1-1-2020

Patch-based gaussian mixture model for scene motion detection in the presence of atmospheric optical turbulence

Richard L. Van Hook
Air Force Research Laboratory

Russell C. Hardie
University of Dayton, rhardie1@udayton.edu

Patch-based Gaussian mixture model for scene motion detection in the presence of atmospheric optical turbulence

Richard L. Van Hook[a,b] and Russell C. Hardie[b]

[a]Air Force Research Laboratory, 2241 Avionics Circle, Wright-Patterson Air Force Base, United States
[b]University of Dayton, 300 College Park Drive, Dayton, United States

ABSTRACT

In long-range imaging regimes, atmospheric turbulence degrades image quality. In addition to blurring, the turbulence causes geometric distortion effects that introduce apparent motion in acquired video. This is problematic for image processing tasks, including image enhancement and restoration (e.g., superresolution) and aided target recognition (e.g., vehicle trackers). To mitigate these warping effects from turbulence, it is necessary to distinguish between actual in-scene motion and apparent motion caused by atmospheric turbulence. Previously, the current authors generated a synthetic video by injecting moving objects into a static scene and then applying a well-validated anisoplanatic atmospheric optical turbulence simulator. With known per-pixel truth of all moving objects, a per-pixel Gaussian mixture model (GMM) was developed as a baseline technique. In this paper, the baseline technique has been modified to improve performance while decreasing computational complexity. Additionally, the technique is extended to patches such that spatial correlations are captured, which results in further performance improvement.

Keywords: motion segmentation, turbulence mitigation, long range imaging, atmospheric turbulence simulator, anisoplanatic imaging, patch-based Gaussian mixture model

1. INTRODUCTION

As sensors and their corresponding imaging systems continue to improve, the effective range of such imaging systems is beginning to be limited by the physical environment of the sensor and its encompassing field of view. Local variations in temperature cause the formation of air pockets with different densities. The density fluctuations create variations in the indices of refraction along optical paths. These air pockets change location, size, shape, and density over time, resulting in atmospheric optical turbulence. Therefore, as light from a point source (for example) propagates through the atmosphere, adjacent air pockets will refract the light slightly differently. The end effects of such turbulence are a combination of blurring as well as spatially- and temporally-correlated geometric distortions.

These effects can cause substantial degradation to images captured by optical sensors, particularly for imaging systems where the field of view of the sensor is much larger than the isoplanatic angle. In order for image processing software to function effectively, these effects must be mitigated. There has been substantial work in the field. Much of this work involves the fusion of many short exposure images. In the case of a static scene, all apparent motion can be treated as turbulence. However, if true scene motion is present, care must be taken to preserve this motion. The focus of this paper is the detection and segmentation of true scene motion in the presence of atmospheric optical turbulence. The true motion mask provided by this work can be used to prevent turbulence mitigation algorithms from inadvertently eliminating or degrading the moving scene content.

A well-explored approach to removing the turbulence-induced geometric distortions is through the use of motion fields, including optical flow,[1–5] patch-based registration,[6–12] and pyramid-based deformation.[13] Bispectrum speckle imaging has been recently applied to the problem of turbulence mitigation,[14–19] though Carrano

---

Further author information: (Send correspondence to Richard L. Van Hook)
Richard L. Van Hook: E-mail: Richard.VanHook@us.af.mil, Telephone: 1 937 713 8589
Russell C. Hardie: E-mail: rhardie1@udayton.edu, Telephone: 1 937 229 3611

et al conducted initial research in the early 2000s.[20] Additionally, background subtraction techniques have been examined in depth.[5, 21–25]

Deshmukh et al[21] employ a Gaussian mixture model (GMM) based on background subtraction within a sliding temporal window to maintain performance compatible with (future) real-time implementations. In our January 2020 paper, the current authors also used a Gaussian mixture model approach, but additionally provided a quantitative performance analysis based upon a novel synthetic data creation process that injects physically-accurate atmospheric turbulence effects.[26] In this paper, we use the same synthetic data set with known ground truth to quantitatively compare two new techniques that are based on the original motion classification algorithm. The first new technique simplifies our January 2020 technique for better detection results with reduced processing time. The second new technique extends the January 2020 technique to patches, further improving detection performance.

