Long-range Turbulence Mitigation: A Large-scale Dataset and A Coarse-to-fine Framework
Supplementary Material

Shengqi Xu, Run Sun, Yi Chang ⋆, Shuning Cao, Xueyao Xiao, and Luxin Yan

National Key Lab of Multispectral Information Intelligent Processing Technology,
Huazhong University of Science and Technology, China

1 Overview

The supplementary material is organized as follows:

(**Section 2**) We conduct visual comparisons and analysis on existing public real-world atmospheric turbulence benchmarks.

(**Section 3**) We provide additional analysis and details about the proposed Coarse-to-fine framework.

- We provide more analysis of pixel displacements statistical prior of turbulence in more long-range turbulence scenes in Sec. 3.1.
- We introduce the solution of low-rank tensor refinement model in Sec. 3.2.
- We introduce the details of data-driven blur removal process in Sec. 3.3.

(**Section 4**) We conduct further discussions on the effectiveness and performance of the proposed framework CDSP and frequency-aware reference frame.

- We study the effectiveness of CDSP towards severe distortions in Sec. 4.1.
- We conduct an user study to compare existing methods with CDSP on proposed dataset RLR-AT in Sec. 4.2.
- We discuss the robustness of proposed CDSP to long-range turbulence at various distances in Sec. 4.3.
- We analyze the frequency-aware weight parameter σ in Sec. 4.4.
- We validate the robustness of proposed FRF to noises in Sec. 4.5.
- We show more visual comparison to demonstrate the high-quality of proposed FRF in Sec. 4.6.
- We show more visual results of hot-air turbulence generalization in Sec. 4.7.

(**Section 5**) We provide more visual comparisons on proposed real dataset RLR-AT and synthetic turbulence to demonstrate the effectiveness of CDSP.

(**Demo**) We provide a video demo to show our CDSP's potential towards long-range turbulence with severe distortions.
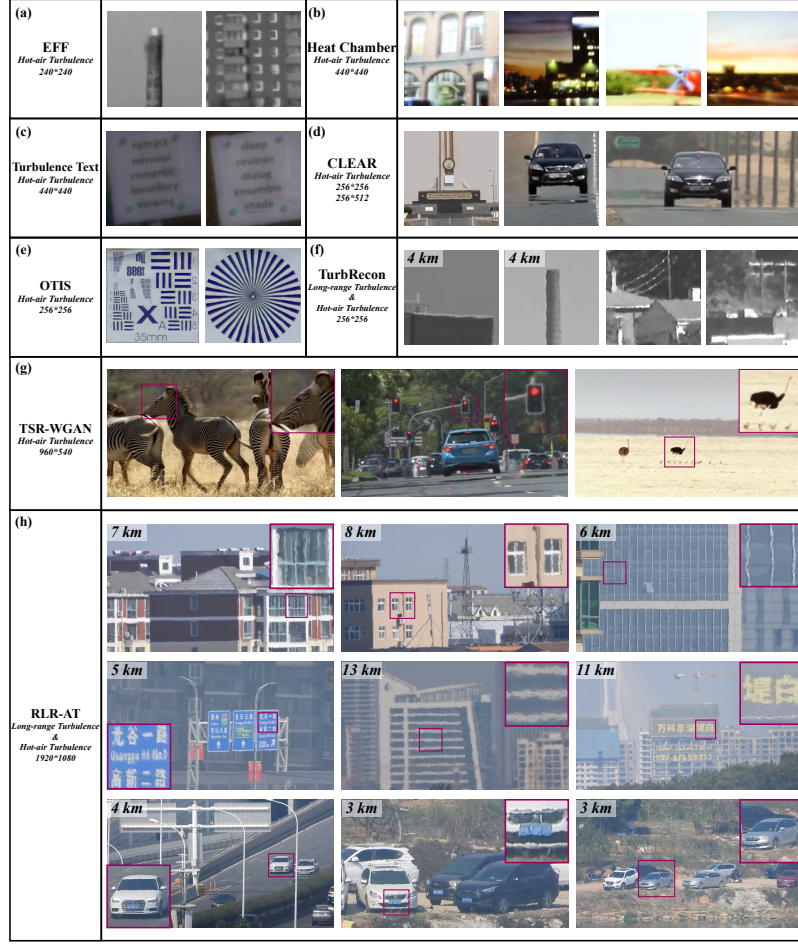

⋆ Corresponding author

Fig. 1: Benchmark analysis of existing public real-world turbulence datasets.

2   Benchmark Analysis

Existing Datasets. We make a brief visual comparison between existing datasets and our dataset RLR-AT in Fig. 1. In Fig. 1(a), EFF [9] provided two widely used hot-air turbulence data: *Chimney* and *Building*. Further, Heat chamber [16] captured hot-air turbulence by imaging the display screen located 20 meters away, and heat chambers are placed between the screen and the camera for generating turbulence effect. These hot-air turbulence datasets were collected with artificial heat burners. Furthermore, OTIS [7], CLEAR [2], Turbulence Text [16] and TSR-WGAN [10] further collected the hot-air turbulence sequences in high-temperature weather conditions near the earth surface. Overall, most existing benchmarks consist of hot-air turbulence, and TurbRecon [14] stands out by capturing 2 turbulence sequences for Building scene at a distance of 4 Km, as
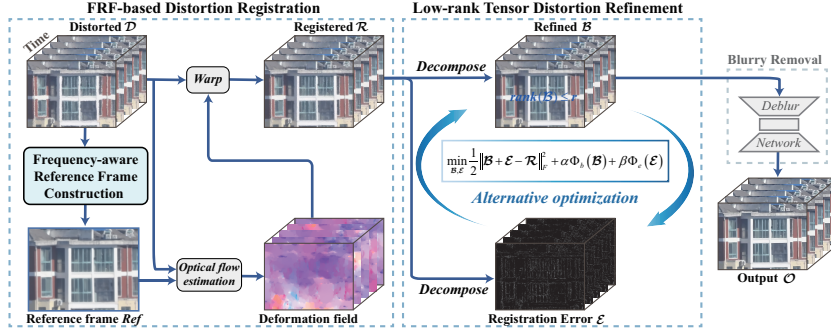
Fig. 2: Illustration of the proposed framework.

shown in Fig. 1(f). Moreover, most existing datasets are not only constrained in terms of distances but also limited in the number and diversity of scenes, as well as the resolution size of images.

