A unified Benchmark for Multi-Frame Image Restoration under Severe Refractive Warping

Supplementary Material

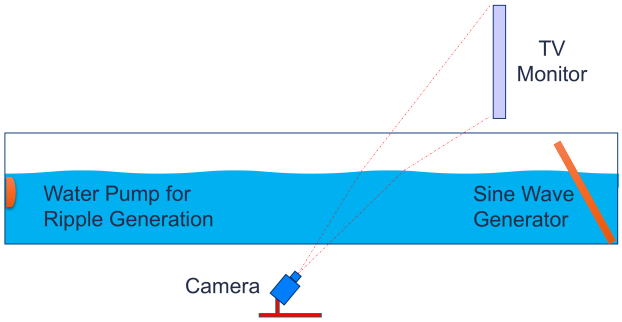

Figure 6. Data collection setup in the lab with water tank and water generators.

6. LAB setup

Figure. 6. shows the laboratory data collection set up with a large water tank (20 × 7 × 3 feet) which was filled with approximately 19-inch depth of water. A TV monitor was placed above the water to display a set of background images. The camera was set up below the water tank pointing towards the TV. During video recording, a wave generator at one end of the water tank produced disturbance on the water surface with a sine wave profile of a frequency between 1.0 to 2.8 Hz and wave amplitude ranging from 3 to 15mm peak-to-peak. Besides the sine wave generator, two water pumps were used to generate additional random ripple-like waves.

7. Wave generation

Table 5 provides parameters used to generate wave profiles. Details of specific wave types are provided below.

7.1. Ocean Wave

For our simulation, we compute the Fast Fourier Transform (FFT) of Gerstner's equations to represent the wave height as a random field over horizontal position and time. The height $h(x,t)$ at the horizontal position $\mathbf{x} = (x,\ z)$ can be expressed as

$$h(\mathbf{x}, t) = \sum_{\mathbf{k}} \tilde{h}(\mathbf{k}, t)\ \exp(i\,\mathbf{k}\cdot\mathbf{x}) \tag{1}$$

Where $\tilde{h}(\mathbf{k}, t)$ denotes the complex spectral coefficient at wavevector $\mathbf{k}$ and time t. This spectral form enables efficient analysis and synthesis of the surface, including fine control over amplitudes and phases across the discretized wavenumber domain.

The wave height field is constructed using a spectral model that accounts for wind direction and speed. The Phillips spectrum [40] is used to define the wave amplitude at different wavenumbers. The spectrum is given by:

$$P_h(\mathbf{k}) = A\,\frac{\exp\left(-\frac{1}{(kL)^2}\right)}{k^4}\,\left|\hat{\mathbf{k}}\cdot\hat{\mathbf{w}}\right| \tag{2}$$

where $P_h(\mathbf{k})$ is the power spectrum at wave vector $\mathbf{k}$, A is an amplitude constant, and for a continuous wind of speed V the largest attainable scale is $L = V^2/g$, where g is gravitational acceleration and $\hat{\mathbf{w}}$ denotes the unit vector in the wind direction.

Water-wave height fields can be modeled as Gaussian random fields whose spatial power follows a prescribed spectrum; the most efficient synthesis assigns the corresponding Fourier coefficients and then transforms to physical space.

$$\tilde{h}_0(\mathbf{k}) = \frac{1}{\sqrt{2}}\,(\xi_r + i\,\xi_i)\,\sqrt{P_h(\mathbf{k})} \tag{3}$$

The terms ξ_r and ξ_i are two independent random numbers drawn from a standard normal (Gaussian) distribution with mean 0, variance 1. After incorporating the Phillips spectrum, conjugate-symmetry, and dispersion, the Fourier amplitudes of the wave field realization [40] at time at time t in the frequency domain is:

$$\tilde{h}(\mathbf{k}, t) = \tilde{h}_0(\mathbf{k})\,e^{i\,\omega(\mathbf{k})t} + \tilde{h}_0^*(-\mathbf{k})\,e^{-i\,\omega(\mathbf{k})t} \tag{4}$$

Here, $\tilde{h}_0$ is initial Fourier amplitude (at $t = 0$). The height field in Eq. 4 preserves the complex conjugation property by propagating waves "to the left" and "to the right". To obtain the spatial field $h(\mathbf{x}, t)$, we apply the inverse FFT to Eq. 4 and then compute the spatial gradients of the resulting height map to estimate surface normals and warping displacements. These normals and displacements are subsequently used to render wave-motion effects on the image.

