Supplemental Material for 'Restoration of Non-rigidly Distorted Underwater Images using a Combination of Compressive Sensing and Local Polynomial Image Representations'

Jerin Geo James
IIT Bombay
jeringeo@cse.iitb.ac.in

Pranay Agrawal
IIT Bombay
pranay.agr09@gmail.com

Ajit Rajwade
IIT Bombay
ajitvr@cse.iitb.ac.in

1. Contents of the Supplemental Material

The supplemental material contains this pdf, and also the following folders:

1. A folder 'Collage_MeanImages' containing mean images of videos restored by various methods:

   (a) Our CS method (CS)

   (b) Our Polynomial Expansion Optical Flow (PEOF) method

   (c) Our two stage approach consisting of CS followed by PEOF (CS-PEOF)

   (d) Our two stage approach consisting of CS followed by PEOF and followed by an RPCA step (CS-PEOF-RPCA)

   (e) The two stage method of [3] (SBR-RPCA)

   (f) The deep learning based approach in [2] (DL)

   (g) The method based on learned water bases from [6].

   Besides this, we also have the ground truth (static) image (acquired under perfectly still water in case of Real1, and provided by the authors of [6] in case of Real2), and the mean image of the distorted videos. *All nine images are displayed as a collage for easy comparison.* We also show separate collages containing local SSIM maps superimposed on the mean images. The local SSIM map consists of values of the form $1 - s(i)$ computed at the i^{th} pixel, where $s(i)$ is the local SSIM value. This map is displayed in red color and the bright regions show clearly where the distortion is high. Besides numerical values, *the local SSIM map also shows the better performance of CS, PEOF and CS-PEOF over other methods*.

2. A folder 'Restoration_Videos' containing videos restored by each of the methods. The videos are well annotated, and also contain the mean image with a local SSIM map superimposed. Besides numerical values, *the local SSIM map also shows the better performance of CS, PEOF and CS-PEOF over other methods*.

3. A folder 'CS_MotionReduction' containing a few videos giving a clear idea of the level of motion reduction achieved by the CS step. The folder contains a README.txt file.

4. A folder 'Flow_Visualization_Videos' containing 2 real and 2 synthetic videos, for visual comparison of MVF estimated by CS, PEOF and CS+PEOF. Videos are in form of collages for easy comparison. Along with the flows estimated by different methods, synthetic videos additionally have ground truth flows in the collages. Flow visualization uses known convention from [1]. The folder contains a README.txt file.

This document contains the following sections. A demonstration of the convergence of the COT of the point trajectories is given in Sec. 2. Our acquisition setup for acquiring real videos is detailed in Sec. 3. A comparison between PEOF and the EpicFlow algorithm [4] (a state of the art optical flow method) is given in Sec. 4. A description for the Siamese network architecture for salient feature point tracking is given in Sec. 5.

