

# Image Reconstruction of Static and Dynamic Scenes through Anisoplanatic Turbulence

Zhiyuan Mao, *Student Member, IEEE*, Nicholas Chimitt, *Student Member, IEEE*,  
and Stanley H. Chan, *Senior Member, IEEE*

**Abstract**—Ground based long-range passive imaging systems often suffer from degraded image quality due to a turbulent atmosphere. While methods exist for removing such turbulent distortions, many are limited to static sequences which cannot be extended to dynamic scenes. In addition, the physics of the turbulence is often not integrated into the image reconstruction algorithms, making the physics foundations of the methods weak. In this paper, we present a unified method for atmospheric turbulence mitigation in both static and dynamic sequences. We are able to achieve better results compared to existing methods by utilizing (i) a novel space-time non-local averaging method to construct a reliable reference frame, (ii) a geometric consistency and a sharpness metric to generate the lucky frame, (iii) a physics-constrained prior model of the point spread function for blind deconvolution. Experimental results based on synthetic and real long-range turbulence sequences validate the performance of the proposed method.

**Index Terms**—Atmospheric turbulence, reference frame, lucky region, blind deconvolution

## I. INTRODUCTION

Ground-based long-range passive imaging systems often suffer from degraded image quality due to a turbulent atmosphere. The cause of such degradation is complicated, as the atmospheric turbulence is affected by temperature, wind velocity, humidity, air pressure, and many other factors [1]. From the image quality point of view, the influence of turbulence on an image is often demonstrated through the random warping and blurring of the image. If the distortions are spatially and temporally varying, we refer to the turbulence as *anisoplanatic* [2], [3]. Anisoplanatic turbulence is common in ground-to-ground systems where the object distance is long and the field of view is large.

Image processing techniques for recovering images from anisoplanatic turbulence have been studied for decades. However, while the literature is rich, existing methods have several limitations: (i) Many methods are designed for the easier case of isoplanatic turbulence, i.e., the blur is spatially invariance. Methods for anisoplanatic turbulence are considerably fewer, among which most are designed for static scenes only. Moving objects remains very challenging. (ii) There is generally a disconnect between the reconstruction algorithm and the physics of the turbulence. Some existing image processing methods use over-simplified models that are not justified and explainable by physics. (iii) Existing benchmark evaluation sequences are often collected by short-range hot-air burners.

The turbulence in these sequences is usually isoplanatic. The long-range anisoplanatic effects are not captured.

The goal of this paper is to present a robust image reconstruction algorithm for anisoplanatic turbulence. Our method applies to both static and dynamic scenes and is evaluated using a comprehensive set of real and synthetic data. Figure 1 summarizes the three contributions of this paper:

- 1) We propose a new method to construct a reliable reference frame. Our method is based on a non-local space-time averaging concept. Compared to existing approaches that are based on temporal averaging, the new method is able to preserve the moving content while stabilizing the turbulent background. (See Section 3.A.)
- 2) We propose a new method to generate the lucky frame by combining two image quality metrics: A geometric consistency metric and a sharpness metric. Compared to existing lucky fusion methods that generate artifacts, our method offers significantly better object consistency. (See Section 3.B.)
- 3) We propose a new prior for the point spread functions (PSFs) to improve the blind deconvolution step. Our prior is based on a linear expansion model of the PSFs through simulation in the Zernike space. Compared to generic blind deconvolution methods, our method offers better reconstruction quality and robustness. (See Section 3.C.)

Besides presenting the ideas, this paper also provides detailed experimental results using synthetic data and real data. For real data, we do not only compare with the commonly used short-range hot-air turbulence sequences but also long-range sequences. These results will be shown in Section 4.

This paper focuses on incoherent imaging. By incoherence we meant that no active light sources are used to illuminate the scene. The alternative to incoherent imaging is coherent imaging, e.g., using adaptive optics [6] and speckle imaging techniques [7]–[10]. Coherent imaging requires dedicated instrumentation, e.g., telescopes in astronomy applications, which are not always feasible for systems constrained by size, weight, and power.

## II. BACKGROUND

As light propagates through a turbulent medium, the wavefront is distorted and so the images are warped and blurred. In this section, we provide a brief overview of the image degradation process and explain the origins of several mainstream image processing methods.

The authors are with the Department of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, 47906 USA. Email: {mao114, nchimitt, stanchan}@purdue.edu.

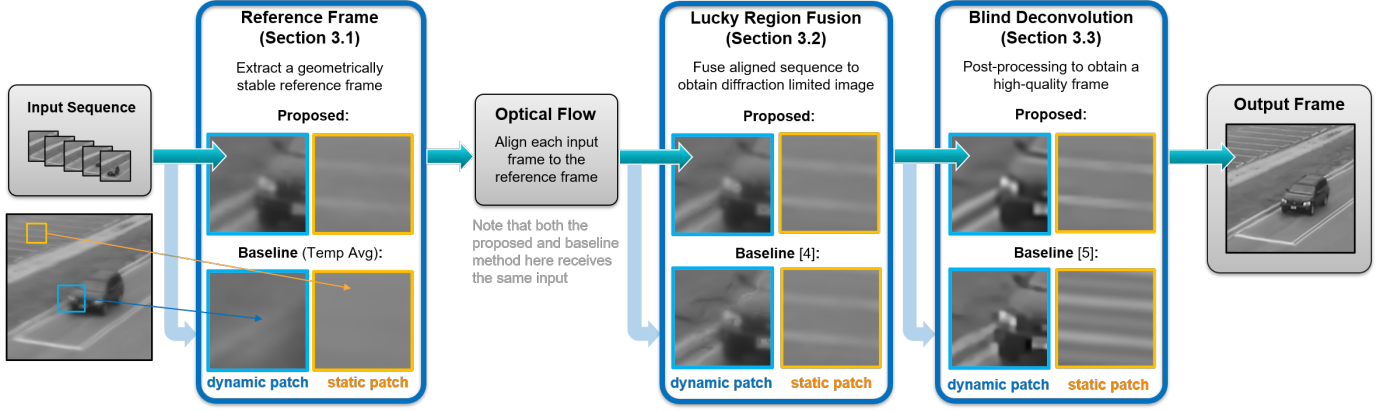


Fig. 1: **Contributions of this paper.** (1) We use space-time non-local averaging to construct a reference frame for optical flow. Our reference frame can maintain the moving objects while suppressing turbulence. The baseline method using temporal averaging tends to wash out the objects. (2) We propose new metrics to extract the lucky regions from the optical warped images. Compared to a baseline method [4], our method is more robust to registration error near the moving object. (3) We propose a new blind deconvolution algorithm based on Zernike decomposition. Compared to the baseline method [5], our method generates fewer artifacts.

### A. The Optics of Imaging through Turbulence

We first discuss the optics of how light propagates through turbulence. Much of the materials in this section can be found in textbooks such as [2], [11]. We highlight a few key concepts in order to explain our algorithm.

**Statistics of the phase.** The optical distortion due to turbulence is caused by the inhomogeneous refractive index of the medium. As hot and cold air moves, the refractive index changes spatially and temporally. When a wave passes through the turbulence, the phase of the wavefront,  $\phi$ , will be distorted, accumulated, and propagated. The physics of such phase distortion is well known in the literature [11, Section 8.3]. It can be shown that the phase is best characterized by the *structure function* [1, Eq. 1.13] which is a variant of the autocorrelation function. By using the Kolmogorov spectrum [12], it follows that the structure function  $\mathcal{D}_\phi$ , assuming wide sense stationarity and spherical symmetry, is [13, Eq. 5.3]

$$\mathcal{D}_\phi(|\mathbf{f} - \mathbf{f}'|) = 6.88(|\mathbf{f} - \mathbf{f}'|/r_0)^{5/3}, \quad (1)$$

where  $\mathbf{f}$  and  $\mathbf{f}'$  are the (2D) spatial frequencies of the phase, and  $r_0$  is called the atmospheric coherence diameter (or the Fried parameter). The Fried parameter measures how strong the turbulence is. It is a function of the refractive index structure parameter  $C_n^2$ , the path length  $L$ , and the wave number  $k^2$  [2, Eq. 3.67].

**Optical transfer function.** The randomness of the phase distortion is translated to the images through the point spread function (PSF) or the optical transfer function (OTF), where OTF is the Fourier transform of PSF. The OTF of turbulence is a product of two components:

$$H(\mathbf{f}) = H_{\text{atm}}^\phi(\mathbf{f}) H_{\text{dif}}(\mathbf{f}), \quad (2)$$

where the first OTF  $H_{\text{atm}}^\phi$  is caused by the random atmosphere, and the second OTF  $H_{\text{dif}}$  is caused by the fixed lens [13, Eq. 3.17].  $H_{\text{atm}}^\phi$  accounts for the random tilts and spatially varying blur, whereas  $H_{\text{dif}}$  accounts for the optics which is spatially

invariant. The PSF of a simulated turbulence is illustrated in Figure 2, where we emphasize the decoupling of the tilt and the blur.

**Long and short exposures.** There are two important observations of  $H_{\text{atm}}^\phi$ . First,  $H_{\text{atm}}^\phi$  is spatially varying because  $\phi$  changes from pixel to pixel. If the field of view is within the isoplanatic angle, then  $\phi$  is not random and so  $H_{\text{atm}}^\phi$  will become spatially invariance [2, Ch. 4].

Second, the exact expression of  $H_{\text{atm}}^\phi(\mathbf{f})$  can be determined using convolutions [2, Eq. 2.44]. Plugging the structure function  $\mathcal{D}_\phi$  into the equation, one can show that the statistical average of  $H_{\text{atm}}^\phi(\mathbf{f})$ , known as the *long-exposure* OTF, is [13, Eq. 3.16]:

$$H_{\text{LE}}(\mathbf{f}) \stackrel{\text{def}}{=} \mathbb{E} [H_{\text{atm}}^\phi(\mathbf{f})] = \exp \left\{ -3.44 \left( \frac{\lambda d |\mathbf{f}|}{r_0} \right)^{5/3} \right\}, \quad (3)$$

where  $d$  is the focal length of the lens,  $\lambda$  is the wavelength, and  $r_0$  is the Fried parameter. Note that  $H_{\text{LE}}(\mathbf{f})$  is spatially *invariant* since it is the average of a random process.

In the turbulence literature, people also consider the statistical average of the OTF when the tilts in the phase distortions are removed. In this case, the tilt-free phase distortion  $\varphi$  will lead to a different OTF, called the *short-exposure* OTF, defined as [13, Eq. 5.9a]

$$\begin{aligned} H_{\text{SE}}(\mathbf{f}) &\stackrel{\text{def}}{=} \mathbb{E} [H_{\text{atm}}^\varphi(\mathbf{f})] \\ &= \exp \left\{ -3.44 \left( \frac{\lambda d |\mathbf{f}|}{r_0} \right)^{5/3} \left[ 1 - \left( \frac{\lambda d |\mathbf{f}|}{D} \right)^{1/3} \right] \right\}, \end{aligned} \quad (4)$$

where  $D$  is the aperture diameter. The difference between the long-exposure and the short-exposure is that the absence of tilts in the short-exposure equation. Therefore, the blur associated with  $H_{\text{SE}}$  is generally much weaker than that of  $H_{\text{LE}}$ . This suggests that if we remove the tilts and then take an average, then the resulting deblurring problem will be easier to solve than the long-exposure case.