2. ALGORITHMS

The technique presented in our original paper[26] has been designated Pixel Temporal Frame-based Technique 1 (PiT-1) and is summarized below. Then, a slightly-modified version of that technique, PiT-2, is presented. In this paper, we also introduce a novel extension of PiT-1 under the designation Local Patch Temporal Frame-based Technique 1 (LPaT-1).

2.1 PiT-1 Algorithm

The PiT-1 classifier operates independently on every pixel in a registered image sequence. That is to say, for an input image of N_1 columns by N_2 rows, the input vector $\mathbf{x}$ refers to the pixel values of a single, arbitrary spatial location throughout the image sequence, as represented by the dark pixels shown in Fig. 1.

An attempt is made to fit a 2-mode GMM to each vector $\mathbf{x}$. Failed attempts are due to low variance within $\mathbf{x}$ (i.e., the values for each pixel location for all frames are very similar). As such, those pixels are assumed to be background and are assigned a score of 0. If the initial 2-mode fit is successful, then additional modes are incrementally added until either the Akaike Information Criterion (AIC) reaches an optimum or a predetermined maximum number of modes is reached.

Because each of the independent $N_1 \times N_2$ GMMs was created based on a video that contains true motion (e.g., moving cars, waving foliage, etc.), it is possible that some GMMs have modes that represent that motion. In order to remove that contamination, each GMM's modes are sorted in descending order by ρ/σ, where ρ is the mode's proportion and σ is the mode's variance. Modes are cumulatively included until a predetermined threshold for proportion is reached, and any modes after that are discarded.

The remaining modes are assumed to model that pixel's background throughout the video sequence, and so a new GMM is directly formed from those modes alone. The probability density function (PDF) of the original input vector $\mathbf{x}$ is calculated based on the background-only GMM, and the negative log likelihood (NLL) of the PDF is taken as the score for the corresponding elements in $\mathbf{x}$.

2.2 PiT-2 Algorithm

A challenge with the PiT-1 technique is determining the threshold when sorting GMM modes into foreground and background modes. For pixels that contain pure background/turbulence throughout all frames, the existence of a threshold nearly always means that one or more modes will be removed. Potentially unique background information is needlessly discarded, resulting in the motion classifier suffer from an increase in false positives. For pixels that contain true motion within their temporal window, that motion is typically sparse and is unlikely to substantially affect the modes within the GMM. Therefore, the PiT-2 technique is identical to PiT-1 up to the point of distinguishing between foreground and background modes. However, it then calculates the negative log likelihood score directly from the contaminated GMM rather than establishing a "background only" GMM.
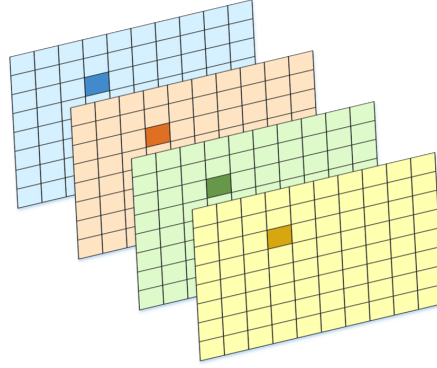
Figure 1. The darker squares represent the same pixel in temporally-adjacent frames of an image sequence, collectively defined as **x**.

2.3 LPaT-1 Algorithm

In many real-world applications, the tip/tilt (i.e., the amount of apparent horizontal and vertical motion) from atmospheric turbulence of even moderate strength exceeds the subtension of a single pixel. Fig. 2 illustrates that tip/tilt displacements of an image subjected to turbulence effects are typically several times the size of a single pixel. As such, any single pixel's intensities in a registered image sequence don't represent the same geophysical location throughout the temporal window. This is easily visualized with an image sequence of a point source propagated through atmospheric turbulence, as shown in Fig. 3.
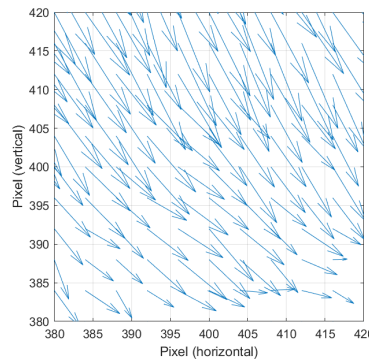

Figure 2. The quiver plot shows typical displacement of pixels from their ideal location to their actual location due to the tip/tilt caused by atmospheric turbulence.