RLR-AT. To solve the limitations of existing datasets and advance this field, we construct a large-scale real-world long-range atmospheric turbulence dataset, RLR-AT. Typical examples are illustrated in Fig. 1(h). The advantages of RLR-AT compared to existing datasets are threefold. Firstly, RLR-AT contains **turbulence with longer-distance and higher-diversity**, covering diverse distortions spanning imaging distances from 1 Km to 13 Km. Secondly, it consists of **larger-scale and greater-variety scenes**, including 1500 turbulence sequences collected across various scenarios, such as text, object, building, etc. Last but not least, RLR-AT is collected by a telephoto camera with **high-resolution (1920*1080 pixels)**. We believe that RLR-AT can serve as a benchmark for future works targeting long-range turbulence mitigation.

3   Method Details

We propose a Coarse-to-fine framework that cooperates dynamic turbulence and static background priors (CDSP) to handle long-range turbulence with severe distortions, as shown in Fig. 2. Specifically, we discover the pixel motion statistical prior of turbulence and propose a frequency-aware reference frame for better registration, which greatly reduces the burden of refinement. Then we align the distorted frames to the proposed reference frame utilizing registration approach based on optical flow [13]. Moreover, we take advantage of the static prior of background and propose a subspace-based low-rank tensor refinement model to refine the registration errors unavoidably left by registration while well preserving details. Finally, we employ a simple data-driven network to further remove the residual blur, and the generation of paired deblurring data is based on the proposed distortion correction framework. The details of SLRTR solution and blur removal process are provided in the following sub-sections, and we also further analyze the pixel motion statistical prior of turbulence in various long-range scenes.
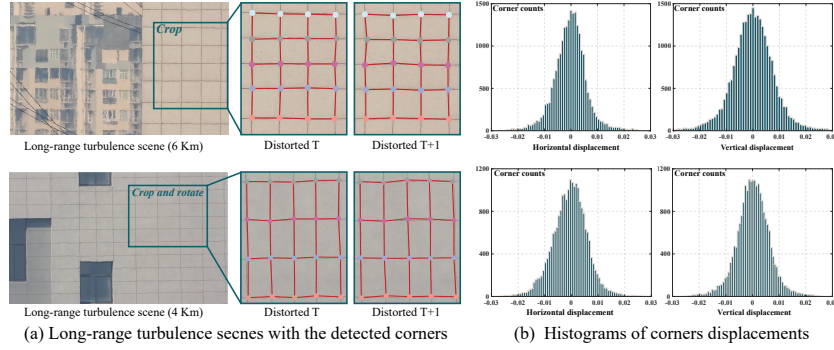
(a) Long-range turbulence secnes with the detected corners   (b) Histograms of corners displacements

Fig. 3: Analysis of pixel motion statistical prior in long-range turbulence scenes. (a) Visual examples of long-range turbulence building scenes with the detected corners. (b) Statistic histograms of corners displacements. It is observed that corners motion in long-range turbulence scenes still approximately follow a zero-mean gaussian distribution, indicating that corners most frequently occur in their original positions.

3.1   Analysis of Pixel Motion Statistics Prior in Long-range Scenes

To further explore the pixel motion statistical prior of turbulence in long-range scenes, we conduct statistical analysis experiments in numerous long-range turbulence scenes (*e.g. Building scenes with wall corners*) as well. Typical visual examples are shown in Fig. 3. We choose the building scenes with wall corners as experiment data and utilize SURF corners detector [3] to detect the corners. Figure 3(a) visualizes the cropped local regions with detected corners. The corners of distorted frames exhibit noticeable motion. The histograms of corner motions are shown in Fig. 3(b). It is observed that corners motion in long-range turbulence scenes still approximately follow a zero-mean gaussian distribution, indicating that corners most frequently occur in their original positions.

3.2   Solution of the Distortion Refinement SLRTR

After aligning the distorted frames to the proposed FRF using the non-grid registration based on optical flow [13], we employ SLRTR to refine the unavoidable registration error. The original optimization problem of SLRTR is shown as follow:

$$
\left\{ \hat{\boldsymbol{\mathcal{B}}}, \hat{\boldsymbol{\mathcal{E}}}, \hat{\boldsymbol{\mathcal{G}}}_i, \hat{O}_i \right\} = \arg\min_{\boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{E}}, \boldsymbol{\mathcal{G}}_i, O_i} \frac{1}{2} ||\boldsymbol{\mathcal{B}} + \boldsymbol{\mathcal{E}} - \boldsymbol{\mathcal{R}}||_F^2
$$
$$
+ \beta ||\boldsymbol{\mathcal{E}}||_1 + \alpha \sum_i \left(\frac{1}{\lambda_i^2} ||\boldsymbol{\mathcal{S}}_i \boldsymbol{\mathcal{B}} \times_3 O_i - \boldsymbol{\mathcal{G}}_i||_F^2 + ||\boldsymbol{\mathcal{G}}_i||_{tnn} \right),
\tag{1}
$$

where $\boldsymbol{\mathcal{R}} \in \mathbb{R}^{h \times w \times t}$ is the registered sequence, $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{h \times w \times t}$ is the undistorted sequence, $\boldsymbol{\mathcal{E}} \in \mathbb{R}^{h \times w \times t}$ represents the registered error, $\boldsymbol{\mathcal{G}}_i$ represents the low-rank approximation, $O_i \in \mathbb{R}^{d \times t}(d \ll t)$ is an orthogonal subspace projection matrix used to capture the temporal low-rank property and $\boldsymbol{\mathcal{S}}_i \boldsymbol{\mathcal{B}} \in \mathbb{R}^{p^2 \times k \times t}$ is the constructed 3-D tensor via the non-local clustering of a sub-cubic $u_i \in \mathbb{R}^{p \times p \times t}$ [5].