7.2. Sine Wave

The sine wave generation model simulates wave surfaces by representing the wave height field as a deterministic sinusoidal function propagating over a 2D spatial domain. This approach simplifies the complex, stochastic nature of waves

into a single-frequency wave, suitable for basic visualization or rendering applications. The wave can propagate horizontally with its motion animated over time to mimic wave dynamics. The wave height h, at position x, z and time t (represented by discrete frames) is modeled as a plane wave using a sine function [18]. For image horizontal propagation, the wave height is:

$$h(x,t) = A\sin(k_x x - \omega t + \phi) \tag{5}$$

where A is the amplitude, ω is the angular frequency, and x and y are spatial coordinates in a 2D domain, the ϕ is the phase offset.

Sine wave is the extended version of the sine wave [19] to generalize to waves propagating in arbitrary directions by rotating $h(x,t)$ with a random angle.

7.3. Shallow Water Wave

The shallow-water equations—derived as a depth-averaged form of the incompressible Navier–Stokes equations [41]—govern conservation of mass and horizontal momentum for a free-surface fluid [19]. Let h_{sh} be the surface height on a Eulerian mesh grid and (u, v) is the 2D velocity, ρ is the fluid density and g is the gravitational acceleration, then the differential equations can be written as

$$\frac{\partial(\rho h_{sh})}{\partial t} + \frac{\partial(\rho h_{sh} u)}{\partial x} + \frac{\partial(\rho h_{sh} v)}{\partial y} = 0 \tag{7}$$

$$\frac{\partial(\rho h_{sh} u)}{\partial t} + \frac{\partial\left(\rho h_{sh} u^2 + \frac{1}{2}\rho g h_{sh}^2\right)}{\partial x} + \frac{\partial(\rho h_{sh} uv)}{\partial y} = 0 \tag{8}$$

$$\frac{\partial(\rho h_{sh} v)}{\partial t} + \frac{\partial(\rho h_{sh} uv)}{\partial x} + \frac{\partial\left(\rho h_{sh} v^2 + \frac{1}{2}\rho g h_{sh}^2\right)}{\partial y} = 0 \tag{9}$$

8. Video generation

The distorted videos are generated by applying 200-frame-long series of precomputed wave normals to a selected background resized to 512×512. We mimic the LAB setup with camera located underwater and assume low field of view (parallel rays coming from the camera). The vector form of Snell's law (Eq. 10) is applied to these rays, $\vec{v}_1$, at the water surface with $\vec{N}$ surface normals to produce refracted rays $\vec{v}_2$. $n1$ and $n2$ are refractive indexes of water and air, respectively.

$$\vec{v}_2 = \frac{n1}{n2}\vec{N} \times \left(-\vec{N} \times \vec{v}_1\right) - \vec{N}\sqrt{1 - \left(\frac{n1}{n2}\right)^2 \left\|\vec{N} \times \vec{v}_1\right\|_2^2} \tag{10}$$

Finally, we compute the lateral displacement of the ray at a given distance to the background from the water surface and generate 2D grid, which then is applied to the background producing the distorted frame. In our experiments we consider four levels of distortion to benchmark model responses in different regimes. We reuse the same wave profiles but scale the degree of deformation both by scaling the distance to the background and surface normals as $\vec{N}' = (1-\alpha)\cdot\vec{N}_0 + \alpha\vec{N}$, where $\vec{N}_0$ is the vertical normal representing a flat surface. This heuristic enables control of the degree of distortion, and we set the coefficients to ensure average std displacement of 0.002, 0.006, 0.018, and 0.054 relative to the image size for low, mid, high, and extreme wave amplitude in all sets regardless the wave type. Since evaluation of refracted rays for each pixel becomes computationally expensive at large image size and long sequence length, we implemented the above procedure on GPU using Pytorch library enabling fast sample generation both in training and inference.

9. Evaluation on the synthetic data

Table 6-Table 9 provide a full summary of the evaluation on ocean, shallow water, sine, and ripple waves at low, mid, high, and extreme levels of distortion. Pixel (PSNR and SSIM) and perception metrics (LPIPS, DINO, CLIP) are used. Entire video setup refers to evaluation of the metric for each frame in the video and then averaging. Comparison of this benchmark to the first frame setup may give a clue about variability over the video.

Table 5. Evaluation for ocean waves. L, M, H, E — low, medium, high, extreme wave amplitude. (*) refers to evaluation on multiple output frames and average the metric.