Additionally, even in video where turbulence is negligible, any arbitrary pixel rarely behaves substantially different than all the neighboring pixels. Even in extreme cases where a pixel is on a sharp, high-contrast corner, there are typically one or two adjacent pixels on the inside of the corner whose values mimic the corner pixel.

Both the geometric distortions and the natural tendency for adjacent pixels to behave similarly indicate the need for a motion classifier that does not operate on each pixel in isolation. As such, the PiT-2 technique has been extended to incorporate patches of pixels, and this extension has been designated LPaT-1.

Each patch is a square box centered on pixel coordinates (n_1, n_2) where b is an odd integer representing the length of the sides of the box. Thus, each patch is a (nominally) b pixel $\times$ b pixel square box. Spatial portions of a patch that lie outside the image are excluded. More formally, for an image with N_1 columns and N_2 rows, a patch includes all pixel locations (p_1, p_2) where $1 \leq p_1 \pm b \leq N_1$ and $1 \leq p_2 \pm b \leq N_2$. This is illustrated in Fig. 4.
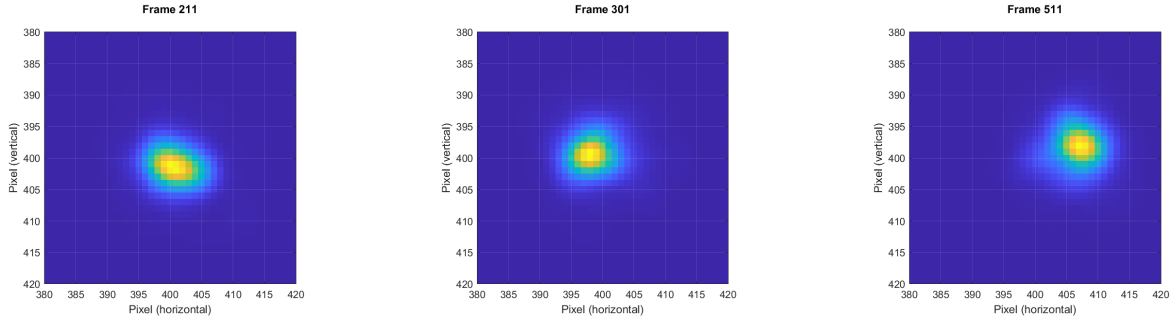
Figure 3. A single point spread function changes both location and shape throughout an image sequence that contains atmospheric turbulence.
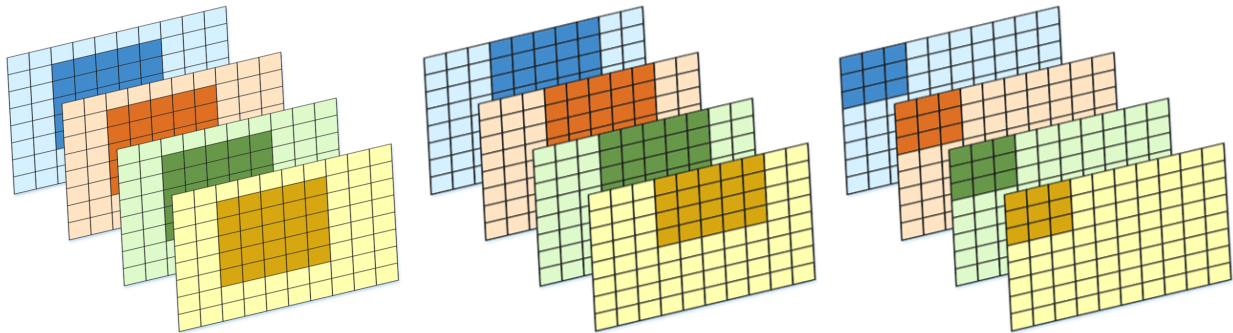

Figure 4. The darker squares represent patches of pixels in adjacent frames of an image sequence, later vectorized as **x**. The patches illustrated here have a box size of 5. From left to right, the illustration shows an interior patch, a semi-exterior edge patch, and a full-exterior corner patch.