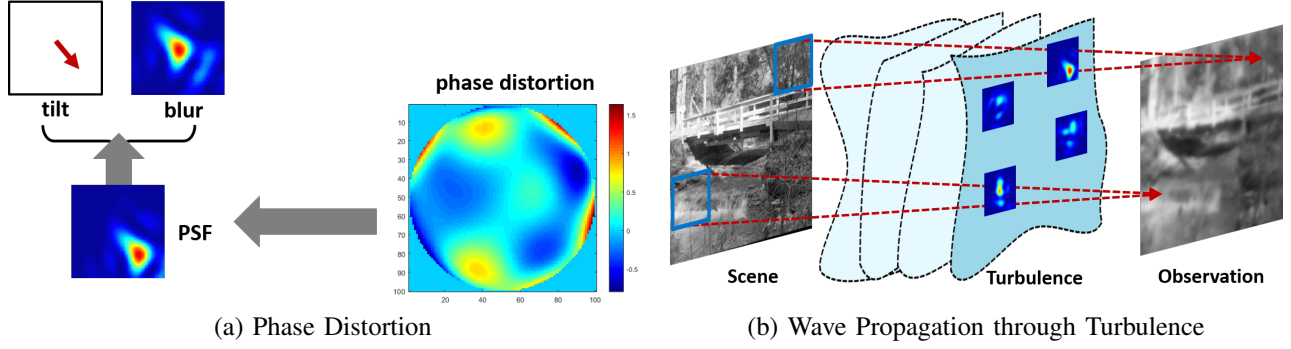


Fig. 2: **Image formation process in a turbulent medium.** (a) The origin of the turbulence distortion is the random phase caused by the changing refractive index. The phase will lead to a point spread function (PSF), which consists of tilt and blur. The tilt shifts the pixels, whereas the blur degrades the image. (b) As light travels from the object plane to the image plane, the PSFs are applied to individual blocks of pixels to generate a spatially-varying distortion.

**Summary.** To summarize, the image formation process in a turbulent medium is originated from the phase. The phase causes tilt and blur, which then leads to a pixel-wise PSF, as shown in Figure 2(a). As light travels from the object plane to the image plane, the spatially varying PSFs will interact with the pixels in the object plane to create the distortion effect, as shown in Figure 2(b). The analysis explains the principle behind many image processing algorithms for turbulence. The idea is to remove the tilt, construct the short exposure expectation by taking an average, and then deblur. Our proposed method is also based on these steps.

### B. Image Reconstruction and Current Limitations

Having discussed the image formation process, we now summarize the typical approaches in existing image reconstruction methods and comment on their limitations.

**Step 1. Removing tilts.** The first step of a reconstruction method is to mimic the short-exposure procedure by removing the tilts. This is often done using the optical flow [14]. However, since all input frames of a turbulence-distorted sequence are randomly warped, we need to first construct a geometrically stable reference frame.

When the scene is static, one can take the temporal average as the reference frame like [4], [15]–[19]. More advanced approaches such as robust principal component analysis (RPCA) can be used to extract the low-rank part as the reference frame. However, RPCA generally has comparable performance to temporal averaging as shown in section IV. B.

Improvement of the reference frame extraction methods exists, e.g., the variational method by Xie et al. [20], and frame selection schemes by Anantrasirichai et al. [17] and Lau et al. [21]. However, these techniques are designed for static scenes. Images containing moving objects still cannot be recovered.

Reference extraction methods for moving scenes are available but scattered. Halder et al. [22] assume that a certain segment of the video sequence does not contain any moving object so that a reference can be extracted. Oreifej et al. [23] and Anantrasirichai et al. [24] use background segmentation to separate the moving object from the static background. However, since turbulence-distorted images are usually blurry,

the object boundaries of the segmentation are difficult to be determined precisely which can be seen from Figure 20 in the experiment section.

Our proposed method does not require any segmentation. It is based on a space-time non-local averaging which can stabilize background while preserving moving objects.

**Step 2. Lucky image fusion.** While optical flow can align the pixels so that we can approximate the short-exposure frames, a simple averaging over all these short-exposure frames is problematic because the optical flow algorithms are not perfect. In addition, it is known in turbulence literature that a sharp region will occasionally appear because the random phase distortion is zero-mean [25]. Therefore, the most reasonable step here is to compare the spatial blocks over an extended period of frames, and pick the sharpest block. This step is known as the lucky image fusion.

The goal of the lucky fusion step is to combine sharp image blocks to form a diffraction limited image. The fusion can be done in the wavelet domain as proposed by Anantrasirichai et al. [17], or in the spatial domain by computing the local gradient, variance or distortions as proposed by Aubailly et al. [15], Zhu and Milanfar [4], and Xie et al. [20]. For robust PCA approaches such as Lau et al. [21] and He et al. [26], the lucky image fusion step is done by refining the sparse components and adding them to the low-rank parts of the image.

In our proposed method, the lucky image fusion is done by a new image quality metric that measures both the sharpness and consistency of the patches. The sharpness metric helps extracting the sharp patches from the sequence, whereas the consistency metric improves the robustness by rejecting any residual jittering pixels from the previous steps.

**Step 3. Blind deconvolution.** After the lucky region fusion, the residual blur that is remaining in the image is largely caused by the diffraction limit of the lens, plus some minor effects due to any averaging process and artifacts of the previous steps. Therefore, a blind deconvolution algorithm is applied at last to remove these residual blurs.

Many turbulence mitigation methods, e.g. [4], [17], [20], [21], [26], use off-the-shelf blind deconvolution algorithms such as the classical method by Shan et al. [5]. However, the priors they used to define the point spread function are

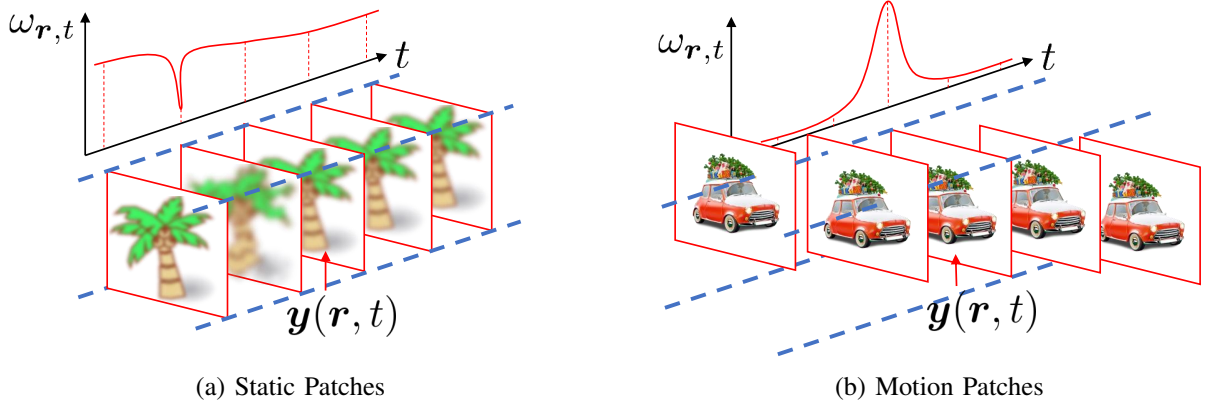


Fig. 3: **How the proposed method in Section 3.A can handle both static and dynamic scenes.** (a) For static scene, jittering patches with similar weight are averaged while bad patches with significant blur or geometric distortion are rejected by a much smaller weight. (b) For dynamic scene, the motion patches will have large weights only at its neighbors. Thus, the weighted average will not wash out the object.

generic, e.g.,  $\ell_1$  or total variation. These priors have almost no connection with the statistical behavior of turbulence, in particular the 5/3-power kernel proved by Fried [27]. A more recent approach by Nieuwenhuizen et al. uses deep neural networks [28]. However, with limited publicly available datasets to train the network, generalization of the method is limited. A physically more principled approach is by Hardie et al. [19], where they construct the ideal short-exposure point spread function by estimating the turbulence parameters. However, since the optical flow is not perfect and the lucky fusion sometimes introduces extra blur [4], [15], the residual errors can contaminate the predicted ideal PSF in [19].

Our proposed method is to construct the prior via basis expansion, similar to the analysis of Fried [25] but straightly in the spatial domain. It is a semi-supervised learning integrating numerical simulation and sparse modeling.

### III. PROPOSED METHOD

The proposed method consists of three ideas, each targeting at a specific step outlined in Figure 1. In this section we present the details of these ideas.

#### A. Space-time Non-Local Averaging

As shown in Figure 1, the first step of the turbulence mitigation process is to remove the tilts by using an optical flow algorithm. However, since the optical flow algorithm requires a reference image which is not available from the raw data, we need to construct it. Existing reference frame generation methods are largely based on temporal averaging. These methods fail to handle moving objects. Our goal is to develop a technique that can generate geometrically stable reference frames for both static and dynamic scenes.

**Proposed approach.** We propose to consider two criteria for the reference frame. First, the reference image needs to keep any moving objects intact. A temporal average is not a good option here because the moving objects will be washed out. Second, the reference image needs to remove the pixel jittering caused by turbulence. This ensures that the reference frame is sufficiently stabilized for the optical flow algorithm.

Let  $\mathbf{y}(\mathbf{r}, t) \in \mathbb{R}^d$  be a  $d$ -dimensional patch located at pixel  $\mathbf{r} = (i, j)$  and time  $t$ . Surrounding the patch we create a spatial-temporal search window  $\Omega_{\mathbf{r}} \times \Omega_t$  such that the possible patches contained are  $\{\mathbf{y}(\mathbf{r} + \Delta\mathbf{r}, t + \Delta t) \mid \Delta\mathbf{r} \in \Omega_{\mathbf{r}}, \Delta t \in \Omega_t\}$ . For each patch within this window, we define a pairwise distance with  $\mathbf{y}(\mathbf{r}, t)$ . This gives an array of distances measured at each space-time index  $(\mathbf{r}, t)$ :

$$\delta_{\mathbf{r},t}(\Delta\mathbf{r}, \Delta t) = \|\mathbf{y}(\mathbf{r}, t) - \mathbf{y}(\mathbf{r} + \Delta\mathbf{r}, t + \Delta t)\|^2. \quad (5)$$

Now, for every  $\Delta t$  (i.e., every frame in the temporal window) we scan through the spatial search window  $\Omega_{\mathbf{r}}$  and pick the  $\Delta\mathbf{r}$  which minimizes the distance:

$$\bar{\delta}_{\mathbf{r},t}(\Delta t) = \min_{\Delta\mathbf{r}} \delta_{\mathbf{r},t}(\Delta\mathbf{r}, \Delta t). \quad (6)$$

Consequently, we construct a space-time non-local weight

$$w_{\mathbf{r},t}(\Delta t) = \exp\{-\beta \bar{\delta}_{\mathbf{r},t}(\Delta t)\}, \quad (7)$$

which is a function of the frame index  $\Delta t$ , and  $\beta$  is a hyper-parameter controlling the decay rate of the weight. The estimated reference frame is the weighted average:

$$\hat{\mathbf{y}}_{\text{ref}}(\mathbf{r}, t) = \frac{\sum_{\Delta t \in \Omega_t} w_{\mathbf{r},t}(\Delta t) \mathbf{y}(\mathbf{r}, t + \Delta t)}{\sum_{\Delta t \in \Omega_t} w_{\mathbf{r},t}(\Delta t)}. \quad (8)$$

For overlapping patches, we take the average of the overlapping spatial locations to construct the final reference image.