Optimization. To solve this difficult problem, we apply the ADMM [12] to decouple the Eq. (1) into several sub-problems.

1) *Registered Error Estimation* $\boldsymbol{\mathcal{E}}$: Fixed other variables, we can get following subproblem:

$$\hat{\boldsymbol{\mathcal{E}}} = \arg\min_{\boldsymbol{\mathcal{E}}} \beta \parallel \boldsymbol{\mathcal{E}} \parallel_1 + \frac{1}{2} \parallel \boldsymbol{\mathcal{B}} + \boldsymbol{\mathcal{E}} - \boldsymbol{\mathcal{R}} \parallel_F^2 . \tag{2}$$

The Eq. (2) can be solved by introducing the soft-thresholding operator [12]:

$$\Psi_v(x) = \begin{cases} x - v, & if \quad x > v \\ x + v, & if \quad x < -v \\ 0, & otherwise \end{cases} \tag{3}$$

Then, we can easily get its closed-form solution:

$$\hat{\boldsymbol{\mathcal{E}}} = \Psi_\beta \left(-\boldsymbol{\mathcal{B}} + \boldsymbol{\mathcal{R}} \right) . \tag{4}$$

2) *Subspace Projection* O_i: We enforce the orthogonal constraint on $O_i^T O_i = I$ with the following subproblem:

$$\hat{O}_i = \arg\min_{O_i^T O_i = I} \frac{1}{\lambda_i^2} ||\boldsymbol{\mathcal{S}}_i\boldsymbol{\mathcal{B}} \times_3 O_i - \boldsymbol{\mathcal{G}}_i||_F^2, \tag{5}$$

According to [18], we perform a model-3 unfolding and singular value decomposition on $\boldsymbol{\mathcal{S}}_i\boldsymbol{\mathcal{B}}$, that is $(\boldsymbol{\mathcal{S}}_i\boldsymbol{\mathcal{B}})^{(3)} = USV^T$. Then let the $\hat{O}_i = (U(:, 1 : d))^T$, where $U(:, 1 : d)$ means to select the first d left singular vectors corresponding to the d largest singular values, where the d is the measurement of the intrinsic subspace of the temporal dimension, we empirically set ($d \leq 4$).

3) *Low-rank Approximation* $\boldsymbol{\mathcal{G}}_i$: Dropping the irrelevant variables, we can get following subproblem:

$$\hat{\boldsymbol{\mathcal{G}}}_i = \arg\min_{\boldsymbol{\mathcal{G}}_i} \frac{1}{\lambda_i^2} \parallel \boldsymbol{\mathcal{S}}_i\boldsymbol{\mathcal{B}} \times_3 O_i - \boldsymbol{\mathcal{G}}_i \parallel_F^2 + ||\boldsymbol{\mathcal{G}}_i||_{tnn}, \tag{6}$$

where λ_i^2 is the noise variance, $\parallel \boldsymbol{\mathcal{G}}_i \parallel_{tnn} = \sum_p \left| \sigma_p(\boldsymbol{\mathcal{G}}_i^{(2)}) \right|_1$ is the tensor nuclear norm, $\boldsymbol{\mathcal{G}}_i^{(2)}$ denotes the model-2 unfolding matrix of the $\boldsymbol{\mathcal{G}}_i$, and $\sigma_p(\boldsymbol{\mathcal{G}}_i^{(2)})$ represents the p-th singular value of the $\boldsymbol{\mathcal{G}}_i^{(2)}$. This minimization problem is usually solved by the singular value thresholding algorithm [4,5], we adopt the weighted nuclear norm minimization to boost the performance [8]. Thus, we can get the following formula:

$$\hat{\boldsymbol{\mathcal{G}}}_i = U\Theta_W(\boldsymbol{\Sigma})V^T, \tag{7}$$

where $U\boldsymbol{\Sigma}V^T$ is the singular value decomposition of the mode-2 unfolding matrix of the $\boldsymbol{\mathcal{S}}_i\boldsymbol{\mathcal{B}} \times_3 O_i$, and $\Theta_W(\boldsymbol{\Sigma})_{pp} = \max(\boldsymbol{\Sigma}_{pp} - W_{pp}, 0)$ is the generalized soft-thresholding operator with the weight vector W. After we obtain the $\boldsymbol{\mathcal{G}}_i$, we conduct the tensor folding to transform it into the 3D tensor $\boldsymbol{\mathcal{G}}_i$.