Setup (ocean waves)	**PSNR↑**	**SSIM↑**	**LPIPS$_{VGG}$↓**	**LPIPS$_{Alex}$↓**	**DINO↓**	**CLIP↓**
First frame						
L	23.43	0.813	0.075	0.016	0.339	0.066
M	17.92	0.585	0.186	0.097	0.827	0.177
H	14.47	0.437	0.356	0.237	1.981	0.416
E	11.76	0.343	0.527	0.457	3.462	0.749
*Entire video**						
L	21.78	0.754	0.096	0.055	0.432	0.086
M	16.71	0.524	0.227	0.125	1.077	0.227
H	13.39	0.389	0.425	0.310	2.515	0.530
E	10.88	0.305	0.586	0.544	3.980	0.888
Pixel average						
L	27.69	0.866	0.172	0.199	0.729	0.160
M	21.26	0.628	0.417	0.442	2.251	0.490
H	17.40	0.469	0.608	0.628	3.606	0.793
E	14.71	0.417	0.660	0.735	3.824	0.862
Grid deformation						
L	25.19	0.813	0.215	0.182	0.736	0.205
M	19.51	0.608	0.367	0.360	1.664	0.405
H	15.87	0.481	0.481	0.231	2.493	0.560
E	12.70	0.387	0.590	0.615	3.778	0.861
*Grid registration**						
L	25.44	0.874	0.071	0.042	0.301	0.060
M	20.65	0.721	0.144	0.079	0.655	0.137
H	15.08	0.466	0.366	0.251	2.067	0.454
E	11.39	0.324	0.572	0.520	3.842	0.870
*DATUM**						
L	26.19	0.863	0.125	0.078	0.396	0.089
M	21.09	0.718	0.229	0.147	0.902	0.198
H	15.60	0.468	0.457	0.353	2.533	0.533
E	12.05	0.312	0.621	0.584	4.023	0.842
V-cache A5						
L	23.95	0.789	0.134	0.077	0.045	0.407
M	23.46	0.772	0.138	0.077	0.397	0.120
H	21.16	0.672	0.186	0.100	0.640	0.171
E	16.54	0.502	0.331	0.207	1.594	0.372
V-cache A3						
L	24.86	0.813	0.126	0.075	0.373	0.114
M	24.27	0.798	0.120	0.073	0.289	0.092
H	22.22	0.717	0.157	0.089	0.449	0.129
E	18.11	0.559	0.157	0.268	1.102	0.263

Table 6. Evaluation for shallow water waves. L, M, H, E — low, medium, high, extreme wave amplitude. (*) refers to evaluation on multiple output frames and average the metric.

Setup (shallow water waves)	**PSNR↑**	**SSIM↑**	**LPIPS$_{VGG}$↓**	**LPIPS$_{Alex}$↓**	**DINO↓**	**CLIP↓**
First frame						
L	21.71	0.733	0.099	0.059	0.489	0.101
M	16.65	0.492	0.264	0.149	1.370	0.296
H	13.27	0.374	0.503	0.399	3.160	0.692
E	10.84	0.292	0.610	0.580	4.244	0.899
*Entire video**						
L	20.68	0.682	0.113	0.064	0.512	0.108
M	16.07	0.462	0.273	0.150	1.384	0.293
H	12.92	0.362	0.501	0.392	3.171	0.682
E	10.52	0.283	0.616	0.593	4.345	0.936
Pixel average						
L	25.67	0.800	0.204	0.228	0.821	0.197
M	20.03	0.536	0.446	0.427	2.468	0.509
H	16.52	0.432	0.612	0.606	3.606	0.793
E	14.08	0.404	0.671	0.750	3.921	0.875
Grid deformation						
L	20.76	0.648	0.328	0.326	1.370	0.354
M	17.29	0.496	0.413	0.386	1.991	0.466
H	14.21	0.413	0.551	0.525	3.401	0.735
E	11.64	0.352	0.636	0.668	4.377	0.937
*Grid registration**						
L	24.81	0.846	0.076	0.045	0.318	0.068
M	19.51	0.646	0.180	0.096	0.855	0.183
H	13.94	0.397	0.464	0.336	2.791	0.628
E	10.88	0.297	0.607	0.575	4.246	0.925
*DATUM**						
L	23.43	0.806	0.137	0.026	0.420	0.098
M	19.51	0.646	0.289	0.174	1.225	0.275
H	13.66	0.374	0.533	0.434	3.306	0.689
E	11.24	0.295	0.640	0.634	4.435	0.920
V-cache A5						
L	23.35	0.772	0.135	0.078	0.394	0.119
M	20.85	0.673	0.165	0.090	0.521	0.151
H	17.45	0.514	0.256	0.137	1.025	0.258
E	13.89	0.394	0.454	0.321	2.613	0.579
V-cache A3						
L	23.92	0.788	0.125	0.077	0.359	0.110
M	20.83	0.674	0.151	0.090	0.428	0.127
H	17.17	0.507	0.247	0.140	0.966	0.241
E	14.17	0.406	0.415	0.280	2.220	0.496

Table 7. Evaluation for sine waves. L, M, H, E — low, medium, high, extreme wave amplitude. (*) refers to evaluation on multiple output frames and average the metric.

