Depth from Optical Turbulence

Yuandong Tian
Carnegie Mellon University
yuandong@cs.cmu.edu

Srinivasa G. Narasimhan
Carnegie Mellon University
srinivas@cs.cmu.edu

Alan J. Vannevel
U.S. Naval Air Warfare Center
alan.vannevel@navy.mil

Abstract

Turbulence near hot surfaces such as desert terrains and roads during the summer, causes shimmering, distortion and blurring in images. While recent works have focused on image restoration, this paper explores what information about the scene can be extracted from the distortion caused by turbulence. Based on the physical model of wave propagation, we first study the relationship between the scene depth and the amount of distortion caused by homogenous turbulence. We then extend this relationship to more practical scenarios such as finite extent and height-varying turbulence, and present simple algorithms to estimate depth ordering, depth discontinuity and relative depth, from a sequence of short exposure images. In the case of general non-homogenous turbulence, we show that a statistical property of turbulence can be used to improve long-range structure-from-motion (or stereo). We demonstrate the accuracy of our methods in both laboratory and outdoor settings and conclude that turbulence (when present) can be a strong and useful depth cue.

1. Introduction

The visual manifestations of clear air turbulence occur often in our daily lives — from hot kitchen appliances like toasters and ovens, to plumes of airplanes, to desert terrains, to roads on hot summer days, to the twinkling of stars at night. The shimmering and distortion observed are caused by random fluctuations of temperature gradients near warm surfaces. In this case, the image projection of a scene point viewed through turbulence is no longer a deterministic process, and often leads to poor image quality.

Several works in remote sensing and astronomical imaging have focused on image correction through turbulence. For atmospheric turbulence, the distorted wavefronts arriving from stars can be optically corrected using precisely controlled deforming mirror surfaces, beyond the angular resolution limit of telescopes [16]. For terrestrial imaging applications, recent works have proposed to digitally post-process the captured images to correct for distortions and to deblur images [6, 3, 4, 9, 26]. Optical flow based methods have been used further to register the image sequences to achieve modest super-resolution [18].

While previous works have focused on what turbulence does *to* vision, this article addresses the question of what
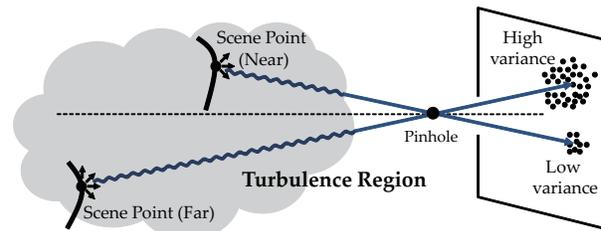

Figure 1. Random fluctuations in the refractive index of a medium cause the perturbation of a light wave radiating from a scene point. The resulting image projections of the scene point over time are also random. The longer the distance of a scene point from the camera, the greater the variance of its image projection.

turbulence can do *for* vision. In other words, what information about the scene can be extracted when viewed through turbulence? Based on the physical model of wave propagation, we study the relationship between the scene depth and the amount of distortion caused by homogenous turbulence over time (see an intuitive illustration in Fig 1). Then, we extend this relationship to more practical scenarios of finite extent and height-varying turbulence, and show how and in what scenarios we can estimate depth ordering, depth discontinuity and relative depths. Although general non-homogenous turbulence does not directly yield depth information, its statistical property can be used along with a stereo camera pair to improve long-range depth estimation.

The input to our techniques is a sequence of short exposure images captured from a stationary camera (or camera pair). Depth cues are obtained by first tracking image features and then by computing the variances of tracker displacements over time. Any feature tracking algorithm can be applied, such as that based on template matching. We verify our approaches in both laboratory and outdoor settings by comparing against known (ground truth) distances of the scene from the camera. We also analyze how the depth cue estimation is influenced by the parameters of the imaging system, such as aperture, exposure time and the number of frames. The depth information computed is surprisingly accurate, even when the scene and camera are not within the turbulent region. Hence, we believe that turbulence should not be only viewed as "noise" that an imaging system must overcome, but also as an additional source of information about the scene that can be readily extracted[1].

---

[1]While not the focus of this work, the short exposure (noisy) and distorted input images can be combined using a dense image alignment approach [21] to improve image quality (see Supplementary material).

1.1. Related Work

Characterizing the structure of turbulence is one of the open problems in physics, with a long research history, starting from the early methods of Kolmogorov [10]. For this work, we reference multiple textbooks by Kopeika [12], Tatarskii [20], Ishimaru [8] and Roggemann [16]. To our knowledge, the key physical model (Eqn. 3) in these texts has not been exploited by the computer vision community.

Direct measurement of turbulent media has received much attention in fluid dynamics. Shadowgraph and Schlieren imaging [17, 24] techniques are often used to capture the complex airflow around turbines, car engines and airplanes wings. Image displacement observed in turbulent media has been shown to be proportional to the integral of the refractive index gradient field. This property is exploited in a tomographic approach [5] with many views to compute the density field of the medium from image displacements of known backgrounds. This approach, called Background Oriented Schlieren (BOS) imaging [23, 22, 15], has emerged as a new technique for flow visualization of density gradients in fluids. Such approaches have also been used to render refractive volumes of gas flow [2]. Similarly, there has been work [13] that aims to estimate the shape of a curvy refractive interface between two media (water and air, for example) using stereo and known backgrounds. In contrast, our work exploits image displacements to extract the depths cues of an unknown scene using an image sequence captured from a single viewpoint.