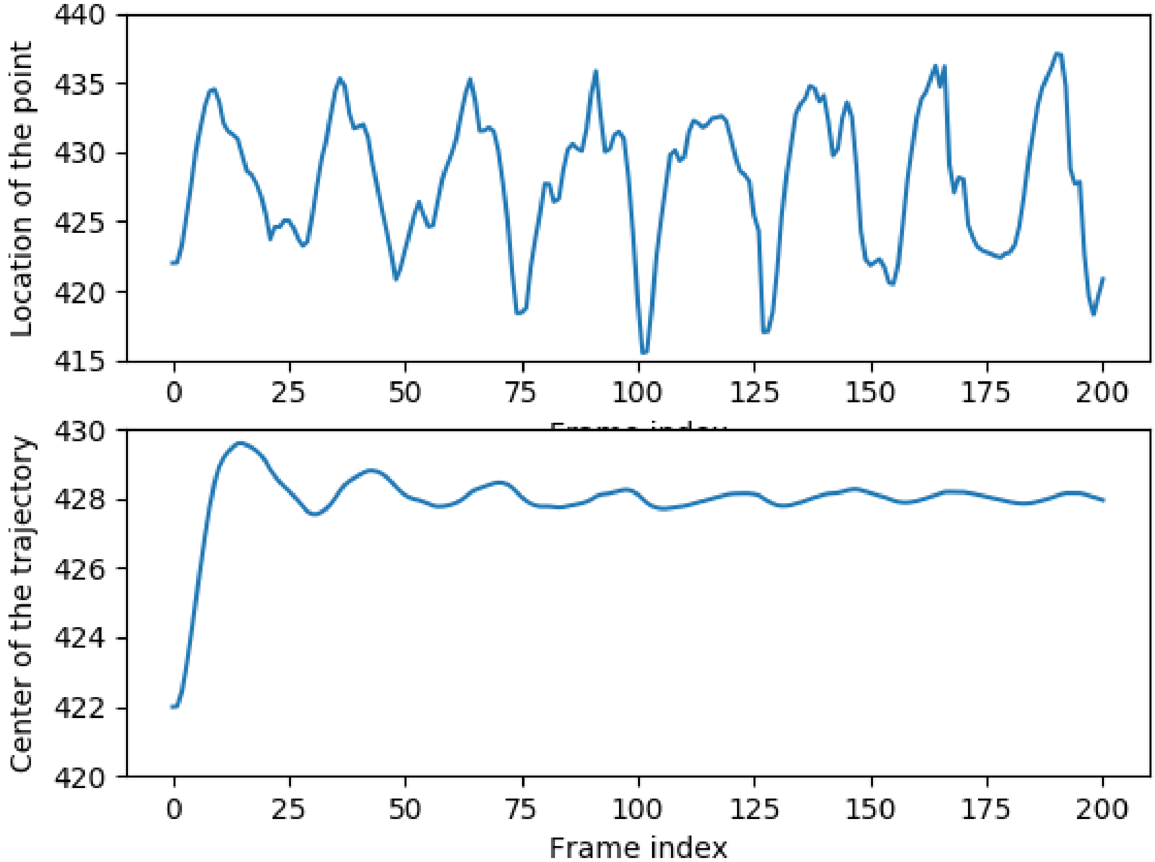

Figure 1. Convergence of COT (second from top) of a salient feature point trajectory (topmost) from a real video sequence

2. Demonstration of Convergence of Center of Trajectory

An example of COT convergence on a real sequence is shown in Fig. 1.

3. Description of Acquisition Setup

We also gathered real video sequences using a standard 50 fps camera from a water tank with facilities for generation of waves using mechanical paddles. The setup shown in the adjoining figure consisted of a trapezoidal water tank of size 4 feet long, 2 feet deep and width ranging from 1.5 feet to 3 feet. It was filled with water up to 25 cm above its flat and horizontal bottom surface. The waves were generated using mechanical paddles at one end of the flume, and also using a rod. We prepared laminated posters of various types of images (texture, text, images of homogeneous and heterogeneous objects). The posters were placed at the bottom of the water-filled flume, and video sequences were gathered from a video camera located 1 metre above the water surface, with its optical axis facing vertically downward. Pictures of our setup can be seen in Fig. 2. Visual inspection revealed that most frames of the acquired videos were quite sharp with blur being only occasionally present. We henceforth refer to this dataset as **Real1**, just as in the main paper. To enable quality assessment, we also acquired a single image of the same poster under still water with the same camera and keeping all settings the same. This produced a ground-truth video denoted as I_0 here.