Setup (sine waves)	**PSNR↑**	**SSIM↑**	**LPIPS$_{VGG}$↓**	**LPIPS$_{Alex}$↓**	**DINO↓**	**CLIP↓**
First frame						
L	21.37	0.716	0.099	0.055	0.413	0.085
M	16.61	0.507	0.213	0.115	0.956	0.209
H	13.41	0.399	0.380	0.276	2.223	0.455
E	10.90	0.320	0.544	0.501	3.639	0.804
*Entire video**						
L	21.38	0.716	0.099	0.055	0.417	0.086
M	16.64	0.506	0.214	0.312	0.962	0.209
H	13.42	0.397	0.381	0.278	2.238	0.452
E	10.81	0.315	0.547	0.507	3.683	0.808
Pixel average						
L	25.82	0.809	0.196	0.172	0.984	0.431
M	20.20	0.568	0.383	0.065	2.217	1.777
H	16.78	0.446	0.521	0.468	3.285	3.110
E	14.18	0.398	0.618	0.636	3.946	4.039
Grid deformation						
L	20.76	0.649	0.326	0.323	1.349	0.351
M	17.24	0.505	0.393	0.366	1.769	0.423
H	14.14	0.417	0.491	0.458	2.746	0.584
E	11.64	0.353	0.595	0.609	3.942	0.842
*Grid registration**						
L	25.84	0.872	0.067	0.040	0.287	0.054
M	21.41	0.739	0.113	0.149	0.478	0.103
H	14.96	0.457	0.327	0.217	1.831	0.383
E	11.20	0.324	0.542	0.490	3.629	0.811
*DATUM**						
L	26.02	0.855	0.121	0.077	0.380	0.084
M	19.68	0.644	0.245	0.087	0.937	0.209
H	14.94	0.441	0.454	0.338	2.515	0.521
E	11.65	0.319	0.613	0.575	3.994	0.840
V-cache A5						
L	22.95	0.747	0.152	0.091	0.501	0.190
M	22.68	0.734	0.150	0.312	0.467	0.137
H	20.54	0.640	0.190	0.105	0.680	0.176
E	15.70	0.489	0.343	0.237	1.834	0.399
V-cache A3						
L	23.87	0.769	0.153	0.093	0.545	0.153
M	23.12	0.756	0.141	0.086	0.430	0.126
H	22.07	0.715	0.150	0.091	0.448	0.128
E	16.70	0.532	0.288	0.198	1.340	0.304

Table 8. Evaluation for ripples waves. L, M, H, E — low, medium, high, extreme wave amplitude. (*) refers to evaluation on multiple output frames and average the metric.

Setup (ripples)	**PSNR↑**	**SSIM↑**	**LPIPS$_{VGG}$↓**	**LPIPS$_{Alex}$↓**	**DINO↓**	**CLIP↓**
First frame						
L	20.66	0.712	0.129	0.077	0.678	0.138
M	16.12	0.502	0.327	0.218	1.865	0.418
H	13.07	0.384	0.521	0.430	3.434	0.746
E	10.49	0.259	0.622	0.604	4.500	0.913
*Entire video**						
L	20.71	0.712	0.127	0.076	0.670	0.137
M	16.15	0.502	0.326	0.217	1.688	0.387
H	13.06	0.383	0.521	0.429	3.345	0.740
E	10.49	0.260	0.621	0.602	4.489	0.910
Pixel average						
L	25.01	0.806	0.214	0.185	1.203	0.222
M	19.69	0.564	0.400	0.326	2.459	0.481
H	16.41	0.443	0.556	0.485	3.665	0.730
E	13.86	0.360	0.642	0.678	4.481	0.885
Grid deformation						
L	21.59	0.682	0.337	0.339	1.485	0.376
M	18.24	0.555	0.433	0.420	2.379	0.541
H	14.37	0.428	0.564	0.561	3.711	0.795
E	11.36	0.321	0.641	0.683	4.678	0.936
*Grid registration**						
L	25.93	0.888	0.063	0.039	0.263	0.056
M	16.89	0.533	0.313	0.196	1.688	0.387
H	13.12	0.380	0.517	0.413	3.345	0.740
E	11.02	0.310	0.571	0.523	3.953	0.857
*DATUM**						
L	22.07	0.753	0.172	0.103	0.591	0.136
M	17.49	0.549	0.362	0.252	1.858	0.403
H	13.88	0.389	0.553	0.471	3.593	0.752
E	11.35	0.258	0.661	0.681	4.629	0.923
V-cache A5						
L	23.69	0.778	0.142	0.081	0.458	0.133
M	21.94	0.708	0.168	0.091	0.567	0.160
H	17.73	0.529	0.323	0.178	1.616	0.380
E	12.40	0.332	0.576	0.527	4.127	0.885
V-cache A3						
L	24.02	0.784	0.142	0.085	0.492	0.138
M	22.79	0.741	0.148	0.083	0.460	0.132
H	18.32	0.567	0.274	0.151	1.235	0.281
E	12.96	0.361	0.538	0.462	3.705	0.775