**Intuition.** The intuition of the proposed method can be seen from the illustration shown in Figure 3. In this figure we consider two cases: A patch of a static object, and a patch of a moving object.

When the patch is static, the distortion caused by turbulence will be mostly random jittering. The jittering is a random variable with zero mean. Therefore, if we take the current patch  $\mathbf{y}(\mathbf{r}, t)$  and search for its neighbors in the adjacent frames, we are likely to find many neighbors. This can be seen from the high (and steady) values of the weights in Figure 3(a), except some severely jittered patches which give small weights. The usage of the minimum operation is critical



when defining  $\bar{\delta}_{\mathbf{r},t}(\Delta t)$ . If the minimum-distant neighboring patch (measured in terms of the  $\ell_2$  distance) still does not match well with the current patch, then this neighboring patch is likely to be either severely distorted or is a moving patch. In this case  $w_{\mathbf{r},t}(\Delta t)$  will be small.

When a patch contains moving objects, the distance  $\delta_{\mathbf{r},t}(\Delta \mathbf{r}, \Delta t)$  will be large for all  $(\Delta \mathbf{r}, \Delta t)$  because we are not able to find a good match within the spatial-temporal window. If this happens,  $\bar{\delta}_{\mathbf{r},t}(\Delta t)$  will remain large and so the weight  $w_{\mathbf{r},t}(\Delta t)$  will be small. As illustrated in Figure 3(b), the weight is large at the current time index, and is small for other time indices. By computing the weighted average, the moving object will remain intact because we do not perturb the current patch.

**Window size.** In principle, the spatial window size should be chosen according to the strength of the turbulence. However, the strength of the turbulence is not observable and any values obtained are essentially estimates, we take a different approach of choosing the window size empirically. Through our experimentation, we found that the window size of 11 pixels generally leads to satisfactory reconstruction results.

**Hyper-parameter  $\beta$ .** The hyper-parameter  $\beta$  in (7) is chosen according to the turbulence strength. Turbulence strength is usually quantified by the aperture-to-coherence ratio  $D/r_0$ . Here,  $D$  is the aperture diameter and  $r_0$  is the Fried parameter [2]. A typical value of  $D/r_0$  in incoherent imaging ranges from 1 to 5 [25], with 1 being weak turbulence and 5 being strong.

The parameter  $\beta$  is empirically determined as follows. For a specific  $D/r_0$  value, we synthetically generate random tilts to an image containing  $64 \times 64$  point sources using the tilt statistics (via the Zernike decompositions) [29]. We choose to work with synthetic point sources because it is content independent. We sweep through a range of  $\beta$ 's in the space-time non-local averaging operation, and pick the  $\beta$  that minimizes the error  $\|\hat{\mathbf{y}}_{\text{ref}}(\mathbf{r}, t) - \mathbf{y}_{\text{true}}(\mathbf{r}, t)\|$ , where  $\mathbf{y}_{\text{true}}$  is the ground truth image.

The result of the experiment is shown in Figure 4, where we plot the optimal  $\beta$  as a function of  $D/r_0$ . The result gives an empirically optimal  $\beta$ , which suggests that as the turbulence becomes stronger (larger  $D/r_0$ ), more frames are required to average out the randomness (hence  $\beta$  drops).

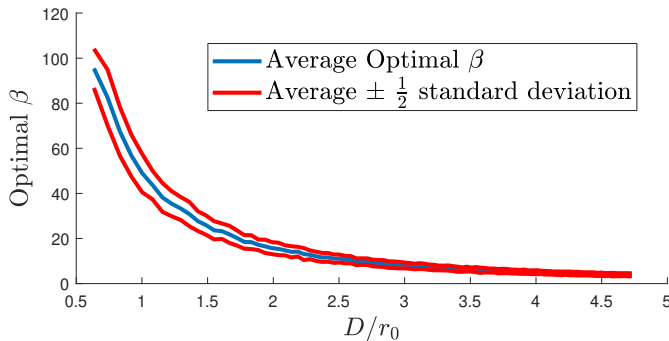


Fig. 4: **Hyper-parameter  $\beta$ .** We choose the optimal  $\beta$  according to the turbulence strength  $D/r_0$ . The blue curve is averaged over 100 trials.

## B. Lucky Region Fusion

The goal of the lucky region fusion is to extract sharp patches from the optical-flow aligned frames to form an image that is roughly diffraction limited. Conventional lucky region fusion techniques [4], [15] achieve the goal by computing the sharpness of a patch and pick the one with the highest sharpness score along the time axis. However, the sharpest patch may not necessarily be consistent with its neighbors. The winner-take-all approach is also not ideal because it creates discontinuity along object boundaries.

**Proposed approach.** The proposed lucky region fusion includes two new ideas. First, we introduce a geometric consistency metric in addition to the sharpness metric used in existing approaches. The geometric consistency metric ensures that the sharp patch is not due to artifacts. The second idea is to replace the winner-take-all by a weighted average. The weighted average ensures that the reconstructed lucky image has consistent object boundaries.

Let  $\mathbf{y}_{\text{flow}}(\mathbf{r}, t) \in \mathbb{R}^d$  be a  $d$ -dimensional patch of the  $t$ -th optical flow aligned image centered at pixel  $\mathbf{r}$ , and let  $\hat{\mathbf{y}}_{\text{ref}}(\mathbf{r}, t) \in \mathbb{R}^d$  be the corresponding patch of the estimated reference frame. We define the geometric consistency between patches  $\mathbf{y}_{\text{flow}}(\mathbf{r}, t)$  and  $\hat{\mathbf{y}}_{\text{ref}}(\mathbf{r}, t)$  as

$$\delta^G(\mathbf{r}, t) = \|\mathbf{y}_{\text{flow}}(\mathbf{r}, t) - \hat{\mathbf{y}}_{\text{ref}}(\mathbf{r}, t)\|^2. \quad (9)$$

The term  $\delta^G$  is called the geometric consistency, because it measures how close the aligned image is compared to the reference. If they are close, then the patch will be used to form the lucky fusion. The other metric measures the sharpness, defined as

$$\delta^S(\mathbf{r}, t) = \|\nabla \mathbf{y}_{\text{flow}}(\mathbf{r}, t)\|_1, \quad (10)$$

where  $\nabla$  is the spatial gradient operator. If an image is sharp, then the gradient  $\mathbf{y}_{\text{flow}}(\mathbf{r}, t)$  will have a large magnitude.

The overall impact of the two metrics is defined through the weight

$$w_{\mathbf{r},t}(\Delta t) = \underbrace{\exp\{-\alpha_1 \delta^G(\mathbf{r}, t + \Delta t)\}}_{\text{geometric}} \times \underbrace{\exp\{\alpha_2 \delta^S(\mathbf{r}, t + \Delta t)\}}_{\text{sharpness}}, \quad (11)$$

where  $\alpha_1$  and  $\alpha_2$  are two user defined parameters controlling the cutoff of the inliers and outliers. The typical values of  $\alpha_1$  and  $\alpha_2$  are determined empirically by verifying the percentile of the scores. One may increase the value of  $\alpha_1$  if the optical flow output contains more errors. Given the weights, the lucky region patch at pixel  $\mathbf{r}$  and time  $t$  is then

$$\hat{\mathbf{y}}_{\text{lucky}}(\mathbf{r}, t) = \frac{\sum_{\Delta t \in \Omega_t} w_{\mathbf{r},t}(\Delta t) \mathbf{y}_{\text{flow}}(\mathbf{r}, t + \Delta t)}{\sum_{\Delta t \in \Omega_t} w_{\mathbf{r},t}(\Delta t)}. \quad (12)$$

**Intuition.** Figure 5 illustrates the intuition of the weight  $w_{\mathbf{r},t}(\Delta t)$ . By virtue of optical flow, large motions are compensated. This means that the remaining distortions are mostly jitter. If the jittering in the patch  $\mathbf{y}_{\text{flow}}(\mathbf{r}, t)$  is still strong, then  $\delta^G(\mathbf{r}, t)$  will be large because  $\mathbf{y}_{\text{flow}}(\mathbf{r}, t)$  is different from  $\hat{\mathbf{y}}_{\text{ref}}(\mathbf{r}, t)$ . However, if the jittering is weak, then the patch

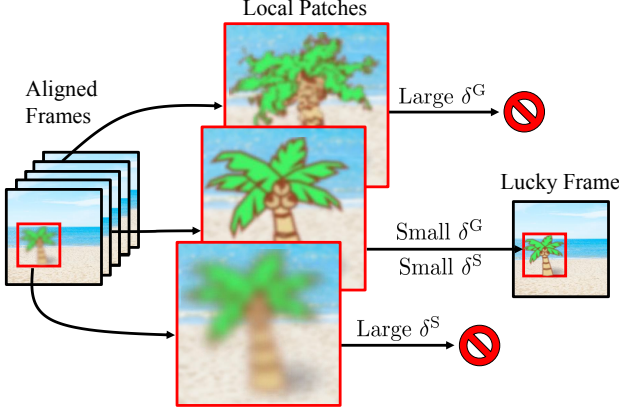


Fig. 5: **Lucky region metrics.** The optical flow aligned frames are assumed to be motion compensated. When determining a lucky patch, the method checks two scores: (i) the geometric consistency score which asks the candidate patch to look similar to the reference, and (ii) the sharpness score which asks the candidate patch to be sharp. Weighted average is formed across all the optical flow aligned frames.

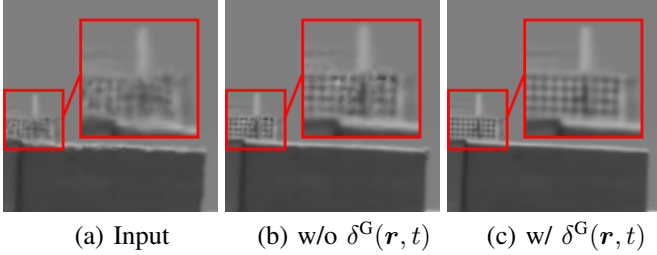


Fig. 6: **Effect of  $\delta^G(\mathbf{r}, t)$ .** Results of lucky region fusion generated by using  $\delta^G(\mathbf{r}, t)$  and without  $\delta^G(\mathbf{r}, t)$ . Note the improved stabilization of the edges, and geometric consistency.

is reliable and we should use it. Thus, a candidate patch is accepted to be used in the weighted average only when it has little geometric warping and it is sharp.