2. Characterization of Turbulence

Turbulence causes random fluctuations of the refractive index $n(\mathbf{r}, t)$ at each location $\mathbf{r}$ in the medium and at time t. From Kolmogorov's seminal work [10, 11], $n(\mathbf{r}, t)$ forms a random field in space-time and can be characterized by a structure function $D(\mathbf{r}_1, \mathbf{r}_2, t)$ that computes the expected squared difference of refractive index at two distinct spatial locations $\mathbf{r}_1$ and $\mathbf{r}_2$:

$$D(\mathbf{r}_1, \mathbf{r}_2, t) = \langle |n(\mathbf{r}_1, t) - n(\mathbf{r}_2, t)|^2 \rangle. \quad (1)$$

For *stationary* turbulence, the structure function is constant over t, *i.e.*, $D(\mathbf{r}_1, \mathbf{r}_2, t) = D(\mathbf{r}_1, \mathbf{r}_2)$. Stationary turbulence is *homogeneous* if $D(\mathbf{r}_1, \mathbf{r}_2) = D(\mathbf{r})$, where $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$. This means that the structure function depends only on the relative displacement of the locations. Homogenous turbulence is *isotropic* if the structure function is spherically symmetric, *i.e.*, $D(\mathbf{r}) = D(r)$, where $r = \|\mathbf{r}\|$. From dimensional analysis, Kolmogorov shows that the structure function follows a 2/3 power law [20]:

$$D(r) = C_n^2 r^{2/3}, \quad (2)$$

where, the constant C_n^2 reflects the strength of turbulence. For non-homogeneous turbulence, C_n^2 is a function of absolute location. A non-turbulent atmosphere has $C_n^2 = 0$.
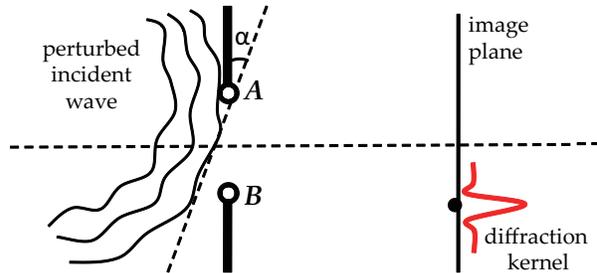

Figure 2. The phase difference of an incident wave (*e.g.*, the phase of point B *leads* that of A) at the aperture determines the angle-of-arrival α (AoA) and in turn, the center of the diffraction kernel, *i.e.*, the location of the projected scene point in the image plane.

In general, the strength C_n^2 of turbulence depends on a variety of environmental and physical factors, such as temperature, pressure, humidity and wavelength of light, which in turn depend on the time of day (less during sunset and sunrise, more at mid-day), cloud cover (less during cloudy day and more during cloudy nights), and wind patterns. An empirical relationship between these factors and refractive index changes can be found in Kopeika's textbook [12].

3. Image Formation through Turbulence

When an electromagnetic wave propagates through a turbulent medium, it undergoes random fluctuations in both amplitude and phase. The perturbed phase determines the angle-of-arrival (AoA) of the light incident at the camera, which in turn fixes the projected location of the scene point in the image (Fig. 2). Mathematically, the propagation of an electric field under the influence of the turbulence structure function in Eqn. 2 can be obtained by solving Maxwell's equations. Since most surfaces and sources of interest to computer vision are at finite distances from the camera and produce divergent waves, we will consider the propagation of spherical waves. Then, following the derivations in [12, 8], the variance $\langle \alpha^2 \rangle$ of the angles-of-arrival of the waves from a scene point at distance L from the camera is obtained by integrating along the line of sight:

$$\langle \alpha^2 \rangle = 2.914 D^{-1/3} \int_0^L C_n^2(z) \left(\frac{z}{L} \right)^{5/3} \mathrm{d}z, \quad (3)$$

where, D is the diameter of the aperture. The actual fluctuation $\langle \delta^2 \rangle$ of the projected image location can be computed using the relation $\delta = f \tan \alpha$ where, f is the focal length of the camera. For small angles, $\delta \approx f\alpha$.

In the following, we will discuss three important special cases of the above image formation model. We will address the general case of non-homogeneous turbulence in Section 7. First, consider a scenario where both the camera and the scene of interest are immersed in a homogeneous turbulence medium (for example, a road scene with vehicles on a hot summer day), as illustrated in Fig. 3(a). Since
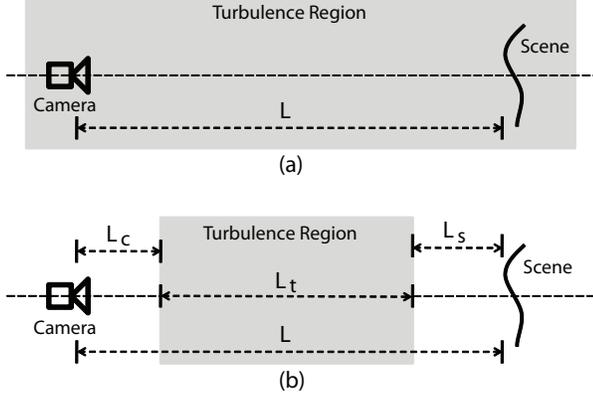
Figure 3. Image formation through turbulence. **(a):** Both the camera and the scene are immersed within a homogeneous turbulence region. **(b):** The camera and/or scene are outside the turbulence region.