The intensity values for all patch locations throughout the image sequences are vectorized as the input **x**. The particular order of the elements within **x** is immaterial, and the GMM creation process is the same as PiT-2. Note that the LPaT-1 technique with a box size of $m = 1$ is identical to the single-pixel PiT-2 technique.

3. DATA

In order to provide a quantitative evaluation of the subject motion detection algorithms, per-pixel truth data is needed. Since it is not feasible to get the necessary truth from real-world data, synthetic data is created. The input to the data creation process is a sequence of raw images that are spatially registered using Speed Up Robust Features (SURF) to establish a similarity transformation between each frame and a reference frame. It is assumed that the scene does not contain turbulence, but moderate turbulence is acceptable provided the blurring is minimal. A rank filter (in this case, a median filter) is applied to the registered frame stack to create a background model. The use of a rank filter removes in-scene motion, both local movers as well as the geometric warping from mild-to-moderate turbulence. It also mitigates some of the blurring effects from turbulence. The left image in Fig. 5 shows a sample frame from the input video sequence, and the right image shows the reference frame created from a stack of 600 input frames.

To create synthetic moving objects that are realistic in both appearance and behavior, motion is extracted from an individual frame and injected into the reference frame. For each image, after conversion to grayscale, the image is subtracted from the reference frame. The resulting frame difference is binarized via thresholding, and then the morphological open operation is applied to fill in holes. Finally, small clusters are assumed to be noise and are removed. For all remaining pixels in the binary map, the color values from the corresponding pixel locations are copied from the original color input frame onto the reference frame. An example of a resulting image is shown in Fig. 6.

Figure 5. The left image is a typical raw image from a video, and the right image is the result of applying a median filter to the 600-frame image registered temporal median. Note the white car on the left is eliminated on the right due to its motion.

Note that even though the example image with injected motion isn't a perfect reconstruction of the original input image, it does capture the majority of the moving pixels. More importantly, the per-pixel truth from this step is the key enabler for quantitative analysis of the motion classification algorithms.

Figure 6. Artificial movers are injected into the median-filtered background image and are circled in red.

The final step in the synthetic data creation process is to degrade each frame with the effects of atmospheric optical turbulence. To accomplish this, the anisoplanatic turbulence simulator from Hardie *et al*[27] is used as it is believed to be the most-validated anisoplanatic atmospheric turbulence simulator currently available.

A grid of sinc-Gaussian point sources is created. Each point source is processed individually via standard split-step propagation through multiple phase screens with Fresnel Angular Spectral Propagation (ASP). Each phase screen accounts for the effects from the turbulence between it and the previous phase screen (or, in the case of the first screen, between it and the point source). The phase screens are modeled with a Modified von Karman (MVK) power spectral density, but also include a sub-harmonic technique to include low frequencies that would otherwise be absent due to the small lateral extent of each phase screen.

After each point is propagated to the pupil plane as a point spread function (PSF), it is cropped based on the aperture size and downsampled to the focal plane pixel spacing. The PSFs are then convolved with the corresponding pixels in the input image to obtain the turbulence-degraded output image, an example of which is shown in Fig. 7.
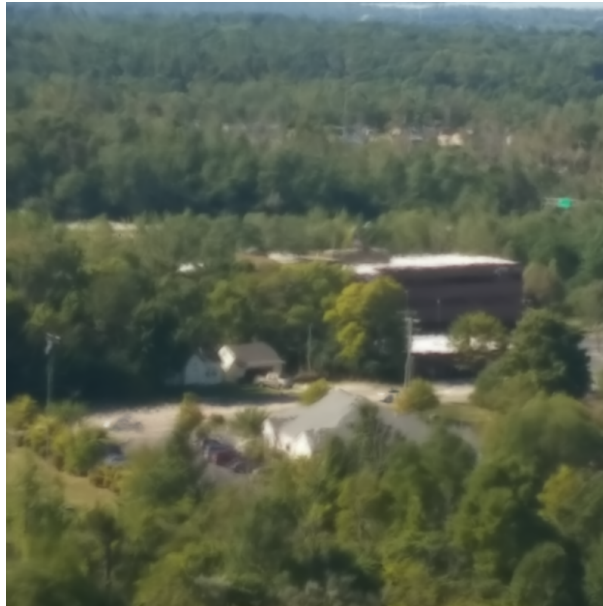

Figure 7. The final synthetic image contains artificial movers and exhibits realistic effects of atmospheric turbulence.