In Figure 6 we show a comparison between the lucky region fusion results generated with and without the geometric consistency term  $\delta^G(\mathbf{r}, t)$ . (The necessity of the sharpness metric is clear in the literature, for example [4], [15].) As we can observe from Figure 6, the geometric consistency term helps rejecting patches that are sharp but contain jittering artifacts.

### C. Blind Deconvolution

The objective of the blind deconvolution step is to remove the residual blur in the lucky frame. This blur comes from three sources: (i) The diffraction limit of the optical system, which is usually not severe. (ii) Weighted temporal averaging in the lucky region fusion step. (iii) Limited number of available frames, which does not allow us to pick the sharpest lucky frame.

**Proposed approach.** Following most of the literature on blind deconvolution, our proposed method starts with a pair of alternating minimization steps. The first step update the

latent image, and the second step updates the unknown point spread function (PSF):

$$\mathbf{z}^{k+1} = \underset{\mathbf{z}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{h}^k \otimes \mathbf{z}\|^2 + \lambda g(\mathbf{z}), \quad (13)$$

$$\mathbf{h}^{k+1} = \underset{\mathbf{h}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{h} \otimes \mathbf{z}^{k+1}\|^2 + \gamma r(\mathbf{h}). \quad (14)$$

In this pair of equations,  $\mathbf{y}$  is the entire lucky image which strictly speaking should be denoted by  $\hat{\mathbf{y}}_{\text{lucky}}(t)$ . We dropped the time index because blind deconvolution is performed per frame. The variable  $\mathbf{h}$  is the unknown PSF, and  $\mathbf{z}$  is the latent clean image. The functions  $g(\cdot)$  and  $r(\cdot)$  are regularizations for  $\mathbf{z}$  and  $\mathbf{h}$ , respectively. In this paper, we adopt the Plug-and-Play prior [30] for  $g(\cdot)$  with BM3D as the denoiser. The regularization  $r(\cdot)$  will be discussed next. Additional details of the Plug-and-Play prior can be found in [31]–[33].

**Modeling  $\mathbf{h}$ .** A major difference between a turbulence deconvolution and a standard deconvolution is that we have a well-defined model for the turbulence PSF as we discussed in the Background section. Specific to the context of turbulence, we highlight a few key points:

- The turbulence PSF is defined through the random phase distortion function  $\phi(\mathbf{f})$  in the frequency domain. See [29, Eqs 6-9].
- The random phase distortion  $\phi(\mathbf{f})$  can be represented in the Zernike space via [29, Eq. 15]:

$$\phi(\mathbf{f}) = \sum_{j=1}^N a_j Z_j(\mathbf{f}). \quad (15)$$

with  $Z_j$  being the  $j$ -th Zernike Basis, and  $a_j$  is the corresponding Zernike coefficient.

- The Zernike coefficients have physical interpretations. E.g.,  $a_2$  and  $a_3$  are the tilts, and  $a_4$  is the defocus, etc. The statistics of  $a_j$ 's are defined through intermodal and spatial correlations. See [29, Eq 20, Eq 33].

In order to encapsulate the physics of the PSFs (which are defined in the phase space), we propose the following linear expansion scheme by translating the distortions in the phase domain to the principal components in the spatial domain.

Consider a simulator that generates a vector of random Zernike coefficients  $\mathbf{a} = \{a_j\}_{j=1}^N$ . Suppose the simulator has generated a set of  $M$  vectors,  $\{\mathbf{a}_1, \dots, \mathbf{a}_M\}$ . Since every  $\mathbf{a}_i$  gives to a random phase, and every random phase corresponds to a random PSF, by generating  $M$  Zernike coefficient vectors we will obtain a collection of  $M$  random PSFs  $\{\mathbf{h}_1, \dots, \mathbf{h}_M\}$ . By decomposing the random PSFs using the principal component analysis (PCA), we can extract a set of basis representations  $\{\mathbf{u}_j\}_{j=1}^p$  via

$$\{\mathbf{u}_j\}_{j=1}^p = \text{PCA}\{\mathbf{h}_1, \dots, \mathbf{h}_M\}.$$

Thus for every  $\mathbf{h}$ , there exist  $\{w_j\}_{j=1}^p$  such that

$$\mathbf{h} = \sum_{j=1}^p w_j \mathbf{u}_j, \quad (16)$$

where  $w_j$  is the basis expansion coefficient. In other words, we converted the linear expansion of the phase numerically

to a linear expansion of the PSF. This is a data-driven procedure, where the data comes from the physical model of the turbulence.

**Removing the tilts.** The above linear expansion model is valid when  $\mathbf{y}$  is the raw turbulent distorted image. However, since  $\mathbf{y}$  is the lucky frame, additional modeling of the other parts of the pipeline is needed.

Referring back to the random phase concept in (15), we note that since the optical flow has compensated for the pixel movements, the tilts of the phase can be approximately dropped. More precisely, we consider a random phase  $\phi(\mathbf{f})$  generated by a set of Zernike coefficients  $\{a_1, \dots, a_N\}$ . Using [29, Eq. 13], we define the tilt-compensated phase as

$$\varphi(\mathbf{f}) = \phi(\mathbf{f}) - \alpha^T \mathbf{f}, \quad (17)$$

for some linear terms  $\alpha$ . The new PSFs defined through the tilt-compensated phase  $\varphi(\mathbf{f})$  are center-aligned, which can be interpreted as outputs of the optical flow. When synthesizing the simulated training data, the tilts can be removed by setting  $a_2 = a_3 = 0$ .

After removing the tilts, the lucky fusion step is used to pick select patches according to two criteria we described in the lucky fusion section. To model this during the training data synthesis, we note that the “wildness” of the PSF is determined by the magnitude of the Zernike coefficients as shown in Figure 7. Therefore, to obtain good patches, we perform an importance sampling by rejecting large Zernike coefficients. Specifically, we define  $\mathbf{a}_{4:N} = \{a_4, \dots, a_N\}$  as the vector of Zernike coefficients from  $a_4$  (because  $a_2 = a_3 = 0$ , and  $a_1$  is a constant which does not affect the shape of the PSF.) We check whether  $\|\mathbf{a}_{4:N}\|^2 \leq \kappa$  for some threshold  $\kappa$ . If the magnitude  $\|\mathbf{a}_{4:N}\|^2$  is above the threshold, we say that the PSF should not be used to form the PCA basis because lucky fusion will likely skip this PSF.

The proposed modeling of the optical flow and the lucky frame is justified by the optics and turbulence literature. In particular, tilt-compensation is a standard practice in adaptive optics where the tilt vector  $\alpha$  is measured by phase offsets. When averaged over many random realizations, the tilt-compensated PSFs will give the short-exposure PSF. The small coefficient approach is also justified. It was first analyzed by Fried in [25], where the threshold was 1 rad<sup>2</sup> for the KarhunenLoève expansion of the phase. In our case, the principle remains but the threshold is set empirically.

**Updating  $\mathbf{h}$ .** The update of the PSF in (14) now becomes

$$\mathbf{w}^{k+1} = \underset{\mathbf{w}}{\operatorname{argmin}} \left\| \mathbf{y} - \left( \sum_{j=1}^p w_j \mathbf{u}_j \right) \circledast \mathbf{z}^k \right\|^2 + \gamma r(\mathbf{w}) \quad (18)$$

$$\mathbf{h}^{k+1} = \sum_{j=1}^p w_j^{k+1} \mathbf{u}_j. \quad (19)$$

The first step (18) defines the weight of the latent PSF. The prior for the weight is denoted as  $r(\mathbf{w})$  for some regularization function  $r(\cdot)$ . The second step (19) updates the PSF using the linear expansion concept. Depending on the prior  $r(\cdot)$ , e.g.,  $\ell_1$  prior, (18) can be solved using standard optimization packages, e.g., CVX [34] or the alternating direction methods

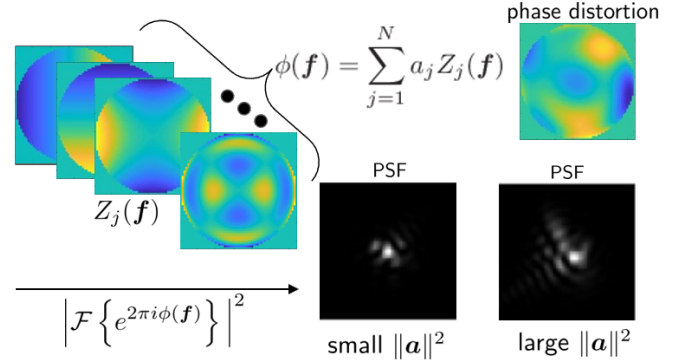


Fig. 7: **Zernike representation of phase.** The phase  $\phi(\mathbf{f})$  can be expressed as a linear combination of the Zernike basis  $Z_j(\mathbf{f})$ , with basis coefficients  $a_j$ . If the coefficient vector  $\mathbf{a} = [a_j]$  has a large magnitude, then the distortion in the resulting PSF will be large.

of multiplier [30], [35].

**Defining  $r(\mathbf{w})$ .** The prior distribution  $r(\mathbf{w})$  can be determined empirically. Consider the set of generated PSFs  $\{\mathbf{h}_1, \dots, \mathbf{h}_M\}$ . For each PSF  $\mathbf{h}_i$ , we seek the *sparsest* representation using the bases  $\{\mathbf{u}_j\}_{j=1}^p$ . We are interested in such representation because during the blind deconvolution step, the current estimate of the PSF is never perfect. Sparse representation entails the most influential bases while rejecting potential residue errors.

The regularization function is defined as

$$r(\mathbf{w}) = \sum_{j=1}^p \frac{|w_j|}{\sigma_j}, \quad (20)$$

where  $\sigma_i$  is the standard deviation of  $w_i$  learned from data. This is equivalent to assuming that the prior distribution of  $\mathbf{w}$  follows a field-of-expert model:

$$p(\mathbf{w}) = \exp \left\{ - \sum_{j=1}^p \frac{|w_j|}{\sigma_j} \right\}, \quad (21)$$

such that  $r(\mathbf{w}) = -\log p(\mathbf{w})$ . Note that the regularization in Equation (20) is a weight  $\ell_1$  norm. The  $\ell_1$  norm is used here as a compromise between model fidelity and complexity. Norms other than  $\ell_1$  can potentially lead to better fitting (as in the mean squared fitting error), but the computational complexity of solving the corresponding optimization will become higher. The choice of the  $\ell_1$  norm is further justified by the shape of the empirical distributions shown in Figure 8. We observe that the distributions indeed follow an exponentially decaying function, and has a tail heavier than a Gaussian.