4) *Undistorted Sequence Estimation* $\boldsymbol{\mathcal{B}}$: We fix the other variables and optimize $\boldsymbol{\mathcal{B}}$ with the following subproblem:

$$\min_{\boldsymbol{\mathcal{B}}} \frac{1}{2}||\boldsymbol{\mathcal{B}} + \boldsymbol{\mathcal{E}} - \boldsymbol{\mathcal{R}}||_F^2 + \alpha \sum_i \frac{1}{\lambda_i^2}||\boldsymbol{\mathcal{S}}_i\boldsymbol{\mathcal{B}} \times_3 O_i - \boldsymbol{\mathcal{G}}_i||_F^2, \tag{8}$$

where the first term is the fidelity term, and the second term is the low-rank constraint. The linear operator $\boldsymbol{\mathcal{S}}$ is used to extract the cubic from the video and

it can be fast solved on a pixel-by-pixel manner. However, the Eq. (8) is difficult to solve directly. To solve this problem, we introduce the auxiliary variable $\mathcal{D} = \mathcal{B}$ to spilt the three terms, and the Eq. (8) can be transformed into the following augmented Lagrangian function:

$$L_\mu(\mathcal{B}, \mathcal{D}) = \frac{1}{2} \parallel \mathcal{B} + \mathcal{E} - \mathcal{R} \parallel_F^2 + \frac{\mu}{2} \parallel \mathcal{D} - \mathcal{B} - \frac{\boldsymbol{\Gamma}}{\mu} \parallel_F^2$$

$$+ \alpha \left(\sum_i \frac{1}{\lambda_i^2} \parallel \mathcal{S}_i \mathcal{D} \times_3 O_i - \mathcal{G}_i \parallel_F^2 + \parallel \mathcal{G}_i \parallel_{tnn} \right), \tag{9}$$

where $\boldsymbol{\Gamma}$ is the Lagrangian multipliers, and μ is a positive scalar. And the Eq. (9) can be separated into several sub-problems:

$$\mathcal{B}^{m+1} = \arg\min_{\mathcal{B}} \frac{1}{2} \parallel \mathcal{B} + \mathcal{E} - \mathcal{R} \parallel_F^2 + \frac{\mu}{2} \parallel \mathcal{D}^m - \mathcal{B} - \frac{\boldsymbol{\Gamma}^m}{\mu} \parallel_F^2, \tag{10}$$

$$\mathcal{D}^{m+1} = \arg\min_{\mathcal{D}} \alpha \sum_i \frac{1}{\lambda_i^2} \parallel \mathcal{S}_i \mathcal{D} \times_3 O_i - \mathcal{G}_i \parallel_F^2 + \frac{\mu}{2} \parallel \mathcal{D} - \mathcal{B}^{m+1} - \frac{\boldsymbol{\Gamma}^m}{\mu} \parallel_F^2, \tag{11}$$

We adopt the fast 3-D Fourier transform for fast calculating the clean video $\mathcal{B}$, the following formula is its closed-form solution.

$$\mathcal{B}^{m+1} = \mathcal{F}^{-1} \left(\frac{\mathcal{F}(\mathcal{R} - \mathcal{E} + \mu(\mathcal{D}^m - \frac{\boldsymbol{\Gamma}^m}{\mu}))}{\mathcal{I} + \mu\mathcal{I}} \right), \tag{12}$$

where $\mathcal{F}$ is the fast 3-D Fourier transform, $\mathcal{F}^{-1}$ is its inverse transform, and $\mathcal{F}^*$ represents its conjugate transform.

As for the auxiliary variable $\mathcal{D}$, we have the following formula:

$$\mathcal{D}^{m+1} = \left(\frac{2\alpha}{\lambda_i^2} \sum_i (\mathcal{S}_i)^T \mathcal{S}_i + \mu\mathcal{I} \right)^{-1}$$

$$\times \left(\frac{2\alpha}{\lambda_i^2} \sum_i (\mathcal{S}_i)^T \mathcal{G}_i \times_3 (O_i)^T + \mu\mathcal{B}^{m+1} + \boldsymbol{\Gamma}^m \right). \tag{13}$$

where $(\mathcal{S}_i)^T \mathcal{G}_i$ is the sum value of all overlapping reconstruction cubics that cover the pixel location, and $(\mathcal{S}_i)^T \mathcal{S}_i$ corresponds to the number of the overlapping cubics. Then the Eq. (13) can be easily solved by the pixel-to-pixel division.

For the Lagrangian multipliers, we update them with the following formulas:

$$\boldsymbol{\Gamma}^{m+1} = \boldsymbol{\Gamma}^m + \mu \left(\mathcal{B}^{m+1} - \mathcal{D}^{m+1} \right). \tag{14}$$

The overall procedure is summarized in **Algo.** 1.

3.3   Data-driven Blur Removal

In this work, after correcting the severe distortions caused by long range turbulence utilizing the proposed CDSP, we further employ a data-driven method to remove the residual blurry in the undistorted image as shown in Fig. 4. To achieve this, we adopt the Uformer [17] as our backbone, due to its excellent performance in image restoration. In addition, we further utilize the turbulence simulator P2S [15] and CDSP to construct the paired blurry training data.
CDSP-based Paired Blurry Data Generation. We've already corrected the severe distortions, so we only need to construct the suitable paired blurry

---

Algorithm 1 Low-rank tensor distortion refinement model

Require: The Registered sequence $\mathcal{R}$
 1: **Initialization:**;
 2:   • Set the regularization parameters β, α, λ_i;
 3:   • Set the temporal subspace dimension $d \leq 4$;
 4: **for** $l=1{:}L$ **do**
 5:     Group similar cubics $\mathcal{S}_i\mathcal{B}^{l+1}$;
 6:     Estimate registration error $\mathcal{E}^{l+1}$ via Eq. (4);
 7:     Estimate subspace projection O_i^{l+1} via Eq. (5);
 8:     Estimate low-rank approximation $\mathcal{G}_i^{l+1}$ via Eq. (7);
 9:     **for** $m=1{:}M$ **do**
10:       Estimate undistorted video $\mathcal{B}^{m+1}$ via Eq. (12);
11:       Estimate auxiliary variable, $\mathcal{D}^{m+1}$ via Eq. (13);
12:       Update lagrangian multipliers Γ^{m+1} via Eq. (15);
13:     **end for**
14: **end for**
Ensure: The undistorted sequence $\mathcal{B}$.

Table 1: The protocol of the simulator for generating paired blurry training data.