C_n^2 is a constant, we can integrate Eqn. 3 to obtain:

$$\langle\alpha^2\rangle = 2.914D^{-1/3}C_n^2\int_0^L\left(\frac{z}{L}\right)^{5/3}\mathrm{d}z$$
$$= \frac{3}{8}K_n^2L, \qquad (4)$$

where, $K_n^2 = 2.914D^{-1/3}C_n^2$. So, the variance of projected positions of the scene point in the image plane over time is directly proportional to the distance L between the scene point and the camera. Setting aside the issue of spatial resolution, this linear relationship determines depth with constant precision for all distances within the turbulence region. By comparison, in stereo, the depth precision falls as the square of the distance from the camera to the scene.

In many scenarios, like the plume of an aircraft or a steaming kettle, the source of turbulence may not extend over the entire line-of-sight from the camera to the scene. In this case, we will assume local homogeneity, *i.e.*, C_n^2 is a constant within a short range and zero elsewhere. For convenience, we decompose L into three parts: $L = L_s + L_t + L_c$, as illustrated in Fig. 3. L_s is the distance between the scene point and the turbulence region, L_t is the path length within the turbulence region and L_c is the distance between the camera and the turbulence region. Once again, we can integrate Eqn. 3 to obtain the analytic form:

$$\langle\alpha^2\rangle = \frac{K_n^2}{L^{5/3}}\int_{L_s}^{L_t+L_s}z^{5/3}\mathrm{d}z$$
$$= K_n^2\frac{3}{8L^{5/3}}\left((L_t+L_s)^{8/3}-L_s^{8/3}\right). \qquad (5)$$

Letting $\langle\alpha^2_{\mathrm{relative}}\rangle = \frac{3}{8L^{5/3}}\left((L_t+L_s)^{8/3}-L_s^{8/3}\right)$ allows us to write in short:

$$\langle\alpha^2\rangle = K_n^2\langle\alpha^2_{\mathrm{relative}}\rangle. \qquad (6)$$
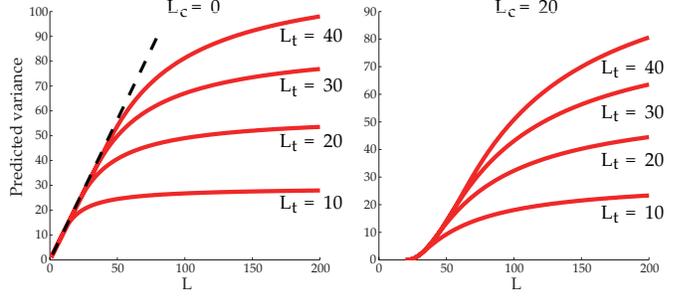

Figure 4. The variance of the angle-of-arrival predicted by Eqn. 6 under different experimental settings. For each curve, the camera and the extent of turbulence (L_c and L_t) are fixed, while the scene depth (L_s) is varied. Each curve represents a monotonically increasing function of scene depth that converges to a fixed variance (at infinity). The dashed black line is the linear relation in the special case when $L_c = L_s = 0$.

If we fix the camera location L_c and the turbulence region L_t, and move the scene point away from the camera, the variance is a monotonically increasing function with respect to L, as shown in Fig. 4. From this, we observe that the variance increases even if the scene moves away from the turbulence region. This is a counter-intuitive result that cannot be explained by ray optics (hence, the usage of "waves" in this article). The variance, however, converges to a fixed value $\langle\alpha^2_\infty\rangle$, when the scene point is infinitely far away from the camera (*e.g.*, a distant star). This can be seen by taking the limit $L_s \to \infty$ in Eqn. 6 to obtain:

$$\langle\alpha^2_\infty\rangle = K_n^2\lim_{L_s\to\infty}\frac{3}{8L^{5/3}}\left((L_t+L_s)^{8/3}-L_s^{8/3}\right)$$
$$= K_n^2L_t. \qquad (7)$$

In this case, the light emitted by the scene point can be modeled as a plane wave.

Height-varying turbulence. In practice, the air turbulence may not be homogenous in the entire field of view. For example on an asphalted road, the turbulence is more severe near the road surface than away from it. We model this effect by writing the strength of turbulence as a smoothly varying function of height h, yielding a separable model:

$$\langle\alpha^2\rangle = K_n^2(h)\langle\alpha^2_{\mathrm{relative}}\rangle. \qquad (8)$$

Typically, $C_n^2(h)$ (or $K_n^2(h)$) decreases with respect to h.

4. Depth cues from an Image Sequence

In this section, we investigate what depth cues can be obtained from the observed variance of image displacements of scene points. The input to all our algorithms is a sequence of images of a stationary scene viewed through turbulence by a fixed camera. Once the images are captured, we track a sparse set of distinctive feature points on the scene. While any feature-tracking algorithm may be used,
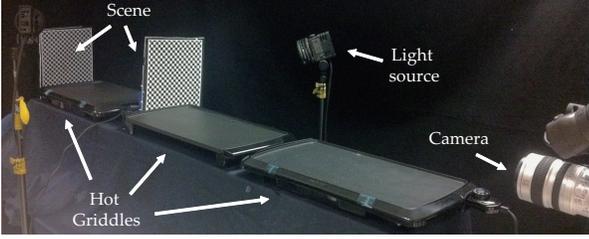
Figure 5. Experimental setup: Three adjacent electric cooking griddles are heated up to 400 degrees Fahrenheit to create hot air turbulence. A camera observes a scene through the turbulence. By varying the temperature, we can emulate a variety of outdoor turbulence strengths and path-lengths of several kilometers.
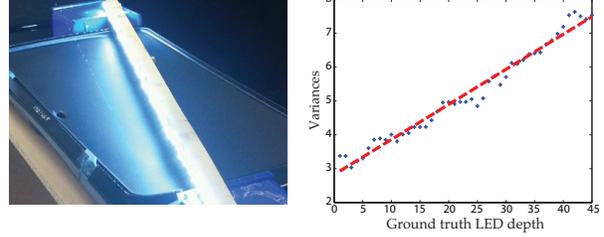