To find the sparse representation, we solve the following problem offline (during the training phase):

$$\mathbf{w}^{(i)} = \underset{\mathbf{w}}{\operatorname{argmin}} \|\mathbf{w}\|_0 \text{ s.t. } \left\| \mathbf{h}_i - \sum_{j=1}^p w_j \mathbf{u}_j \right\|^2 \leq \tau, \quad (22)$$

where  $\tau$  is a pre-define threshold. Note that Equation (22) is a standard  $\ell_0$  basis pursuit denoise problem, which can be solved

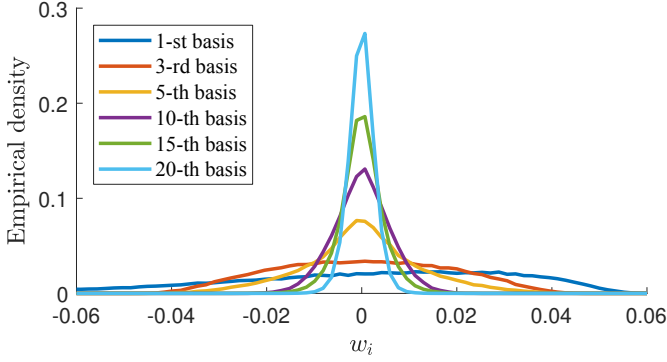


Fig. 8: **Distribution of  $w$ .** Empirical distributions of the weights by solving (22) with 100,000 different simulated PSF  $h$ . The empirical distribution demonstrates a double-sided exponential distribution.

TABLE I: Runtime (sec) for generating 100 images of sizes  $256 \times 256$

Reference Frame	3.17 sec
Optical Flow	260.02 sec
Lucky Fusion	6.71 sec
Blind Deconvolution	336.44 sec
Overall	606.34 sec

using tools such as orthogonal matching pursuit [36]. After solving the problem, we will have a collection of the expansion coefficients  $\{w^{(i)}\}_{i=1}^M$ . Plotting the empirical density of the collection  $\{w_j^{(1)}, \dots, w_j^{(M)}\}$  will give us plots shown in (8), for each of the  $j$ -th basis. The empirical distributions will then define the regularization function  $r(w)$ .

#### D. Computing Time

We report the computing time required to process a sequence with 100  $256 \times 256$  frames in Table I. The optical flow [14] is implemented in C++ and other steps are implemented in MATLAB. The algorithm is tested on an Intel I7-7700HQ CPU with 16 GB memory. It is observed that the bottle-neck of the run time is the optical flow, where 100 frame-to-frame image alignment needs to be done. The other bottleneck is the blind deconvolution, where the alternating minimization is used. Compared to existing methods which also use optical and blind deconvolution, the additional cost of the proposed method is the reference generation step. However, this step contributes to less than 1% of the total runtime.

### IV. EXPERIMENTAL RESULTS

#### A. Evaluation Data

We evaluate the proposed method on three types of data:

- **Synthetic sequences**, where we have ground truths to compute the PSNRs. Our simulation is based on the method presented in [29], with simulation parameters listed in Table II. The refractive index constant  $C_n^2$  ranges from  $1 \times 10^{-15} \text{m}^{-2/3}$  to  $5 \times 10^{-15} \text{m}^{-2/3}$ , which

correspondingly makes the aperture-to-coherence ratio  $D/r_0$  ranges between 1 and 4.

- **Real Hot-air Sequences**, popularly used in the computer vision literature, e.g., Hirsch et al. [37] and Anantrasirichai et al. [17]. These sequences are collected by placing the camera behind a hot-air burner. These hot-air sequences are good surrogates of the true long-range turbulence, but there are limitations.
- **Real Long-range Sequences**, collected on 25 September, 2019. The path length between the camera and the object is approximately 4km at a temperature of 30°C. In addition, we also use the NATO RTO SET/RTG-40 dataset reported by Leonard, Howe and Oxford in [38].

All the sequences used in this paper contain 100 frames, which are consistent with [4], [16], [17], [37].

TABLE II: Simulation Parameters

Parameter	Value
Path length	$L = 4\text{km}$
Aperture Diameter	$D = 0.1\text{m}$
Focal Length	$d = 0.4\text{m}$
Wavelength	$\lambda = 525\text{nm}$
Zernike Phase Size	$64 \times 64$ pixels
Image Size	$512 \times 512$ pixels
Nyquist spacing (object plane) $\frac{L\lambda}{2D}$	$\delta_0 = 1.05\text{mm}$
Nyquist spacing (focal plane) $\frac{d\lambda}{2D}$	$\delta_f = 1.05\mu\text{m}$

We like to comment on the difference between a hot-air sequence and the long-range sequences. Figure 9 shows three video sequences. The first two sequences Chimney [37] and A3 . Back car [17] are created by hot-air burners, and the third sequence New Chimney is a real long-range turbulence captured at 4km. In each of the sub-figures, we show a typical frame and a manually selected “lucky” frame. For the two hot-air sequences, we observe that the “lucky” frames actually have very high quality, so high-quality that even post-processing becomes unnecessary. The reason for such phenomenon is that the turbulence caused by a hot-air burner is isoplanatic which is quite different from the real long-range turbulence. If we look at the real long-range turbulence case in Figure 9, we observe that the best lucky frame is still quite similar to the other frames.

The difference between the hot-air and the long-range turbulence can be quantitatively compared. Consider the *normalized gradient*

$$\mathcal{E}_t \stackrel{\text{def}}{=} \frac{\sum_{\mathbf{r}} \|\nabla \mathbf{y}(\mathbf{r}, t)\|_1}{\max_t \sum_{\mathbf{r}} \|\nabla \mathbf{y}(\mathbf{r}, t)\|_1}, \quad (23)$$

which is a metric measuring the sharpness of the frames (relative to the sharpest frame in the sequence). If we plot  $\mathcal{E}_t$  as a function of time, we observe that long-range turbulence has substantially different behavior than hot-air. In particular, the curves in Figure 10 show that “lucky frames” generally do not appear often in long-range sequences, but happens occasionally in hot-air sequences. As a result, we argue that it is necessary to evaluate a method using not just the hot-air sequences but also the real long-range turbulence sequences.



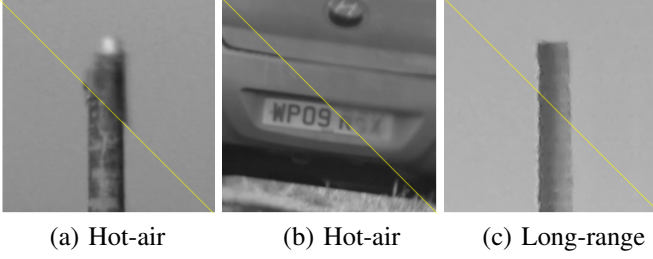


Fig. 9: **Comparing hot-air and real turbulence.** Hot-air sequences can have really “lucky” frames, but a real turbulence sequence is consistently degraded. Thus, just using hot-air sequences for evaluation is inappropriate. For each sub-figure, the upper right is a typical frame randomly picked from a sequence, and the bottom left is a manually selected lucky frame from the same sequence. (a) Chimney [37] (hot-air), (b) A3. Back car [17] (hot-air), and (c) New Chimney (long-range).

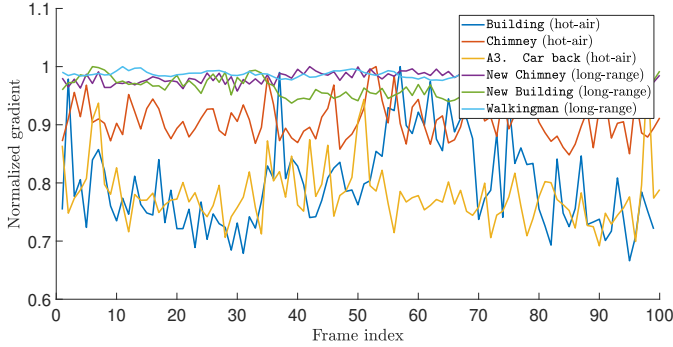


Fig. 10: **Normalied gradient for hot-air and long range sequences.** Short range hot-air sequences exhibit a strong fluctuation in the the temporal normalized gradients, whereas long range turbulence sequences do not.

### B. Evaluating Reference Frames

We evaluate the effectiveness of the proposed reference frame extraction method using two sets of experiments.

**Resolution enhancement.** The first experiment is a synthetic experiment that tests the minimal resolvable distances. We consider a square image consisting of two horizontal lines separated by a distance of 10 pixels, as shown in Figure 11(a). We apply the simulator presented in [29], with parameters specified in Table II and  $C_n^2 = 1.5 \times 10^{-15} \text{m}^{-2/3}$ . Figure 11(b) shows one random realization of the turbulent distorted image. The reference frames generated by competing methods are shown in Figure 11(c) - Figure 11(f). For the RPCA method by Lin et al. [39], default parameters are used. For the methods by Lau et al. [21], we retain 25% of the original frames as suggested by the authors. For our proposed method, we use a  $7 \times 7$  space-time search window. As can be seen in the figures, the proposed method has the best recovery.

A more quantitative result can be seen in Figure 11(g), where we plot the cross-section of the estimated reference frames. In Figure 11(h) we further plot the dynamic range of the recovered pattern, which is defined as the gap between the peak and the valley of the intensity plots. As we can see in

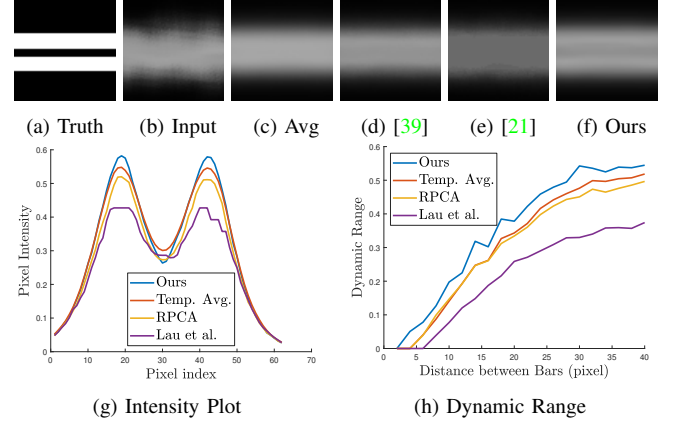


Fig. 11: **Evaluating the reference frame method.** The objective of this experiment is to evaluate which reference frame reconstruction method can achieve better resolution. (a) Ground truth. Spacing between bars is 4 pixel. (b) A simulated frame (c) Temporal average. (d) RPCA (e) Lau’s reference frame (f) Our reference frame (g) Dynamic range of the reference frame with varying spacing (h) The cross-section at center of the bars when the spacing is 10.