Strength	Probability	D(m)	D/r_0	Distance(m)	$Corr$	Kernel Size
Weak	0.3	U(0.001, 0.005)	[0.4, 0.8, 1.2, 1.5]	U(800,1500)	[-1, -0.1, -0.5, -0.05]	33
Medium	0.5	U(0.04, 0.1)	[0.8 , 1, 1.6]	U(2000,3000)	[-1, -0.1, -0.5, -0.05]	33
Strong	0.2	U(0.1, 0.2)	[1.6, 2, 2.4]	U(4000,5000)	[-1, -0.1, -0.5, -0.05]	33

training data. As shown in Fig. 4, we first employ the turbulence simulator to generate the synthetic turbulence. Next, we apply the proposed CDSP to correct the distortion in the synthetic turbulence, obtaining undistorted image that can be paired with clean image to create the training data for blurry removal.

Description of Training Dataset. We randomly select 60000 images in the place dataset [19] as input of the simulator. For the configuration of turbulence simulator, we set turbulence strength in 3 levels: weak, medium and strong. The protocol of the simulator is shown in Table 1. From top to bottom, the turbulence strength becomes higher. D denotes the aperture diameter of the sensor, D/r_0 denotes the ratio of D and the fried parameter r_0, $Corr$ denotes the spatial correlation. U(a, b) denotes uniform distribution in the range(a, b) and [a, b, c, d] denotes random choice with equal probability.

Implementation Details. We train the network using the Charbonnier loss [6] supervised by the ground-truth image:

$$\mathcal{L}(\hat{I}, I) = \sqrt{\| \hat{I} - I \|^2 + \epsilon^2}, \tag{15}$$

where $\hat{I}$ is the clean image and I is the output image, and ϵ is empirically set to 10^{-3}. The network is trained with two RTX 3090 GPUs. The images are
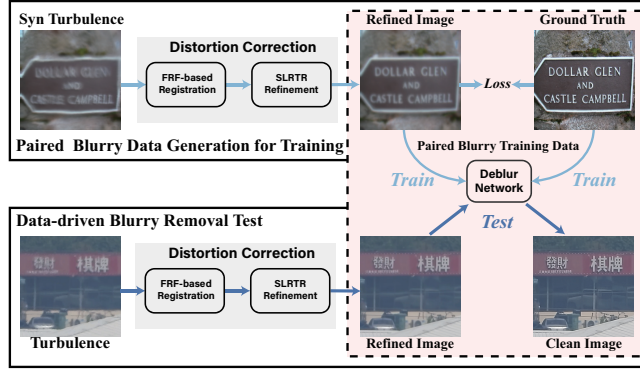
Fig. 4: Paired blurry data generation for training and data-driven blurry removal test process. We apply the proposed CDSP to correct the distortion of synthetic turbulence images, generating undistorted image that could be paired with the ground-truth image to construct the training dataset for blurry removal. The network trained with this generated paired dataset could effectively remove the residual blur from the refined image, producing a clean image.
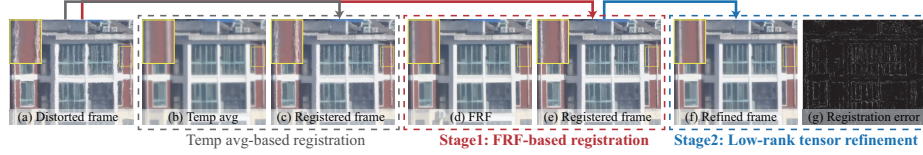

Fig. 5: Visualization of each stage. (a) Distorted frame. (b) Baseline of reference frame: Temporal average (Temp avg). (c) Registered frame obtained utilizing Temp avg. (d) Frequency-aware reference frame obtained based on dynamic turbulence prior. (e) Registered frame obtained utilizing the proposed FRF in stage 1. (f) Refined frame via low-rank tensor distortion refinement based on static background prior. (g) Estimated registration error in stage2.

randomly cropped into 128*128 for training. The learning rate is set as 0.0002. The Adam optimizer is adopted for optimization with a batch size of 32.

4   Analysis and Discussion

4.1   Effectiveness of Coarse-to-fine Framework towards Severe Distortions

We illustrate the effectiveness of our Coarse-to-fine framework in Fig. 5. Our framework jointly employs both dynamic turbulence and static priors, where dynamic and static priors complement to each other to better remove the severe distortions. On the one hand, we propose a frequency-aware reference frame (FRF) based on pixel motion statistical prior of turbulence. Compared to existing baseline: Temp avg, FRF possesses more superior visual quality and sharper
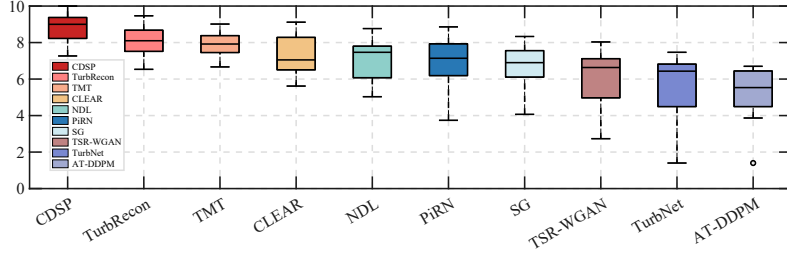
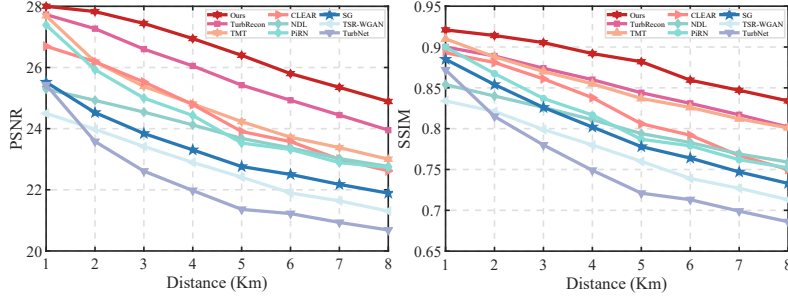**Fig. 6:** The user study on real long-range turbulence mitigation on RLR-AT.