4. RESULTS

4.1 Input Parameters

Tab. 1 lists the parameters used in the turbulence simulation. The strength of turbulence, commonly referred to as the C_n^2 profile, is conspicuously absent from the list. This is because different turbulence profiles and strengths may have the same end effect, statistically speaking. Therefore, it is most appropriate (and more concise) to ensure variety in the end effect (i.e., the tip/tilt of the point spread functions) rather than in the input conditions. Tab. 2 lists the effective turbulence strength (assuming a uniform profile) and the corresponding tip/tilt that such turbulence produces. Note that configuration 0 represents conditions where no turbulence is presence.

4.2 Computer Specifications

All processing is performed on a Hewlett-Packard Zbook 17 G6 laptop with an Intel Core i7-9850H processor running at 2.6 GHz. The laptop contains 64 GB of DDR4 RAM operating at 2400 MHz. It has an integrated Intel UHD Graphics 630 graphics card and an Nvidia Quadro T1000 graphics card, although neither is (currently) used for processing.

Each run is done on a clean boot with all networking and user-controlled background processed turned off. The processing time is measured via Matlab's 'tic' and 'toc' commands, and only includes the time from when the images are first passed to the motion classification function until that function returns the scores.

Table 1. Parameters used for synthetic data generation.

Parameter	Value
Aperture	$D = 0.2034$ m
Focal length	$f = 1.2$ m
F-number	$f/\# = 5.9$
Wavelength	$\lambda = 525$ nm
Object distance	$Z = 7$ km
Number of phase screens	10
Pixel resolution	801×801 pixels
Pixel sampling	$1.55\,\mu$m
Number of frames	600
Framerate	60 frames / second
Wind speed	1 m / second

Table 2. Specifications of each turbulence configuration

Config #	$C_n^2(\mathrm{m}^{-2/3})$	Tip/tilt (m)	Tip/tilt (pixels)
0	0	0	0
1	3.205e-17	7.744e-7	0.5
2	1.282e-16	1.549e-6	1
3	5.128e-16	3.097e-6	2
4	2.051e-15	6.195e-6	4
5	8.204e-15	1.239e-5	8

4.3 PiT-2 Results

Performance is quantified primarily via a receiver operating characteristic (ROC) curve. Fig. 8 shows the performance of PiT-2 under the aforementioned turbulence configurations. As expected, the general trend is for stronger turbulence to result is less performance. However, while the configuration with no turbulence might have been anticipated to fare the best, it exhibits a strong knee close to $P_d = 0.88$ and quickly underperforms even strong turbulence cases. This is likely due to the lack of variance in a pixel's intensity throughout the temporal window, resulting in a failure to fit a GMM even when motion is present.

Tab. 3 lists the area under the curve (AUC) and processing time for each turbulence configuration. The best performance occurs for turbulence configuration 2 (average tip/tilt of 1 pixel) and quickly degrades as turbulence strength is increased. Processing time varied between approximately 60 and 100 minutes for the 600-frame image stack.

4.4 LPaT-1 Results

The patch-based LPaT-1 technique showed similar behaviors to PiT-2 (as shown in Fig. 8), including the sharp knee for the zero-turbulence case and the general trend of decreasing performance with increased turbulence strength. Tab. 4 again lists the processing time and AUC. The best performance occurs for turbulence configuration 1, but configuration 2 has nearly identical performance; this is not surprising as the addition of another half-pixel of tip/tilt is not substantial.