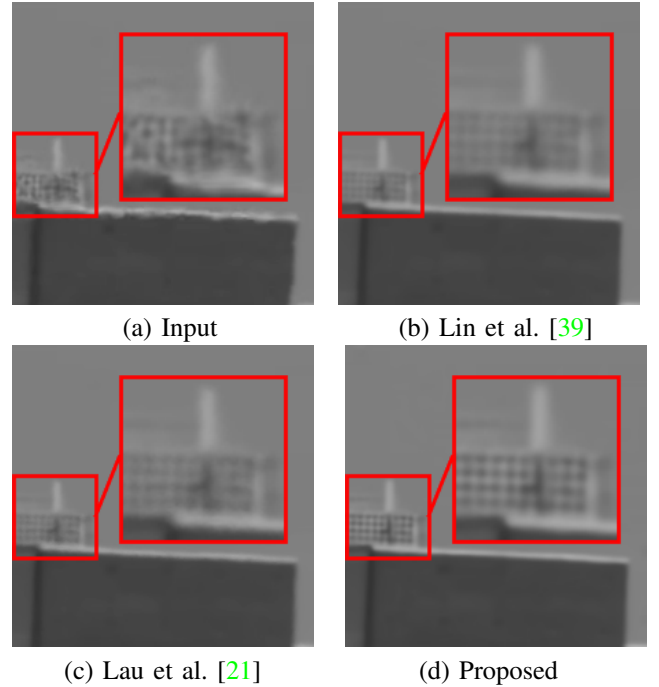


Fig. 12: **Evaluating the reference frame method using real static data.** This figure shows the reconstructed references frames by competing methods.

the intensity plot and the dynamic range plot, the proposed method generates the best contrast of the patterns. The RPCA by Lin et al. [39] and the frame selection method by Lau et al. [21] both adopt a low rank approximation approach. As such, the recovered patterns have lower intensity values because the sparse components are subtracted from the images.

**Test on real data.** In Figure 12 we show a visual comparison of the extracted reference frames from a real turbulence

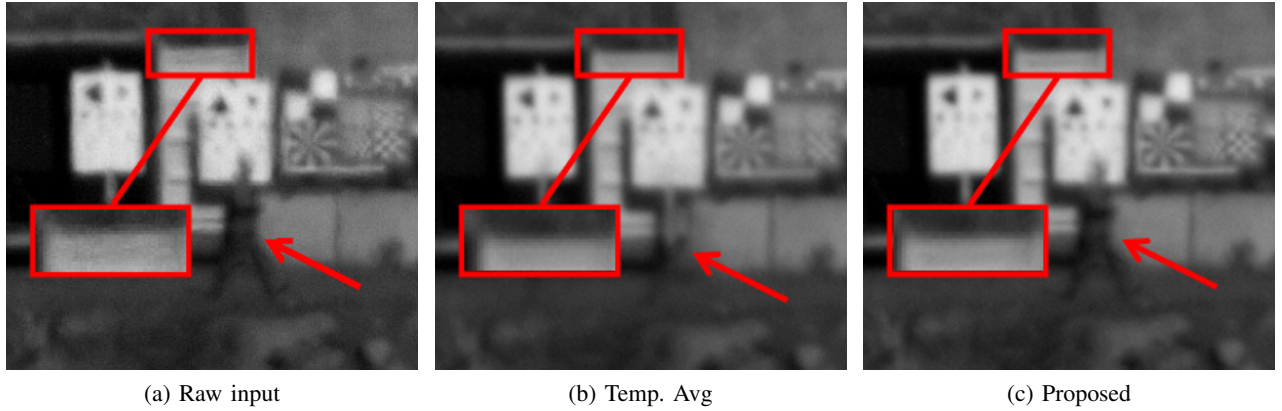


Fig. 13: **Evaluating the reference frame method using real dynamic data.** (a) Raw input. (b) Temporal averaging. Observe that the man is blurred over 100 frames. (c) Ours. Observe that the man remains in the image while the background is stabilized.

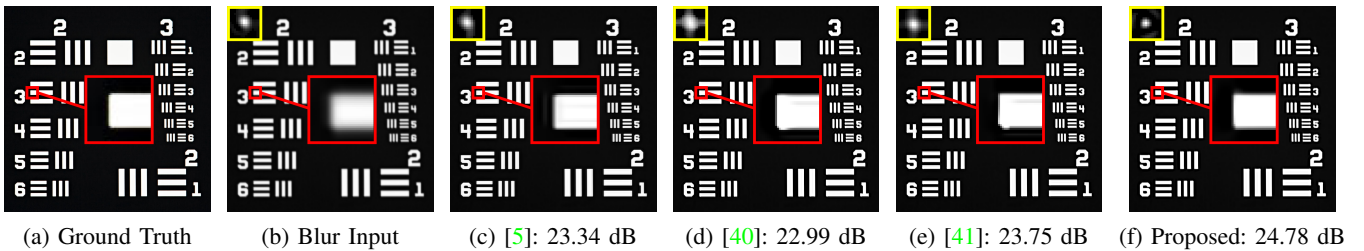


Fig. 14: **Comparing blind deconvolution using synthetic data.** In this example we use the USAF resolution chart. We blur the ground truth using an ideal short-exposure PSF, and reconstruct the images using different methods.

sequence, where we compare the RPCA by Lin et al. [39], the frame selection by Lau et al. [21], and the proposed method. As we can see in the results, the proposed method generates a noticeably sharper reference frame.

In Figure 13 we show a sequence containing a moving object. Because of the moving object, RPCA essentially becomes the temporal average when we consider the largest principal component. Our result in Figure 13 shows that the temporal average will wash out the moving object. In contrast, the proposed method can retain it.

### C. Evaluating Blind Deconvolution

We decouple the blind deconvolution from the rest of the pipeline and evaluate the effectiveness of the proposed basis expansion idea. In this experiment, we compare with a classic method by Shan et al. [5] and two recent deep neural networks by Chakrabarti [40] and Xu et al. [41]. The codes and models were obtained from the authors’ websites. We use pre-trained models for the two neural networks, and fine-tune the parameters in [5]. For our proposed method, we fixed  $\lambda = 0.05$  and  $\gamma = 1 \times 10^{-3}$  for all experiments reported in this section.

In Figure 14 we show a visual comparison using the USAF resolution chart with  $D/r_0 = 1.4$ . We blur the ground truth image using an ideal short-exposure PSF, and apply various blind deconvolution methods to recover the blur and the image. As we can see, the proposed method generates an image containing the least amount of artifacts. The estimated PSF is also more structured and interpretable than the neural network methods.

We evaluate the result using the 24 images in the Kodak image dataset. Each image is blurred with 50 random point spread functions, where each one is constructed by averaging 10 short-exposure PSFs under different turbulence levels (with  $D/r_0$  from 1 to 2.6 with step size 0.4). This creates a total of 1200 testing images. The average PSNR value with respect to the ground truth under different turbulence levels is reported in Figure 15. It is evident from the table that the proposed method has considerably better reconstruction PSNR than the competing methods. We argue that this is due to the improved physics-inspired prior for the point spread functions.

### D. Evaluating the Overall Pipeline

We now evaluate the effectiveness of the overall proposed algorithm by comparing it with several existing methods: Sobolev gradient flow (SG) by Lou et al. [16], near-diffraction-limit (NDL) by Zhu and Milanfar [4] with deblurring using Shan et al. [5], and wavelet fusion method (CLEAR) by Anantrasirichai et al. [17]. The codes are provided by the original authors and the internal parameters are tuned according to the best of our knowledge.

**Synthetic dataset.** We synthetically generate turbulent sequences for 10 “standard” images (e.g., Lena, House, Cameraman, etc), where each sequence consists of 100 frames. The turbulence levels are set as  $D/r_0 = 1.4, 2.8$  and  $4.0$  respectively. The results are reported in Figure 16, where each PSNR value corresponds to an average over the 10 testing sequences. A visual comparison of the House sequence is shown in Figure 17. While all the algorithms yield similar

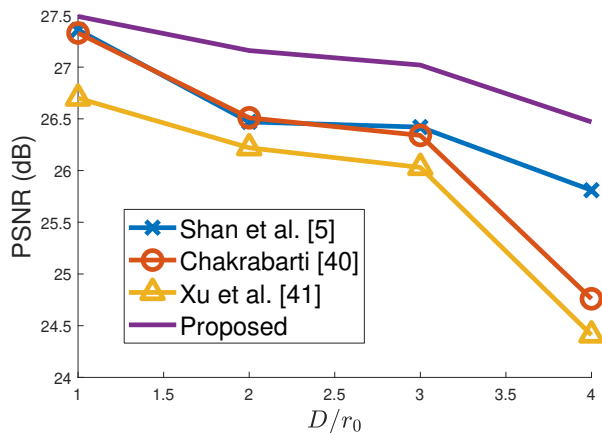


Fig. 15: **PSNR curves of blind deconvolution using synthetic data.** The PSNRs are averaged over 24 images of the Kodak image dataset, and every image is tested 50 times for different random turbulence PSFs. Thus every data point reported in this table is an average over 1200 testing scenarios.

results under a low turbulence level, our approach offers superior performance as the turbulence level increases.

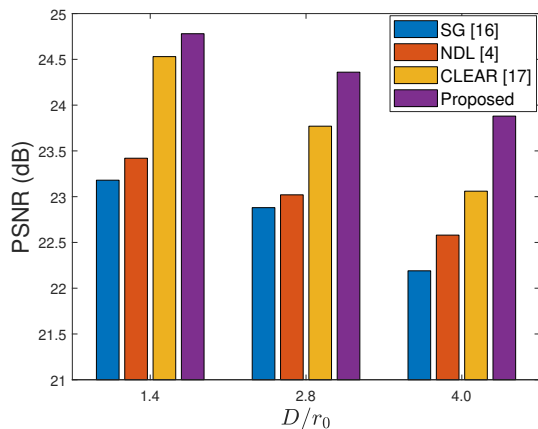


Fig. 16: **PSNR of the entire pipeline using synthetic data.** The PSNRs reported here are averaged over 10 sequences.

**Real sequences.** We compare the proposed method with existing methods on two sets of real turbulence sequences: Hot-air turbulence shown in Figure 18, and a real long-range turbulence shown in Figure 19. We include both comparisons to demonstrate the universality of the proposed method. As we can see in both figures, the proposed method has more superior performance than the competing methods, both in terms of sharpness and visual consistency.

**Moving object sequences.** We finally compare methods for sequences containing moving objects. In particular, we compare our method with a very recent approach by Anantrasirichai et al. [24], which is a generalization of the wavelet fusion previously presented in [17] to moving sequences. The dataset we used consists of two synthetic video sequences and a real video sequence from the NATO RTO SET/RTG-40 dataset [38].

The results shown in Figure 20 suggest that static methods

such as CLEAR [17] does not handle motion as it creates ghosting artifacts. If the object is moving quickly, e.g., the last row of Figure 20, the method will completely wash out the object. Segmentation-based method such as [24] rely heavily on the quality of the masks, which are generally of sub-optimal quality. The proposed method demonstrates the most reliable reconstruction among the three, as is evident from the figure. Additional reconstruction results of the proposed method are shown in Figure 21.