Figure 6. **Left:** LEDs are immersed in the turbulence region. **Right:** The relationship between the variance of LED projections and their ground truth depths is very close to linear (correlation coefficient is 0.987), and is consistent with the model (Eqn. 6).

we adopt a simple frame-to-frame template matching approach. To handle image blurring caused by turbulence, we also add blurred versions of the templates. The variance of the image location of each tracked point is then computed.

For a fixed configuration of camera and extent of the homogeneous turbulence region, the model (Eqn. 6) is a monotonic smooth function of scene depth. Thus, both depth ordering and discontinuities (like two buildings far apart) of the scene can be readily obtained from variances. In particular, detecting such discontinuities can be useful to segment the scene into different depth layers (planes).

On the other hand, a more quantitative measurement, such as relative depth between scene points, requires additional assumptions. Note that absolute depth cannot be computed without knowing the turbulence strength, C_n^2. Thus, without loss of generality, we will assume $L_t = 1$. When the camera and scene are immersed in turbulence ($L_c = L_s = 0$), the linear variance-depth relationship (Eqn. 4) allows us to obtain relative depth by taking variance ratios to eliminate the unknown constant K_n^2. In general, if L_c, L_s and K_n^2 are known, depth can be obtained by inverting Eqn. 6. By monotonicity of the model, only a unique depth can be obtained from a given variance.

However, in practice, these constants are usually unknown. Thus, for N scene points we have $N + 3$ unknowns (N depths plus L_c, L_s and K_n^2) but N equations. Consider a scene with repetitive patterns (windows on a building, street lamps, cars parked on a street), then the depths $\{L_i\}_{i=1}^N$ of the N points follow an arithmetic sequence:

$$L_i = L_0 + i\Delta L \qquad (9)$$

Thus, N depths $\{L_i\}_{i=1}^N$ are parameterized by 2 variables, L_0 and ΔL. As a result, only $2 + 3 = 5$ scene points suffice to estimate (using numerical optimization) both the relative depths and the extent of the turbulence region.

In the case of height-varying turbulence, we need to estimate the height-varying function $K_n^2(h)$ as well as the scene depth. Fortunately, if the height is aligned with the y-axis of the image, then separating depth from height can be achieved by treating each scan-line individually.

5. Laboratory Experiments

We performed several experiments in a controlled laboratory environment to validate the theory. A flat cooking griddle of size 52cm × 26cm is used to produce and maintain uniform heat of up to 400 degrees Fahrenheit, across the flat surface. In the experiment setup (Fig. 5), multiple such griddles are placed side by side to increase the path-length of turbulence. The three griddles set at maximum temperature produce roughly the same shimmering as a kilometer of natural turbulence in the desert. By controlling the number of griddles and the temperature, a wide range of turbulence strengths seen outdoors can be emulated. In all experiments, variances are estimated by capturing a 20-30 seconds long video sequence of the scene at 30 fps.

5.1. Quantitative Evaluation

Variance-depth linearity within turbulence region. 50 equally-spaced LEDs are placed 5 cm above the hot griddle (in the turbulence region). One end of the stick is closer to the camera while the other is farther away. Fig. 6 shows the variances computed for each LED projection onto the image plane averaged over 3 experimental trials. Consistent with the model (Eqn. 6) when $L_s = L_c = 0$, indeed the relationship between the depth (represented by the indices of LED) and the variances is linear with a high correlation coefficient of 0.987. A similar experiment that estimates the depth of a curvy line on a sphere is also shown in Fig. 7.

Identifying depth discontinuity. In this experiment, we place two checker-board patterns vertically at two distinct depths (L^{near} and L^{far}) and measure variances of the key points on the scene. We conducted four experiments with different settings of L^{near} and L^{far} (Table 1). All were captured in the same setting of $f/11$ with exposure $1/2000$, while the zoom was varied to include the entire scene within the field of view. Fig. 8 illustrates the variance discontinuity by two separate parametric fittings of the key points on two checker-boards in one experiment. Clearly, the depth discontinuity can be detected from the variance discontinuity.