Fig. 7: Comparison of long-range turbulence mitigation at different distances.

edge, as shown in Fig. 5(b) and (d). Therefore, FRF-based registration ensures greater accuracy, thereby reducing registration errors in the registered frame and greatly reducing the burden of refinement, as shown in Fig. 5(c) and (e). On the other hand, we further utilize the low-rank prior of static background, and propose a subspace-based low-rank tensor refinement model to eliminate the registration errors inevitably left by registration, as shown in Fig. 5(f) and (g). The correction result of Fig. 5(f) not only avoids severe distortions but also well preserves details, implying the effectiveness of the proposed Coarse-to-fine framework towards long-range turbulence with severe distortions.

4.2 User Study Conducted on RLR-AT

To further compare CDSP with other methods, we conduct a user study on the RLR-AT. We randomly choose 50 real turbulence images and the corresponding restoration results produced by each method. Then we invite 20 people to score each result with the range of 1-10, higher score represents better restoration quality. The distribution of scores for each method is shown in Fig. 6. It is observed that CDSP achieves the highest median score, surpassing 8 in its first quartile, indicating that users tend to rank our results first. The user study results again imply that the proposed method is superior to other methods.
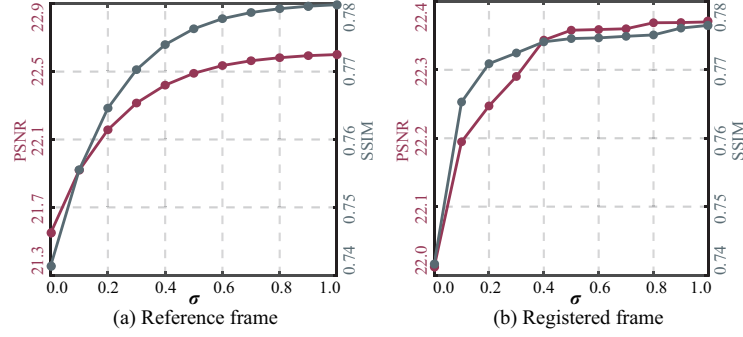
Fig. 8: Analysis of frequency-aware weight parameter σ on the synthetic turbulence at 5Km distance. (a) and (b) show the Quantitative results of reference frame constructed with different numbers of σ and corresponding registered frames.
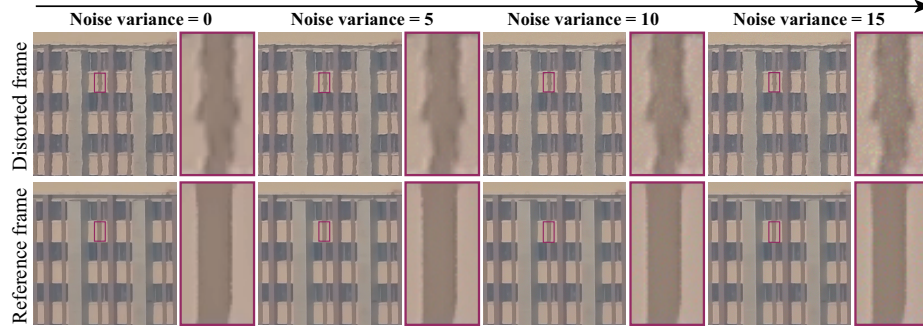

Fig. 9: Robustness of frequency-aware reference frame construction to noises.

4.3   Robustness towards Different Distances

We further analyze the performance of CDSP when handling different distances of turbulence, presenting the quantitative results in Fig. 7. We can observe that our results outperform existing methods by a large margin under each distance. Moreover, the visualization results in Fig. 13 shows that the CDSP can consistently improve image quality for different distances. Both the quantitative and qualitative results verify the robustness of the CDSP towards different distances.

4.4   Analysis of Frequency-aware Weight Parameter σ

The parameter σ determines the sensitivity of weight function to frequency. In Fig. 8, we show the quantitative results of reference frame with different numbers of σ and their corresponding registered frames. It is evident that as σ increases, the quality of results improves gradually before reaching a stable state. In this work, we empirically set σ to 1.0.

Turbulence Image   Temp Avg [22]   Non-local Avg [15]   Lau[11]   FRF (Ours)
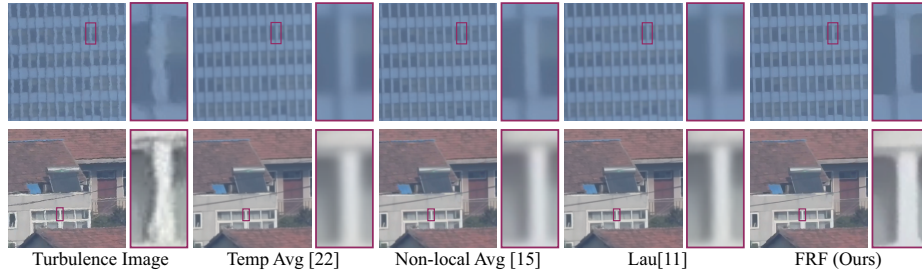
Fig. 10: Visual comparison of FRF with other reference frames.