## V. CONCLUSION

We presented a unified turbulence mitigation algorithm that can restore both static and dynamic scenes. The method includes three new ideas: (i) A novel space-time non-local averaging method that can extract stable reference frames for static and dynamic scenes. (ii) A new lucky region fusion method that is robust against motion registration error. (iii) A new prior model for turbulent PSFs by utilizing the Zernike representation of the phase. Ablation studies of the method show that all components are essential for the overall performance. In particular, the reference extraction step has substantially better results than temporal averaging and robust principal component analysis, and the blind deconvolution outperforms existing deep learning based methods. We evaluated the algorithm using both synthetic and real turbulence image data. Compared to existing methods, our approach offers substantially better reconstruction quality on both static and dynamic scenes.

## ACKNOWLEDGMENT

The work is funded, in part, by the Air Force Research Lab and Leidos, and by the National Science Foundation under grants CCF-1763896 and CCF-1718007. The authors would like to thank Michael Rucci, Barry Karch, Daniel LeMaster and Edward Hovenac of the Air Force Research Lab for many insightful discussions. The authors also thank Nantheera Anantrasirichai, a co-author of [17], and Peyman Milanfar, a co-author of [4], for generously sharing their MATLAB code.

This work has been cleared for public release carrying the approval number 88ABW-2020-0292.

## REFERENCES

- [1] V. I. Tatarskii, *Wave propagation in a turbulent medium*, 1st ed. New York: Dover Publication, 1961. 1, 2
- [2] M. Roggemann, B. Welsh, and B. Hunt, *Imaging Through Turbulence*, ser. Laser & Optical Science & Technology. Taylor & Francis, 1996. 1, 2, 5
- [3] D. L. Fried, "Anisoplanatism in adaptive optics," *J. Opt. Soc. Am.*, vol. 72, no. 1, pp. 52–61, Jan 1982. 1
- [4] X. Zhu and P. Milanfar, "Removing atmospheric turbulence via space-invariant deconvolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 157–170, Jan. 2013. 2, 3, 4, 5, 6, 8, 10, 11, 12, 13
- [5] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *ACM Transactions on Graphics (SIGGRAPH)*, vol. 27, no. 3, Aug. 2008. 2, 3, 10
- [6] R. Tyson, *Principles of Adaptive Optics*, ser. Series in Optics and Optoelectronics. CRC Press, 2010. 1
- [7] A. Labeyrie, "Attainment of Diffraction Limited Resolution in Large Telescopes by Fourier Analysing Speckle Patterns in Star Images," *Astronomy and Astrophysics*, vol. 6, p. 85, May 1970. 1



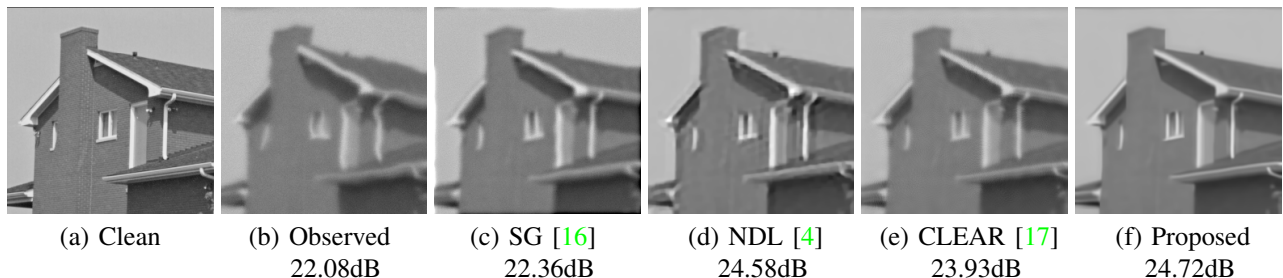


Fig. 17: **Overall comparison using synthetic static data.** The sequence is distorted by a turbulence level of  $D/r_0 = 2.8$ . PSNR comparisons of the entire dataset is shown in Figure 16.

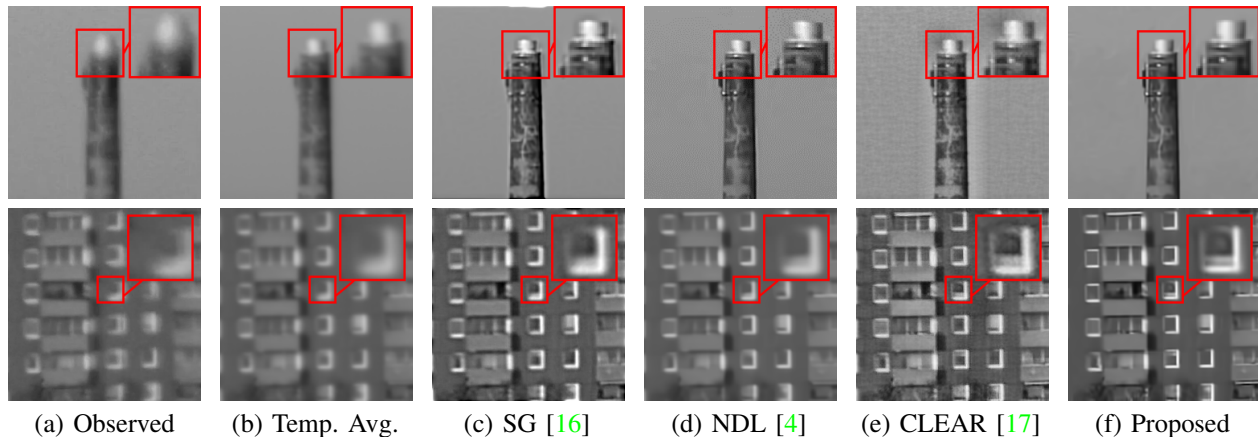


Fig. 18: **Overall comparison using real “hot-air” static sequences.** The sequences were previously reported in [37]. [Top] The chimney. [Bottom] Building. Note that these are real videos where ground truths are not available.

- [8] K. T. Knox and B. J. Thompson, “Recovery of Images from Atmospherically Degraded Short-Exposure Photographs,” *Astrophysical Journal*, vol. 193, p. L45, Oct 1974. 1
- [9] A. W. Lohmann, G. Weigelt, and B. Winitzer, “Speckle masking in astronomy: triple correlation theory and applications,” *Appl. Opt.*, vol. 22, no. 24, pp. 4028–4037, Dec 1983. 1
- [10] C. J. Carrano, “Speckle imaging over horizontal paths,” in *High-Resolution Wavefront Control: Methods, Devices, and Applications IV*, J. D. Gonglewski, M. A. Vorontsov, M. T. Gruneisen, S. R. Restaino, and R. K. Tyson, Eds., vol. 4825, International Society for Optics and Photonics. SPIE, 2002, pp. 109 – 120. 1
- [11] J. Goodman, *Statistical Optics*, 2nd ed. Hoboken, New Jersey: John Wiley and Sons Inc., 2015. 2
- [12] A. Kolmogorov, “The Local Structure of Turbulence in Incompressible Viscous Fluid for Very Large Reynolds’ Numbers,” *Akademiia Nauk SSSR Doklady*, vol. 30, pp. 301–305, 1941. 2
- [13] D. L. Fried, “Optical resolution through a randomly inhomogeneous medium for very long and very short exposures,” *J. Opt. Soc. Am.*, vol. 56, no. 10, pp. 1372–1379, Oct 1966. 2
- [14] C. Liu, “Beyond pixels: Exploring new representations and applications for motion analysis.” Ph.D. dissertation, Massachusetts Institute of Technology, Jan. 2009. 3, 8
- [15] M. Aubailly, M. A. Vorontsov, G. W. Carhart, and M. T. Valley, “Automated video enhancement from a stream of atmospherically-distorted images: the lucky-region fusion approach,” *Proceedings of SPIE*, vol. 7463, p. 74630C, Aug. 2009. 3, 4, 5, 6
- [16] Y. Lou, S. Ha Kang, S. Soatto, and A. Bertozzi, “Video stabilization of atmospheric turbulence distortion,” *Inverse Problems and Imaging*, vol. 7, no. 3, pp. 839–861, Aug. 2013. 3, 8, 10, 12, 13
- [17] N. Anantrasirichai, A. Achim, N. G. Kingsbury, and D. R. Bull, “Atmospheric turbulence mitigation using complex wavelet-based fusion,” *IEEE Transactions on Image Processing*, vol. 22, no. 6, pp. 2398–2408, June 2013. 3, 8, 9, 10, 11, 12, 13
- [18] J. Gilles and S. Osher, “Wavelet burst accumulation for turbulence mitigation,” *Journal of Electronic Imaging*, vol. 25, p. 033003, May 2016. 3
- [19] R. C. Hardie, M. A. Rucci, A. J. Dapore, and B. K. Karch, “Block matching and wiener filtering approach to optical turbulence mitigation and its application to simulated and real imagery with quantitative error analysis,” *Optical Engineering*, vol. 56, no. 7, p. 071503, 2017. 3, 4
- [20] Y. Xie, W. Zhang, D. Tao, W. Hu, Y. Qu, and H. Wang, “Removing turbulence effect via hybrid total variation and deformation-guided kernel regression,” *IEEE Transactions on Image Processing*, vol. 25, no. 10, pp. 4943–4958, Oct 2016. 3
- [21] C. P. Lau, Y. H. Lai, and L. M. Lui, “Restoration of atmospheric turbulence-distorted images via RPCA and quasiconformal maps,” *Inverse Problems*, vol. 35, no. 7, Mar. 2019. 3, 9, 10
- [22] K. K. Halder, M. Tahtali, and S. G. Anavatti, “Geometric correction of atmospheric turbulence-degraded video containing moving objects,” *Optics Express*, vol. 23, no. 4, pp. 5091–5101, Feb 2015. 3
- [23] 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
- [24] N. Anantrasirichai, A. Achim, and D. Bull, “Atmospheric turbulence mitigation for sequences with moving objects using recursive image fusion,” in *IEEE International Conference on Image Processing*, Oct 2018, pp. 2895–2899. 3, 11, 13
- [25] D. L. Fried, “Probability of getting a lucky short-exposure image through turbulence,” *J. Opt. Soc. Am.*, vol. 68, no. 12, pp. 1651–1658, Dec 1978. 3, 4, 5, 7
- [26] R. He, Z. Wang, Y. Fan, and D. Feng, “Atmospheric turbulence mitigation based on turbulence extraction,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 1442–1446. 3
- [27] D. L. Fried, “Optical resolution through a randomly inhomogeneous medium for very long and very short exposures,” *J. Opt. Soc. Am.*, vol. 56, no. 10, pp. 1372–1379, Oct 1966. 4
- [28] R. Nieuwenhuizen and K. Schutte, “Deep learning for software-based turbulence mitigation in long-range imaging,” in *Artificial Intelligence and Machine Learning in Defense Applications*, J. Dijk, Ed., vol. 11169, International Society for Optics and Photonics. SPIE, 2019, pp. 153 – 162. 4
- [29] N. Chimmitt and S. H. Chan, “Simulating anisoplanatic turbulence by