Validation of the physics model (Eqn. 6). As shown in Fig. 8, due to height variations of the turbulence, the vari-
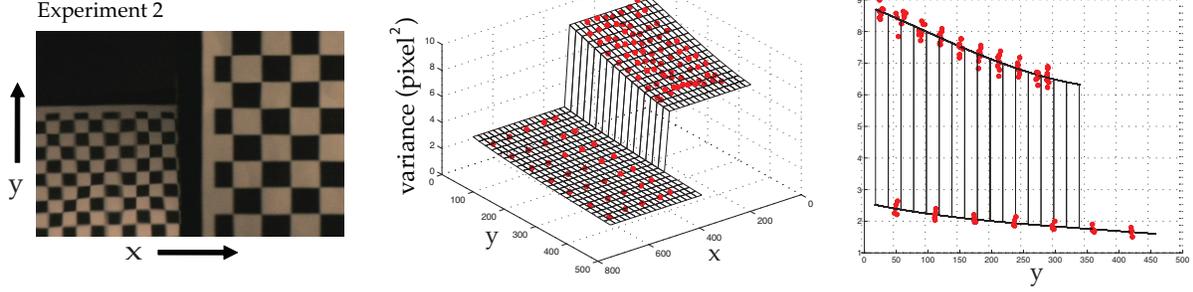
Figure 8. Experiments with two planar checker-boards placed at different distances from the camera. Due to space limit, we only show 1 (Exp.2) of the 4 experiments, and leave the remaining in the supplementary material. The experimental setting can be found in Table 1. The first column shows a sample distorted frame, the second and third columns show two views of the variance distribution of the corners of the checker-boards. From the figures, variances changes due to depth discontinuity and height is obvious. We detect the discontinuity and fit smooth surfaces to the variances. The ratio of variances of the two depth planes are then computed and quantitatively compared to the ground truth (Table 1).
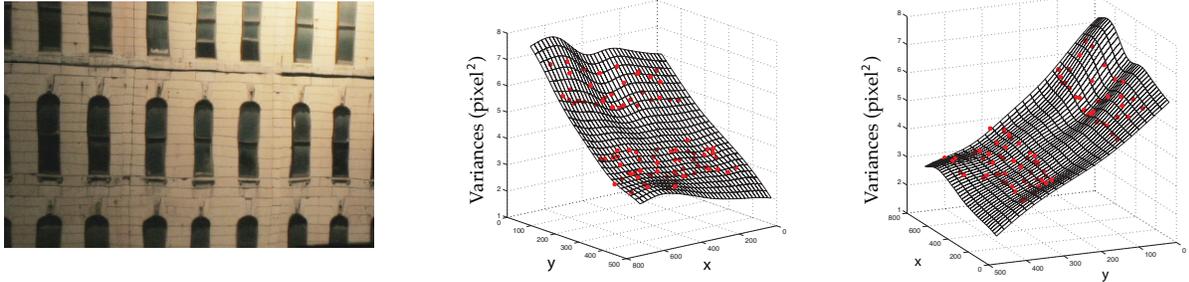

Figure 9. Depth estimation of equally spaced points on an inclined planar scene. A sample distorted frame due to turbulence is shown on the left. On the right, are two views of the variance distribution of a sparse set of key points and a smooth function fit illustrating the near planar geometry. From this, it is possible to predict the relative depths of the scene points (assuming the length of the turbulence region to be 1). For evaluation, the relative depth estimated is converted to actual depth by using the actual length (196cm) of turbulence region. The estimated slope of the plane is 0.517 cm per 10 pixels in horizontal direction, compared to the ground truth value 0.529 cm per 10 pixels.
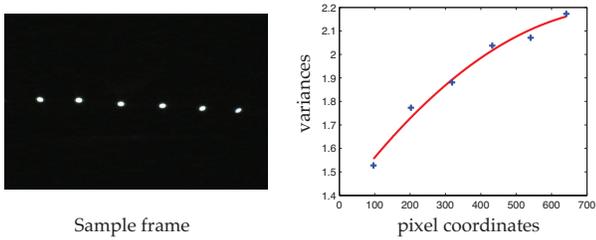

Sample frame

Figure 7. Ellipse fitting on a video sequence capturing a curve on the sphere through turbulence. Ideally the projection of the LEDs forms an ellipse on the image plane. **Left:** A sample frame of the captured video sequence. **Right:** Average error in fitting is 12.6%. The average fitting error between a covariant x and dependent variable y is computed using $\sqrt{\sum_i (\hat{y}_i - y_i)^2}/\sqrt{\sum_i (\bar{y} - y_i)^2}$, where $\hat{y}_i$ is the fitted value of point x_i and $\bar{y}$ is the mean of y.

No.	L_c	L_t	L^{near}	L^{far}	Measured	Predicted
Exp1	54	171	163	225	1.77	1.79
Exp2	74	173	183	382	3.60	3.52
Exp3	74	173	247	382	1.67	1.94
Exp4	74	173	247	320	1.56	1.55

Table 1. Columns 1-4 show the ground truth measurement (in centimeters) for the four checker-board experiments. Columns 5-6 show the comparison between the measured (5th column) variance ratio and that predicted by the model (6th column). In all but one case, the measurements are very accurate.

ance changes smoothly over the y-axis. However, the variance ratio computed by two points on two checker-boards at the same scan-line is independent of height h, amount of turbulence C_n^2 and aperture diameter D. On the other hand, we can compute the theoretical variance ratios using the ground truth value of L, L_c, L_t and model Eqn. 6,. The measurement is consistent with the theory, validating the model in all four settings (Table 1) that covers both cases where the scene is within and outside the turbulence region.