The processing time was noticeably longer than PiT-2 and typically took between 90 and 140 minutes. This increase in time is not surprising as there were more pixels to process for the same number of GMMs. Note
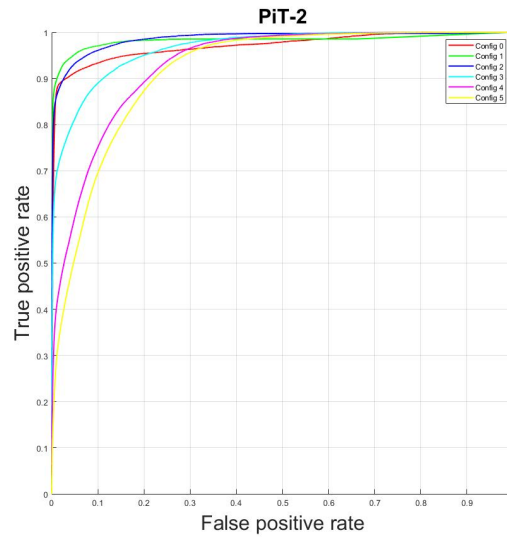
Figure 8. ROC curves for the PiT-2 motion classification technique.

Table 3. AUC and processing time of the PiT-2 motion classification technique.

Config #	AUC	Processing time (sec)
0	0.9705	4041
1	0.9820	3477
2	0.9859	5234
3	0.9651	6399
4	0.9303	6281
5	0.9153	5070

that only a box size of $m = 3$ was evaluated due to the processing time. It is fully expected that the AUC would improve for cases of strong turbulence if patch sizes were increased to account for the increased spatial dispersion.

Table 4. AUC and processing time of the LPaT-1 motion classification technique.

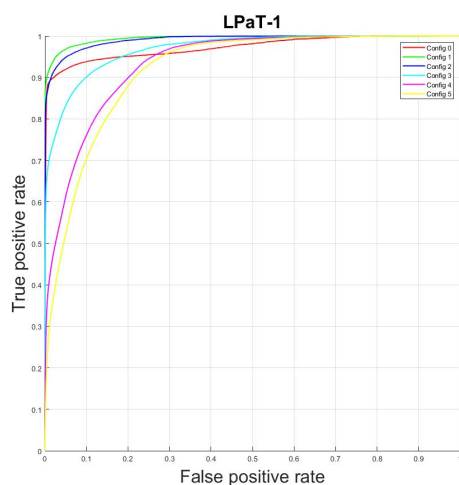| Config #	AUC	Processing time (sec)
0	0.9724	5318
1	0.9939	7192
2	0.9908	8240
3	0.9685	8276
4	0.9331	7849
5	0.9178	6386

Figure 9. ROC curves for the LPaT-1 motion classification technique.

4.5 Comparative Results

Fig. 10 shows the ROC curves for all 3 techniques compared to-date, split into the 6 different turbulence configurations. For configurations 1 through 5, PiT-1 performs the worst, PiT-2 slightly better, and LPaT-1 the best. In the case of configuration 0, the lack of any turbulence degrades motion detection performance across the board. The quantitative analysis of all three motion detection techniques is listed in Tab. 5. The approach of PiT-2 to not sort modes into background and foreground provides a small improvement in speed. More importantly, however, it is one less parameter that must be tuned. Also of note is that while LPaT-1 had the best performance, it took substantially longer (sometimes twice as long) to achieve that result.

Table 5. AUC and processing time of all 3 motion detection techniques

	PiT-1		PiT-2		LPaT-1	
Config #	AUC	Time	AUC	Time	AUC	Time
0	0.9679	4138	0.9705	4041	0.9724	5318
1	0.9810	3532	0.9820	3477	0.9939	7192
2	0.9814	5319	0.9859	5234	0.9908	8240
3	0.9532	6387	0.9651	6399	0.9685	8276
4	0.9134	6382	0.9303	6281	0.9331	7849
5	0.9009	5139	0.9153	5070	0.9178	6386