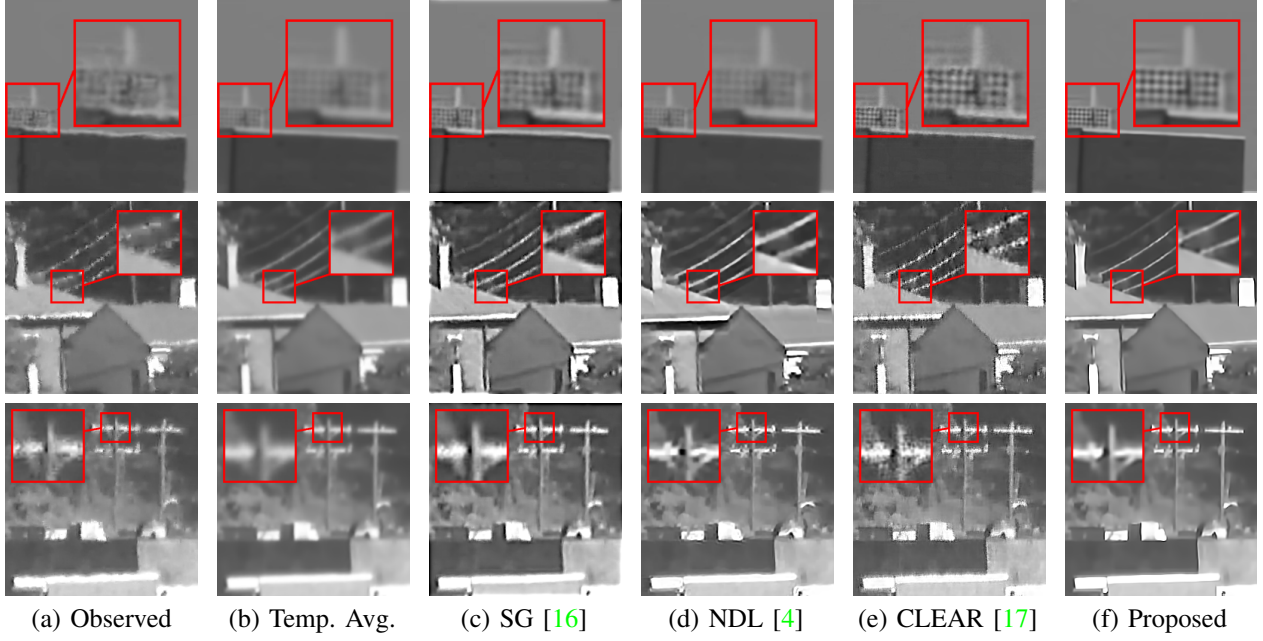


Fig. 19: **Overall comparisons using real long-range static sequences.** The first sequence was obtained from Youtube [42]. The rest were captured by Panasonic Full HD Camcorder HC-V180K (aperture diameter 24mm and focal length 103mm), at a distance of 4km and temperature of 30°C.

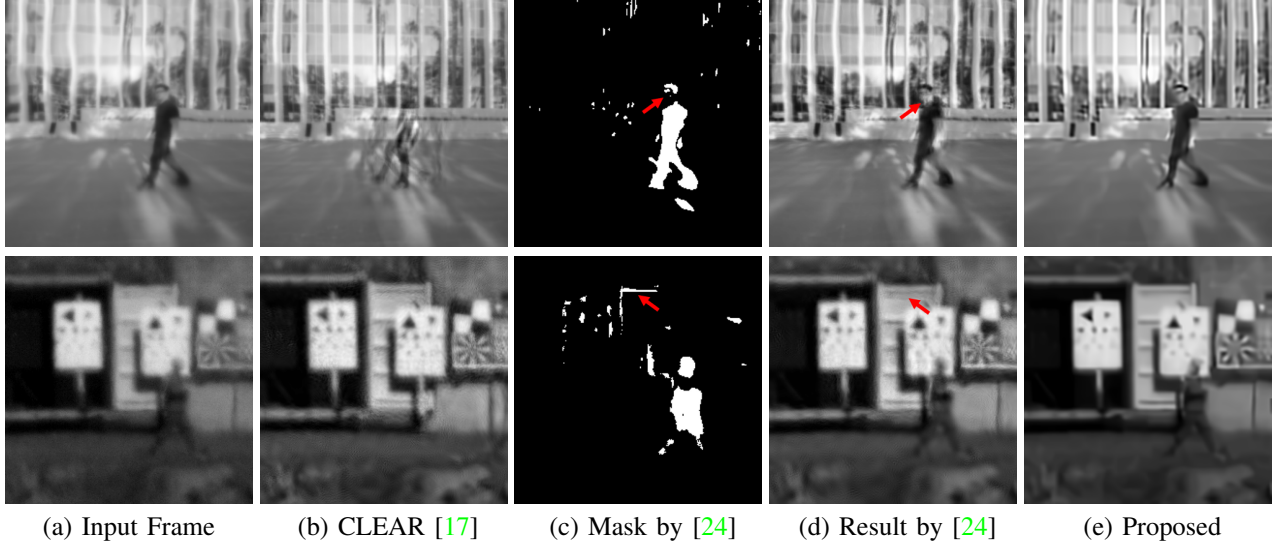


Fig. 20: **Overall comparison with synthetic and real dynamic sequences.** The first video is synthetically generated from clean video using the simulator presented in [29], whereas the last one is a real turbulence sequence from the NATO dataset.

- sampling correlated Zernike coefficients,” in *Proc. IEEE International Conf. Computational Photography*, 2020. 5, 6, 7, 8, 9, 13, 14
- [30] S. H. Chan, X. Wang, and O. A. Elgendy, “Plug-and-play ADMM for image restoration: Fixed-point convergence and applications,” *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 84–98, 2017. 6, 7
- [31] R. Ahmad, C. A. Bouman, G. T. Buzzard, S. H. Chan, E. T. Reehorst, and P. Schniter, “Plug-and-play methods for magnetic resonance imaging,” *IEEE Signal Processing Magazine*, vol. 37, no. 1, pp. 105–116, Jan. 2020. 6
- [32] G. Buzzard, S. H. Chan, S. Sreehari, and C. A. Bouman, “Plug-and-play unplugged: optimization free reconstruction using consensus equilibrium,” *SIAM Imaging Science*, vol. 11, no. 3, pp. 2001–2020, Sep. 2018. 6
- [33] S. H. Chan, “Performance analysis of Plug-and-Play ADMM: A graph signal processing perspective,” *IEEE Transactions on Computational Imaging*, vol. 5, no. 2, pp. 274–286, Jun. 2019. 6
- [34] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 2.1,” <http://cvxr.com/cvx>, Mar. 2014. 7
- [35] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan 2011. 7
- [36] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition,” in *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, Nov 1993, pp. 40–44 vol.1. 8
- [37] M. Hirsch, S. Sra, B. Schölkopf, and S. Harmeling, “Efficient filter flow for space-variant multiframe blind deconvolution,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2010, pp. 607–614. 8, 9, 12
- [38] K. R. Leonard, J. Howe, and D. E. Oxford, “Simulation of atmospheric

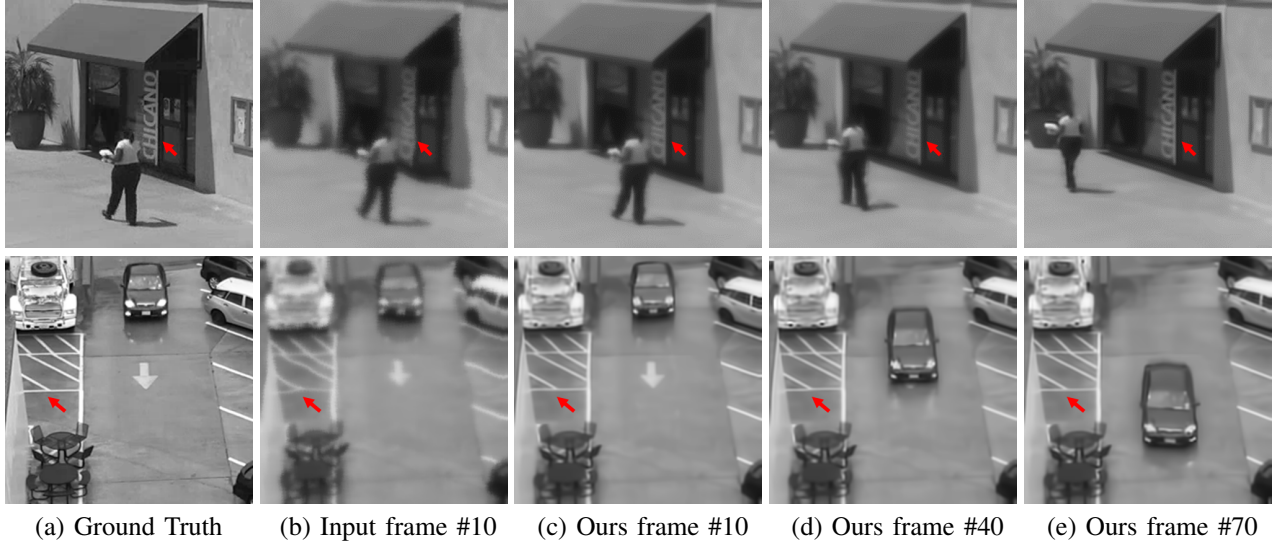


Fig. 21: **Additional results on synthetic dynamic sequences.** The video sequences are synthetically generated from clean videos from [43] using simulator presented in [29]. Note the consistent static background and well-maintained moving objects across frames.

turbulence effects and mitigation algorithms on stand-off automatic facial recognition,” in *Proc. SPIE 8546, Optics and Photonics for Counterterrorism, Crime Fighting, and Defence VIII*, Oct. 2012, pp. 1–18. 8, 11

- [39] Z. Lin, M. Chen, and Y. Ma, “The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices,” *ArXiv*, vol. abs/1009.5055, 2013. 9, 10
- [40] A. Chakrabarti, “A neural approach to blind motion deblurring,” in *European Conference on Computer Vision*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 221–235. 10
- [41] X. Xu, J. Pan, Y. Zhang, and M. Yang, “Motion blur kernel estimation via deep learning,” *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 194–205, Jan. 2018. 10
- [42] CopenhagenSensorTech, “Heat haze reduction demonstration,” [https://www.youtube.com/watch?v=ePPa\\_l9m5CA](https://www.youtube.com/watch?v=ePPa_l9m5CA), June 2016. 13
- [43] S. Oh, A. Hoogs, A. Perera, N. Cuntoor, C. Chen, J. T. Lee, S. Mukherjee, J. K. Aggarwal, H. Lee, L. Davis, E. Swears, X. Wang, Q. Ji, K. Reddy, M. Shah, C. Vondrick, H. Pirsiavash, D. Ramanan, J. Yuen, A. Torralba, B. Song, A. Fong, A. Roy-Chowdhury, and M. Desai, “A large-scale benchmark dataset for event recognition in surveillance video,” in *CVPR 2011*, June 2011, pp. 3153–3160. 14