Depth estimation of equally spaced scene points. We estimate the depths of equally spaced points using the method in Section 4. A horizontally slanted plane with a texture of a building facade is placed behind the turbulence region. Fig. 9 shows the two views of the computed variances at key points and the surface fits that demonstrate the near planar geometry. Assuming $L_t = 1$ (length of turbulence region), relative depths of scene points can be computed. For validation, the estimated relative depths are converted to absolute ones using the actual length ($L_t = 196$ cm) of the turbulence region. The depth slope of the plane (ΔL in Section 4) is estimated as 0.517 cm per 10 pixels in the horizontal direction, compared to the ground truth value of 0.529 cm per 10 pixels. Please see the supplementary material for videos.
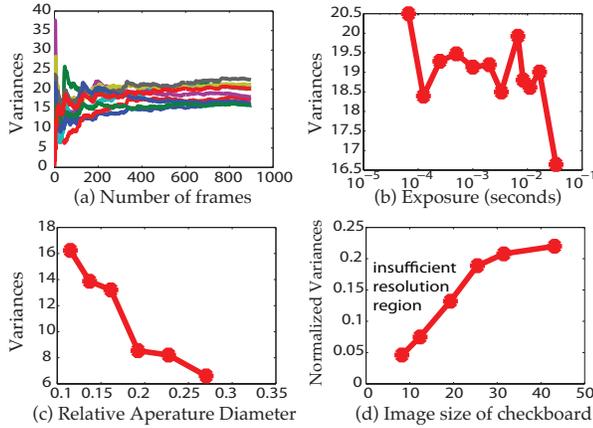
Figure 10. Influence of imaging parameters on the variance of the corners of the checker-board pattern. **(a)** The variance estimates converge with sufficient frames, showing that the turbulence is stationary during measurement. **(b)** The measured variance is similar for different exposures, except for very long exposures where the tracking performance degrades due to motion blur. **(c)** Consistent with the model, the measured variance decreases significantly with aperture size. **(d)** Insufficient image resolution results in much lower variance estimation.

5.2. Influence of Imaging Parameters

The accuracy of the measured variance depends on many imaging parameters. Larger aperture reduces the depth-of-field, higher exposure time adds unwanted motion blur, and low image magnification causes greater quantization of the variance. Here we present an empirical study.

The estimate of the variance converges as the number of captured frames increases. Fig. 10(a) shows the variance of the 10 key points in Exp.1 with different numbers of frames used. In our experiments, stable estimates are achieved using frames captured over 30 seconds using a 30 fps camera. To study the effects of aperture, we vary the $f/\#$ from $f/3.7$ to $f/11$, fix the exposure time at $1/8000s$, and zoom at the highest level. Fig. 10(c) shows a significant decrease in variance as predicted by the model in Eqn. 6. The plot in Fig. 10(b) shows the variances computed by changing only the exposure time. Fig. 10(d) shows the effect of varying focal-length. Here we normalize the variances by the pixel size of the checker-board patterns to remove the effects of image magnification. In these experiments, we have tried to maintain the same noise level in the camera by maintaining similar image brightness (by varying illumination intensity). From these plots, aperture size is the main factor that affects the estimation quality. However, since we take variance ratio as a depth measure, the effect of aperture size is reduced (in theory, independent). Also, for very low magnification (spatial resolution) or long exposure time ($1/30s$), the variance estimate shows large degradation.

In addition, camera shaking can cause false displacements leading to poor variance estimates. In general, this is a hard problem and we will set it aside for future work.
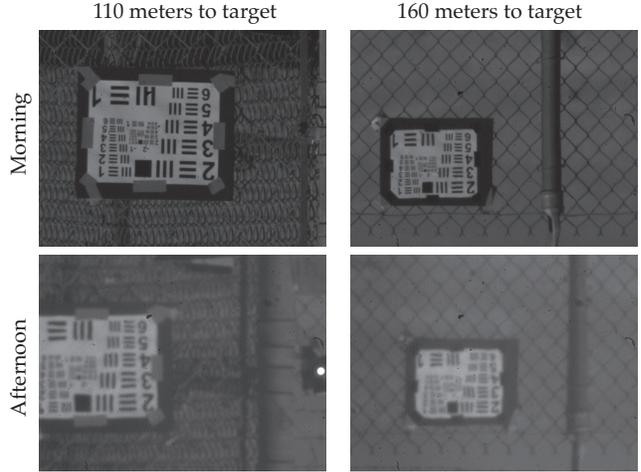

Figure 11. Sample frames of the outdoor experiments in the morning and afternoon. The targets are placed at different distances from the camera.

6. Outdoor Experiments

Besides indoor experiments in a controlled environment, we also conducted experiments outdoors in a desert region. The imaging setup used for the experiments consists of a Prosilica GC1380H camera and a Celestron C6 Telescope. The focal length is 1500mm. A one-to-one ratio optical relay is used between the telescope and the camera, without changing the focal length of the main telescope. We placed two standard contrast targets 110 meters and 160 meters away from the camera and captured sequences of 300-400 images during mild turbulence (morning) and strong turbulence (afternoon). We used a 30mm aperture in the morning and a 10mm aperture in the afternoon, and varied the exposure times between 0.5ms and 1ms.

We tracked a sparse set of points through each image sequence and rejected outliers such as the static trackers of the dirt on the CCD and high-variance erroneous trackers near locally repetitive textures. The computed variances of the trackers converge quickly (see supplementary material).

Table 2 shows the mean variance computed from all the trackers of each image sequence, for each depth and imaging setting. Since the amount of variance is relatively invariant to exposure change, we further averaged the variance over different exposures. If we take the variance ratio between 110m and 160m, we obtain 1.6857 for 30mm aperture in the morning and 1.5785 for 10mm aperture in the afternoon. Both are close to the ratio of two distances $160/110 = 1.4545$, verifying the dependence of turbulence model on the depth. Besides, Table 2 also shows the standard derivation of variances in each video sequence. The low standard deviation shows that the turbulence is indeed homogenous on surfaces that are perpendicular to the optical axis. Note we do not consider the height variation of turbulence, since compared to the laboratory setting, the target occupies a much narrower field of view.