5. CONCLUSIONS

Atmospheric optical turbulence causes blurring and warping in acquired images sequences. The warping from turbulence creates apparent motion that can be mistaken for true scene motion and vice versa. Many turbulence mitigation algorithms average short exposure frames after some registration process. If that registration process treats true scene motion as turbulence, the scene motion will be degraded and scene content will be lost. Thus,
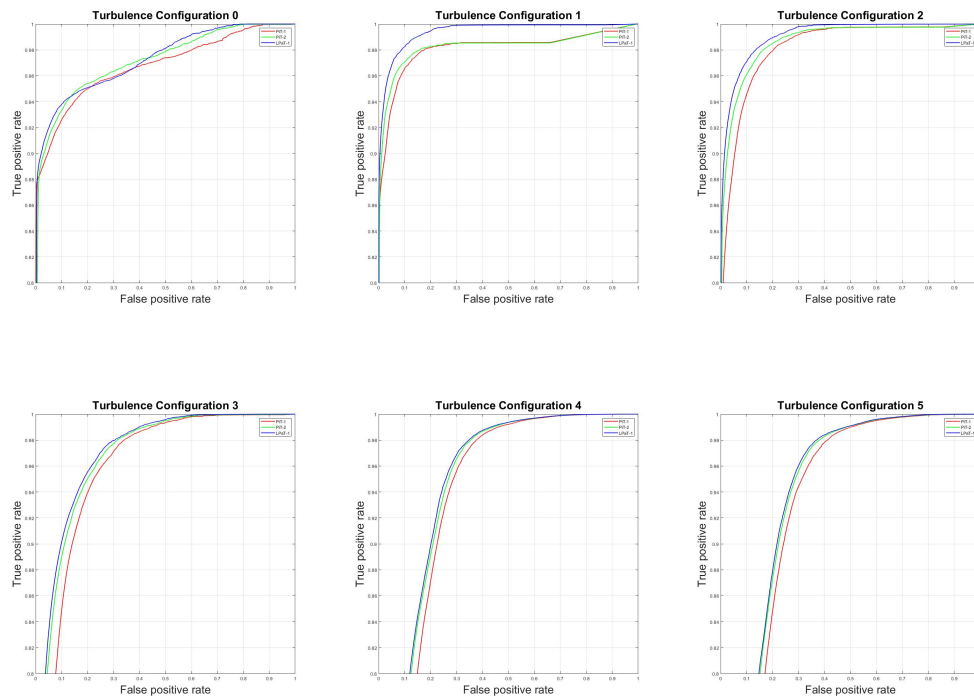
Figure 10. ROC curves of the 3 motion detectors are compared for each of the 6 turbulence configurations.

it is critical to have a robust true scene motion detection capability. We have investigated a number of GMM-based detection methods in this paper. We have extended our previous work that used a single pixel detector to patch based detectors here. The patch detection method is shown to have superior performance as is exploits the spatial correlation that exists with multipixel moving objects as well as spatial correlation in the turbulent warping. In future work, we plan to integrate the detection statistics of the proposed method with turbulence mitigation algorithms.

REFERENCES

[1] Frakes, D. H., Monaco, J. W., and Smith, M. J., "Suppression of atmospheric turbulence in video using an adaptive control grid interpolation approach," in [*2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)*], **3**, 1881–1884, IEEE (2001).

[2] Li, D., Mersereau, R. M., Frakes, D. H., and Smith, M. J., "A new method for suppressing optical turbulence in video," in [*2005 13th European Signal Processing Conference*], 1–4, IEEE (2005).

[3] Li, D., *Restoration of atmospheric turbulence degraded video using kurtosis minimization and motion compensation*, PhD thesis, Georgia Institute of Technology (2006).

[4] He, H. and Kondi, L. P., "A super-resolution technique with motion estimation considering atmospheric turbulence," in [*Passive Millimeter-Wave Imaging Technology VIII*], **5789**, 135–145, International Society for Optics and Photonics (2005).

[5] Robinson, P. E. and Nel, A. L., "Foreground segmentation in atmospheric turbulence degraded video sequences to aid in background stabilization," *Journal of Electronic Imaging* **25**(6), 063010 (2016).

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

[7] Yaroslavsky, L. P., Fishbain, B., Shteinman, A., Gepshtein, S., et al., "Processing and fusion of thermal and video sequences for terrestrial long range observation systems," (2004).

[8] Fishbain, B., Yaroslavsky, L. P., and Ideses, I. A., "Real-time stabilization of long range observation system turbulent video," *Journal of Real-Time Image Processing* **2**(1), 11–22 (2007).

[9] Fishbain, B., Yaroslavsky, L. P., and Ideses, I. A., "Real time turbulent video perfecting by image stabilization and super-resolution," *arXiv preprint arXiv:0704.3447* (2007).

[10] Yaroslavsky, L., Fishbain, B., Shabat, G., and Ideses, I., "Superresolution in turbulent videos: making profit from damage," *Optics Letters* **32**(20), 3038–3040 (2007).