4.5   Robustness of Frequency-aware Reference Frame to Noises

We construct the frequency-aware reference frame under various levels of noises, as shown in Fig. 9. It is observed that the proposed FRF sill maintains accurate structure and sharp edge even as the noise increases, demonstrating the robustness of proposed frequency-aware reference frame to noises.
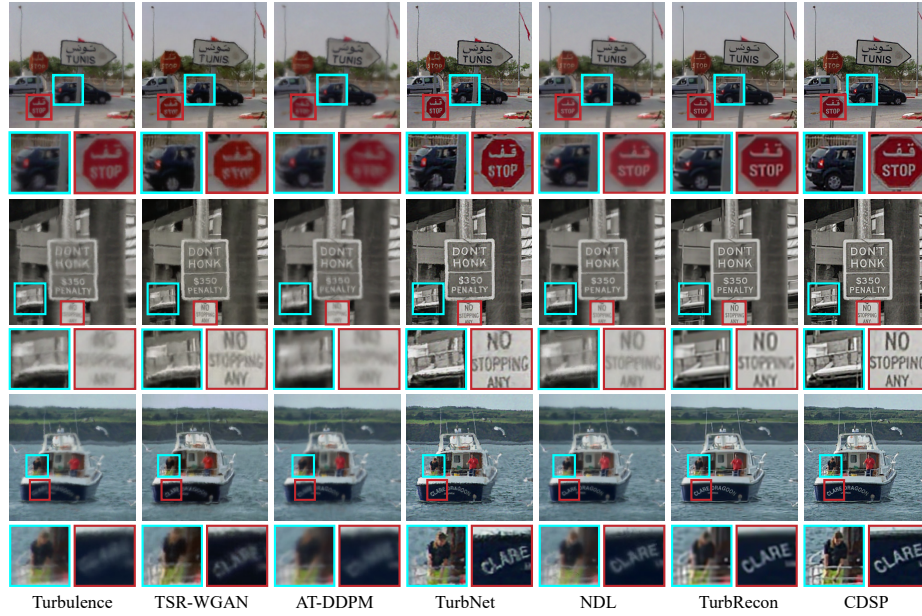

Turbulence   TSR-WGAN   AT-DDPM   TurbNet   NDL   TurbRecon   CDSP

Fig. 11: Visual comparisons on real hot-air turbulence. The proposed method generalizes well for hot-air turbulence which is collected with heat chamber.

Table 2: The protocol of simulator for synthetic turbulence comparison experiment.

Parameter	Value
Distance	$L = 1Km - 8Km$
Fried Parameter	$r_0 = 0.041m - 0.145m$
Aperture Diameter	$D = 0.077m$
Aperture-to-Coherence Ratio	$D/r_0 = 0.53 - 1.85$
Correlation Strength for PSF	$Corr = -0.01$
Image Size	256×256 pixels
Wavelength	$\lambda = 500nm$
Zernike Phase Size	16×16 pixels
Kernel Size	33×33 pixels

4.6   High-quality of Frequency-aware Reference Frame

To further demonstrate the high-quality of the proposed FRF, we compare the existing reference frames (Temp Avg [21], Non-local avg [14], Lau [11]) and proposed FRF in Fig. 10. We can observe that existing reference frames suffer from severe blur, while FRF possesses more favorable image quality and sharper edge, which is beneficial for large scale distortion registration.

4.7   Generalization to Hot-air Turbulence

We further compare with state-of-the-art methods on hot-air turbulence, using UG2+ TurbuText [1]. As shown in Fig. 11, existing methods exhibit either severe geometric distortion or significant loss of image detail. Compared with these methods, CDSP effectively corrects most geometric distortions meanwhile well preserving image details, which demonstrates that the proposed method could effectively handle hot-air turbulence as well.

5   Comparison Experiments

5.1   Configuration of Simulator

The ground-truth images for the synthetic turbulence comparison experiment are randomly selected from ADE20K [20] and differ from the dataset used to generate paired blurry data. We simulate the synthetic turbulence sequences at different distances applying the turbulence simulator P2S. Each strength of synthetic turbulence contains 100 distorted sequences with 100 frames. The details of the simulator parameter are provided in Table 2 and $C_n^2 = 1.0 \times 10^{-15} m^{-2/3}$.

5.2   Qualitative Comparisons on Synthetic Turbulence

Figure 12 shows the visual results on synthetic turbulence. For the supervised methods, they cannot well correct the distortions due to the domain gap. The
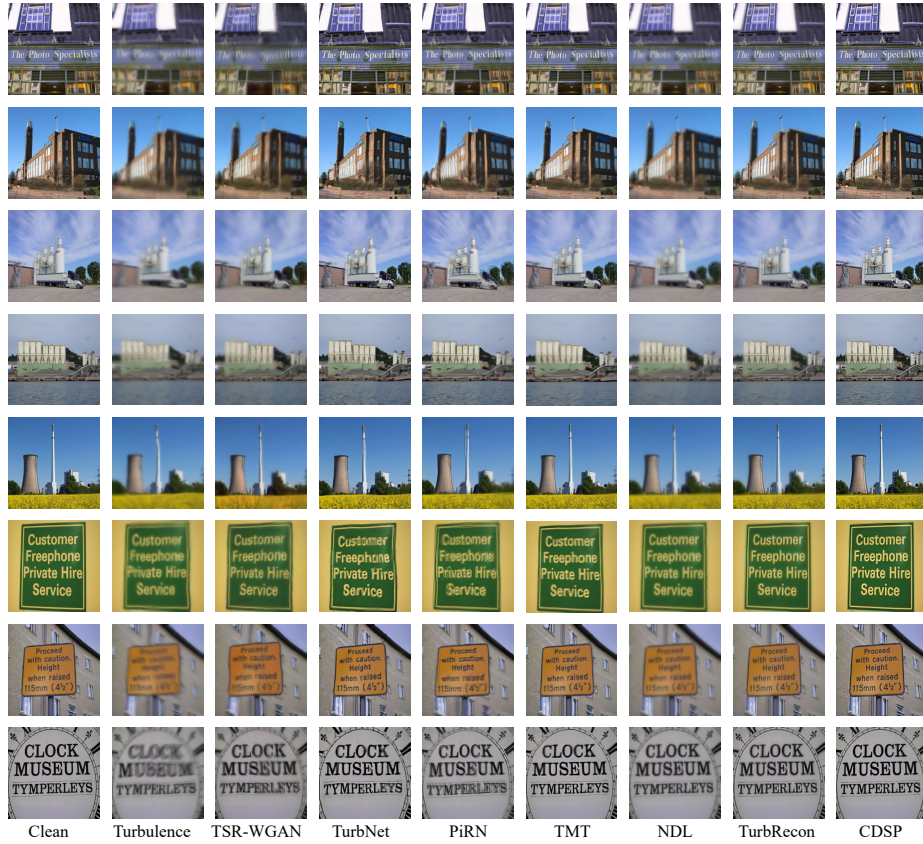
| Clean | Turbulence | TSR-WGAN | TurbNet | PiRN | TMT | NDL | TurbRecon | CDSP |