Morning Capture. Target 110meters away, 30mm aperture.		
0.50ms exposure	0.75ms exposure	1.00ms exposure
4.32 ± 0.49	4.89 ± 0.46	4.79 ± 0.45
Average: 4.67		
Morning Capture. Target 160meters away, 30mm aperture.		
0.50ms exposure	0.75ms exposure	1.00ms exposure
8.23 ± 0.88	7.61 ± 0.67	7.75 ± 0.64
Average: 7.86		
Afternoon Capture. 10mm aperture and 5ms exposure		
Target 110meters away	Target 160meters away	
43.68 ± 7.50	69.37 ± 9.49	

Table 2. Average variance (and its standard derivation) of trackers for outdoor experiment. The variance ratios between 160m and 110m turbulence video are 7.86/4.67=1.6857 (30mm aperture captured in the morning) and 69.37/43.68=1.5785 (10mm aperture captured in the afternoon), close to the distance ratio (160m/110m=1.4545), verifying our model.

7. Jitter-stereo in Nonhomogenous Turbulence

Until now, we have addressed depth estimation under homogenous and simple height-varying turbulence. However, due to unpredictable temperature and humidity fluctuation, turbulence cannot be guaranteed to be homogenous over a large area and a long period of time. In this case, it is impossible to estimate depth using the model without knowing the turbulence structure function. Instead, we will exploit a general statistical property of turbulence along with a stereo camera pair to improve long-range depth estimation.

Recall that binocular stereo estimates the depth of a scene point by computing scene disparity across two views. If a scene point is far away compared to the stereo baseline, the disparity may be less than one pixel and the depth estimation may fail. However, in the presence of turbulence, the location of the scene point "jitters" around its true location in the image. But how do we estimate the true location of a scene point without turbulence? It has been observed that distribution of (even non-homogeneous) turbulence distortions is close to zero-mean. This is true even if the variance of each scene point is different. Thus, the mean positions of the tracked scene points are their most-likely positions when there is no turbulence. Furthermore, estimating mean locations of trackers allows us to obtain the disparity possibly with sub-pixel accuracy, helping in long-range depth estimation with a short baseline. This approach is also similar in spirit to Swirski et. al [19] that estimates correspondences in stereo using underwater caustics patterns.

In order to experimentally verify our approach, we captured two image sequences of a planar scene 110m away from different view points, with baseline of less than 1m. The two sequences were captured at different times when the turbulence was significantly different. We tracked a common sparse set of scene points in the two views and compute the mean locations. We also verified that the distribution of tracker displacements are zero-mean in all
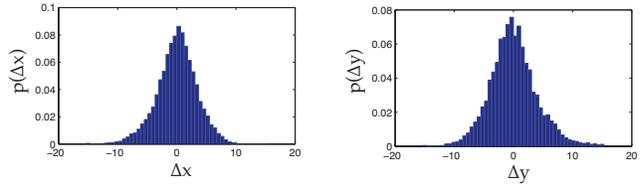

Figure 12. Distribution of x and y displacement of trackers in outdoor image sequence, computed over all trackers and all frames from an image sequence in the afternoon. Consistent with our assumption, they follows a zero-mean distribution.
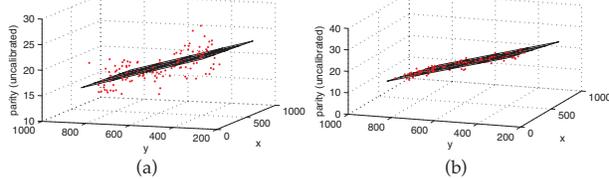

Figure 13. Jitter stereo in turbulence. **(a)** (Uncalibrated) disparity computed from two video frames at a certain time, due to turbulence, the disparity is noisy. Over the 319 frames, the correlation coefficient varies from 0.632 to 0.954 with mean being 0.884 and standard derivation being 0.054. **(b)** (Uncalibrated) disparity using mean tracker locations. The disparity is clearly linear (correlation coefficient is 0.976).

our experiments (see Fig. 12). The disparities between the mean-locations of corresponding trackers are then estimated. Note that the disparity of a scene point on a plane is a linear function with respect to its x and y coordinates on the image. The linear fit is strong with a correlation coefficient of 0.976 and is significantly better than computing disparities on a per-frame basis, as shown in Fig. 13.

8. Comparisons with Depth-from-X

This work introduces turbulence as a new cue for scene depth. So, it is instructive to discuss the parallels and differences between depth from turbulence and other depth-from-X approaches in computer vision.

Structure from motion (SFM): SFM relies on estimating the pixel disparity across different views of the scene that reduces with depth. In the case of turbulence, variance of projected scene point is measured from the same camera position over time and monotonically increases with scene distance. Thus, while SFM is suited for shorter distances (for a fixed baseline), depth from turbulence is better in general for a longer range and/or stronger turbulence. But if the scene is outside the turbulence region, the depth precision degrades in the asymptotic region of the variance curve (Fig. 4). At the same time, both approaches share the same issues with finding and tracking corresponding features.

Depth from defocus or diffusion (DFD): In both cases, the point-spread function varies across the scene. The extent of the observed blur monotonically increases with distance from the sensor in the case of turbulence and distance from the focal/diffuser plane in the case of DFD [7, 25]. Depth from turbulence requires capture of a temporal sequence of

images, and is similar to moving a pinhole across the aperture of the lens to compute depth [1].

