of atmospheric turbulent video containing real motion using rank filtering and elastic image registration," in *12th European Signal Processing Conference*, Sept 2004, pp. 477–480.

[5] Barak Fishbain, Leonid P. Yaroslavsky, and Ianir Ideses, "Real-time stabilization of long range observation system turbulent video," *Journal of Real-Time Image Processing*, vol. 2, pp. 11–22, 2007.

[6] Kalyan Kumar Halder, Murat Tahtali, and Sreenatha G. Anavatti, "Geometric correction of atmospheric turbulence-degraded video containing moving objects," *Optics Express*, vol. 23, pp. 5091–5101, 2015.

[7] A. Deshmukh, G. Bhosale, S. Medasani, K. Reddy, P. H. Kumar, A. Chandrasekhar, P. K. Kumar, and K. Vijayasagar, "Embedded vision system for atmospheric turbulence mitigation," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, June 2016, pp. 861–869.

[8] N. Anantrasirichai, A. Achim, D. Bull, and N. Kingsbury, "Mitigating the effects of atmospheric distortion using DT-CWT fusion," in *IEEE International Conference on Image Processing*, Sept 2012, pp. 3033–3036.

[9] N. Anantrasirichai, A. Achim, N.G. Kingsbury, and D.R. Bull, "Atmospheric turbulence mitigation using complex wavelet-based fusion," *Image Processing, IEEE Transactions on*, vol. 22, no. 6, pp. 2398–2408, 2013.

[10] Artur Loza, David R. Bull, Paul R. Hill, and Alin M. Achim, "Automatic contrast enhancement of low-light images based on local statistics of wavelet coefficients," *Digital Signal Processing*, vol. 23, no. 6, pp. 1856 – 1866, 2013.

[11] 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.

[12] N. Anantrasirichai, J. Burn, and D. Bull, "Terrain classification from body-mounted cameras during human locomotion," *IEEE Transactions on Cybernetics*, vol. 45, no. 10, pp. 2249–2260, Oct 2015.

[13] Xiang Zhu and Peyman Milanfar, "Removing atmospheric turbulence via space-invariant deconvolution," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 1, pp. 157 –170, Jan. 2013.

[14] R. He, Z. Wang, Y. Fan, and D. Fengg, "Atmospheric turbulence mitigation based on turbulence extraction," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2016, pp. 1442–1446.

[15] Claudia S. Huebner, "Software-based turbulence mitigation of short exposure image data with motion detection and background segmentation," in *SPIE 8178, Optics in Atmospheric Propagation and Adaptive Systems XIV*, 2011.

[16] Alessandro Foi, Vladimir Katkovnik, Pavlo Molchanov, and Enrique Sánchez-monge, "Methods and systems for suppressing atmospheric turbulence in images," 2015.

[17] Eric J. Kelmelis, Stephen T. Kozacik, and Aaron L. Paolini, "Practical considerations for real-time turbulence mitigation in long-range imagery," *Optical Engineering*, vol. 56, pp. 1 – 12, 2017.

[18] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1999, vol. 2, p. 252 Vol. 2.

[19] E. A. Wan and R. Van Der Merwe, "The unscented kalman filter for nonlinear estimation," in *Proceedings of the IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium*, 2000, pp. 153–158.

[20] Hany Farid and Jeffrey B. Woodward, "Video stabilization & enhancement," Tech. Rep. TR2007-605, Dartmouth College, Computer Science, 2007.