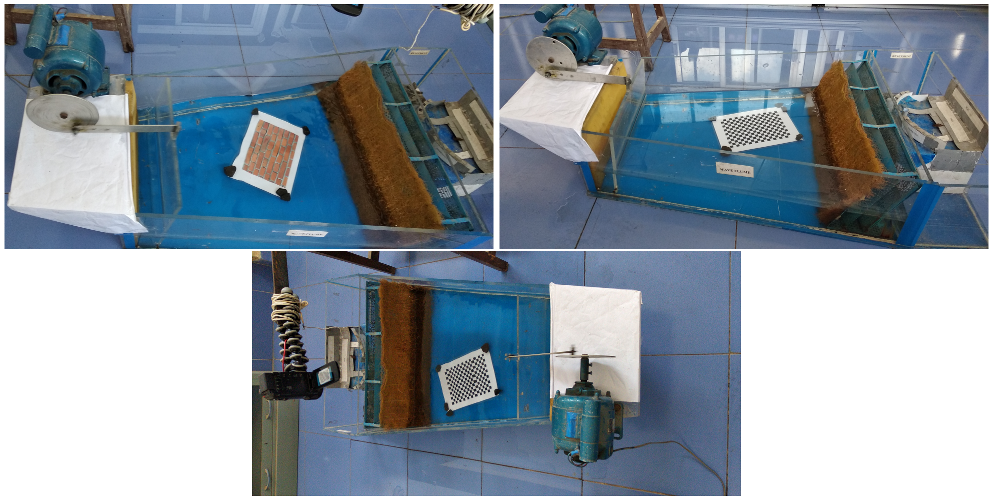

Figure 2. Pictures of the water tank setup, view the bottom-most subfigure by rotating the document clockwise by 90 degrees.

Dataset	**PEOF SSIM**	**EF SSIM**
Real1		
Cartoon	**0.91**	0.85
Checker	**0.89**	0.38
Dices	**0.91**	**0.91**
Bricks	**0.80**	0.75
Elephant	**0.76**	0.66
Eye	**0.98**	0.89
Math	**0.93**	0.90
Synthetic		
BlueTiles	**0.82**	0.29
BrickWall	**0.69**	0.63
Vision	**0.92**	0.40
HandWritten	**0.91**	0.58
Real2		
Middle	**0.85**	0.69
Small	**0.77**	0.61
Tiny	**0.70**	0.28

Table 1. Comparison between PEOF and EpicFlow [4]. Note that CS also clearly outperforms EpicFlow - compare with Table 1 from the main paper.

4. Comparison between PEOF and EF

A comparison between the SSIM values for PEOF and the EpicFlow algorithm (EF) [4], a highly competitive optical flow algorithm, are given in Table 1. The EF method has not been used so far for underwater image restoration, but we are including this comparison for the sake of rigorous comparison. Note that the CS method also outperforms EF.

5. Siamese Network for Salient Feature Point Tracking

We also trained a Siamese network following [5] to learn good feature descriptors. The Siamese network contained two four-layer convolutional neural networks (CNNs) that inferred discriminative feature descriptors from their respective input patches. The training set for this consisted of positive patch-pairs (i.e. structurally similar but spatially distorted w.r.t. each other) and negative patch-pairs (i.e. structurally different). The patch-pairs were extracted from frames of video sequences generated synthetically by simulating refraction of scenes beneath water surfaces modeled as a mixture of periodic waves with randomly generated amplitudes and frequencies. We observed that a network thus trained generally yielded good quality point tracking on unseen synthetic as well as real video sequences. However the performance suffered if there was a moderate amount of blur. Hence we used the KLT (Kanade-Lucas-Tomasi) tracker in the experiments in this paper.

References

[1] S. Baker, D.Scharstein, and J. Lewis. A database and evaluation methodology for optical flow. In *ICCV*, 2007. 1

[2] Z. Li, Z. Murez, D. Kriegman, R. Ramamoorthi, and M. Chandraker. Learning to see through turbulent water. In *WACV*, pages 512–520, 2018. 1

[3] O. Oreifej, G. Shu, T. Pace, and M. Shah. A two-stage reconstruction approach for seeing through water. In *CVPR*, pages 1153–1160, 2011. 1

[4] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid. Epicflow: Edge-preserving interpolation of correspondences for optical flow. In *CVPR*, 2015. 1, 3

[5] E. Simo-Serra, E. Trulls, L. Ferraz, I. Kokkinos, P. Fua, and F. Moreno-Noguer. Discriminative learning of deep convolutional feature point descriptors. In *ICCV*, pages 118–126, 2015. 3

[6] Y. Tian and S. Narasimhan. Seeing through water: Image restoration using model-based tracking. In *ICCV*, pages 2303–2310, 2009. 1