Structure from bad weather: Perhaps depth from fog or haze [14] is most similar in spirit to depth from turbulence. These approaches also use a single viewpoint, provide measures that are more reliable for scenes with long distances, and are (mostly) independent of the scene reflectance properties. That said, there are also fundamental differences. Turbulence is a statistical and temporally varying phenomenon, where depth cues are due to phase variations of the incident light rather than the intensity variations as in fog. The environmental illumination (air light) provides a strong cue for depth from fog, whereas the specific illumination geometry of the environment plays little or no part in depth from turbulence.

9. Conclusion and Future Work

This article is an initial attempt at understanding what depth cues can be extracted from optical turbulence. We derived a simple relation between scene depths and variance of the projected scene points under turbulence. Our experiments showed, somewhat surprisingly, that accurate depth cues can be obtained from optical turbulence. There are several avenues of future work including dense scene reconstruction and image super-resolution from the image sequence under turbulence. We wish to also study several other related physical phenomena. The twinkling of stars is caused primarily due to the changes in amplitude of the incident wave that are distance-related as well. Aside from temperature gradients, the chaotic movement of a medium itself can cause turbulence. This type of phenomenon can occur due to under water currents, due to strong wind flow in the upper atmosphere, and due to air flow around engines. We wish to build upon this work to apply to these scenarios.

References

[1] E. Adelson and J. Wang. Single lens stereo with a plenoptic camera. *IEEE Trans. on PAMI*, 14(2), 1992. 8

[2] B. Atcheson, I. Ihrke, W. Heidrich, A. Tevs, D. Bradley, M. Magnor, and H.-P. Seidel. Time-resolved 3d capture of non-stationary gas flows. *ACM TOG*, 27(5), 2008. 2

[3] J. Gilles, T. Dagobert, and C. Franchis. Atmospheric turbulence restoration by diffeomorphic image registration and blind deconvolution. In *ACIVS*, 2008. 1

[4] S. Harmeling, M. Hirsch, S. Sra, and Schlkopf. Online blind deconvolution for astronomy. In *ICCP*, 2009. 1

[5] G. Herman. *Image reconstruction from projections*. Academic Press, New York, 1980. 2

[6] M. Hirsch, S. Sra, B. Schlkopf, and S. Harmeling. Efficient filter flow for space-variant multiframe blind deconvolution. In *CVPR*, 2010. 1

[7] T. Hwang, J. Clark, and A. Yuille. A depth recovery algorithm using defocus information. In *CVPR*, 1989. 7

[8] A. Ishimaru. *Wave Propagation and Scattering in Random Media*. Wiley-IEEE Press, 1999. 2

[9] N. Joshi and M. Cohen. Seeing mt. rainier: Lucky imaging for multi-image denoising, sharpening and haze removal. *In ICCP*, 2010. 1

[10] A. Kolomogorov. *Dissipation of energy in the locally isotropic turbulence*. Comptes rendus de l'Acadmie des Sciences de l'U.R.S.S. 32: 16C18, 1941. 2

[11] A. Kolomogorov. *The local structure of turbulence in incompressible viscous fluid for very large Reynold's numbers*. Comptes rendus de l'Acadmie des Sciences de l'U.R.S.S. 32: 16C18, 1941. 2

[12] N. S. Kopeika. *A System Engineering Approach to Imaging*. SPIE Publications, 1999. 2

[13] N. Morris and K. Kutulakos. Dynamic refraction stereo. In *ICCV*, 2005. 2

[14] S. Narasimhan and S. Nayar. Vision and the atmosphere. *IJCV*, 48(3), 2002. 8

[15] H. Richard, M. Raffel, M. Rein, J. Kompenhans, and G. Meier. Demonstration of the applicability of a background oriented schlieren method. In *Intl. Symp. on Applications of Laser Techniques to Fluid Mechanics*, 2000. 2

[16] M. C. Roggemann and B. Welsh. *Imaging Through Turbulence*. CRC press, 1 edition, Mar. 1996. 1, 2

[17] G. Settles. *Shlieren and Shadowgraph Techniques*. Springer-Verlag, 2001. 2

[18] M. Shimizu, S. Yoshimura, M. Tanaka, and M. Okutomi. Super-resolution from image sequence under influence of hot-air optical turbulence. In *CVPR*, 2008. 1

[19] Y. Swirski, Y. Schechner, B. Herzberg, and S. Negahdaripour. Stereo from flickering caustics. In *ICCV*, 2009. 7

[20] V. Tatarskii. *Wave Propagation in a Turbulent Medium*. McGraw-Hill Books, 1961. 2

[21] Y. Tian and S. Narasimhan. A globally optimal data-driven approach for image distortion estimation. In *CVPR*, 2010. 1

[22] G. Vasudeva, D. R. Honnery, and J. Soria. Non-intrusive measurement of a density field using the background oriented schlieren method. In *4th Australian Conf. on Laser Diagnostic in Fluid Mechanics and Combustion*, 2005. 2

[23] L. Venkatakrishnan and G. E. A. Meier. Density measurements using the background oriented schlieren technique. *Experiments in Fluids*, 37:237–247, 2004. 2

[24] G. Wetzstein, W. Heidrich, and R. Raskar. Hand-held schlieren photography with light field probes. In *ICCP*, 2011. 2

[25] C. Zhou, O. Cossairt, and S. Nayar. Depth from diffusion. In *CVPR*, 2010. 7

[26] X. Zhu and P. Milanfar. Stabilizing and deblurring atmospheric turbulence. *In ICCP*, 2011. 1