[11] Huebner, C. S., "Software-based turbulence mitigation of short exposure image data with motion detection and background segmentation," in [*Optics in Atmospheric Propagation and Adaptive Systems XIV*], **8178**, 81780K, International Society for Optics and Photonics (2011).

[12] Huebner, C. S., "Turbulence mitigation of short exposure image data using motion detection and background segmentation," in [*Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXIII*], **8355**, 83550I, International Society for Optics and Photonics (2012).

[13] Li, J., Zhang, J., and Sui, Z., "Image restoration from sequences under atmospheric turbulence effects," in [*AOPC 2017: Optical Sensing and Imaging Technology and Applications*], **10462**, 104625G, International Society for Optics and Photonics (2017).

[14] Paolini, A., Tow, D., and Kelmelis, E., "Using atcom to enhance long-range imagery collected by nasa's flight test tracking cameras at armstrong flight research center," in [*Sensors and Systems for Space Applications VII*], **9085**, 908503, International Society for Optics and Photonics (2014).

[15] Paolini, A., Humphrey, J., Curt, P., and Kelmelis, E., "Multi-frame image processing with panning cameras and moving subjects," in [*Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications XI*], **9076**, 907607, International Society for Optics and Photonics (2014).

[16] Paolini, A., Kelmelis, E., Kozacik, S., Bonnett, J., and Fox, P., "Real-time technology for enhancing long-range imagery," in [*Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications XII*], **9460**, 94600C, International Society for Optics and Photonics (2015).

[17] Kozacik, S. T., Paolini, A. L., Sherman, A., Bonnett, J. L., and Kelmelis, E. J., "Comparison of turbulence mitigation algorithms," *Optical Engineering* **56**(7), 071507 (2017).

[18] Kelmelis, E. J., Kozacik, S. T., and Paolini, A. L., "Practical considerations for real-time turbulence mitigation in long-range imagery," *Optical Engineering* **56**(7), 071506 (2017).

[19] Paolini, A., Bonnett, J., Kozacik, S., and Kelmelis, E., "Development of an embedded atmospheric turbulence mitigation engine," in [*Long-Range Imaging II*], **10204**, 102040C, International Society for Optics and Photonics (2017).

[20] Carrano, C. J. and Brase, J. M., "Adapting high-resolution speckle imaging to moving targets and platforms," in [*Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications*], **5409**, 96–106, International Society for Optics and Photonics (2004).

[21] Deshmukh, A. S., Medasani, S. S., and Reddy, G., "Moving object detection from images distorted by atmospheric turbulence," in [*2013 International Conference on Intelligent Systems and Signal Processing (ISSP)*], 122–127, IEEE (2013).

[22] Haik, O. and Yitzhaky, Y., "Effects of image restoration on automatic acquisition of moving objects in thermal video sequences degraded by the atmosphere," *Applied optics* **46**(36), 8562–8572 (2007).

[23] Chen, E., Haik, O., and Yitzhaky, Y., "Detecting and tracking moving objects in long-distance imaging through turbulent medium," *Applied optics* **53**(6), 1181–1190 (2014).

[24] Halder, K. K., Tahtali, M., and Anavatti, S. G., "Turbulence mitigation and moving object detection for underwater imaging," in [*2015 International Conference on Optical Instruments and Technology: Optoelectronic Imaging and Processing Technology*], **9622**, 96220C, International Society for Optics and Photonics (2015).

[25] Elahi, M. T. E. and Halder, K. K., "Detecting moving objects from long-range atmospheric turbulence degraded videos," in [*2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEiCT)*], 260–263, IEEE (2018).

[26] Van Hook, R. and Hardie, R., "Scene motion detection in the presence of atmospheric optical turbulence with performance anslysis using numerical wave simulation data and ground truth," in [*9th International Symposium on Optronics in Defence and Security*], Association Aéronautique et Astronautique de France (2020).

[27] Hardie, R. C., Power, J. D., LeMaster, D. A., Droege, D. R., Gladysz, S., and Bose-Pillai, S., "Simulation of anisoplanatic imaging through optical turbulence using numerical wave propagation with new validation analysis," *Optical Engineering* **56**(7), 071502 (2017).