Fig. 12: Visualization comparisons on synthetic turbulence.

model-based methods NDL and SG are limited to handle severe pixel displacement. The wavelet-based CLEAR generates artifacts which degrade the visualization quality in restoration results. Compared with these methods, CDSP could acquire images with more favorable visual quality and structure, closing to corresponding ground-truth image.

5.3   Qualitative Comparisons on Real Turbulence

We evaluate the performance of the proposed CDSP and other state-of-the-art methods on RLR-AT that contains severe long-range turbulence. The turbulence removal results visualization are shown in Fig. 13. We can observe that the supervised methods have difficulty in dealing with realistic turbulence and leave the distortions unremoved, due to the huge gap between the simplified synthetic turbulence and the complex real turbulence. The artifacts in CLEAR degrade restoration images. NDL continually suffers from pixel displacement due to unsuitable design for long-range turbulence. Albeit TurbRecon can acquire results

Fig. 13: Visual comparison on real long-range turbulence in proposed RLR-AT.

with comparable quality, the results still encounter details loss and artifacts due to the severe pixel displacement. In comparison, CDSP possesses not only higher image quality but also more satisfying distortion correction results thanks to the cooperative dynamic turbulence and static background priors.

References

1. Bridging the gap between computational photography and visual recognition: 6th ug2+ prize challenge, `http://http://cvpr2023.ug2challenge.org/dataset23_t2.html`, track 2
2. Anantrasirichai, N., Achim, A., Kingsbury, N.G., Bull, D.R.: Atmospheric turbulence mitigation using complex wavelet-based fusion. IEEE TIP **22**(6), 2398–2408 (2013)
3. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: ECCV. pp. 404–417. Springer (2006)
4. Cai, J.F., Candès, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. SIAM Journal on optimization **20**(4), 1956–1982 (2010)

5. Chang, Y., Yan, L., Zhong, S.: Hyper-laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising. In: CVPR. pp. 4260–4268 (2017)
6. Charbonnier, P., Blanc-Feraud, L., Aubert, G., Barlaud, M.: Two deterministic half-quadratic regularization algorithms for computed imaging. In: ICIP. vol. 2, pp. 168–172. IEEE (1994)
7. Gilles, J., Ferrante, N.B.: Open turbulent image set (otis). PRL **86**, 38–41 (2017)
8. Gu, S., Zhang, L., Zuo, W., Feng, X.: Weighted nuclear norm minimization with application to image denoising. In: CVPR. pp. 2862–2869 (2014)
9. Hirsch, M., Sra, S., Schölkopf, B., Harmeling, S.: Efficient filter flow for space-variant multiframe blind deconvolution. In: CVPR. pp. 607–614 (2010)
10. Jin, D., Chen, Y., Lu, Y., Chen, J., Wang, P., Liu, Z., Guo, S., Bai, X.: Neutralizing the impact of atmospheric turbulence on complex scene imaging via deep learning. NMI **3**(10), 876–884 (2021)
11. Lau, C.P., Lai, Y.H., Lui, L.M.: Restoration of atmospheric turbulence-distorted images via rpca and quasiconformal maps. Inverse Problems **35**(7), 074002 (2019)
12. Lin, Z., Liu, R., Su, Z.: Linearized alternating direction method with adaptive penalty for low-rank representation. NeurIPS **24** (2011)
13. Liu, C., et al.: Beyond pixels: exploring new representations and applications for motion analysis. Ph.D. thesis, Massachusetts Institute of Technology (2009)
14. Mao, Z., Chimitt, N., Chan, S.H.: Image reconstruction of static and dynamic scenes through anisoplanatic turbulence. IEEE TCI **6**, 1415–1428 (2020)
15. Mao, Z., Chimitt, N., Chan, S.H.: Accelerating atmospheric turbulence simulation via learned phase-to-space transform. In: ICCV. pp. 14759–14768 (2021)
16. Mao, Z., Jaiswal, A., Wang, Z., Chan, S.H.: Single frame atmospheric turbulence mitigation: A benchmark study and a new physics-inspired transformer model. In: ECCV. pp. 430–446. Springer (2022)
17. Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., Li, H.: Uformer: A general u-shaped transformer for image restoration. In: CVPR. pp. 17683–17693 (2022)
18. Xie, Q., Zhao, Q., Meng, D., Xu, Z.: Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery. IEEE TPAMI **40**(8), 1888–1902 (2017)
19. Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A.: Places: A 10 million image database for scene recognition. IEEE TPAMI **40**(6), 1452–1464 (2017)
20. Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., Torralba, A.: Scene parsing through ade20k dataset. In: CVPR. pp. 633–641 (2017)
21. Zhu, X., Milanfar, P.: Removing atmospheric turbulence via space-invariant deconvolution. IEEE TPAMI **35**(1), 157–170 (2012)