

Deskewing by space-variant deblurring

Karthik Seemakurthy
 karthikvjit@gmail.com

Subeesh Vasu
 ee13d050@ee.iitm.ac.in

A.N. Rajagopalan
 raju@ee.iitm.ac.in

Image Processing and Computer Vision
 Lab
 Department of Electrical Engineering
 IIT Madras
 Chennai, India

Abstract

Skew and motion blur are significant challenges when camera and scene of interest are in two different media. Skew occurs due to spatially varying refraction on a dynamic water surface, whereas motion blur results from multiple intensities impinging on the imaging sensor during camera exposure time due to time varying refraction. In this paper, we introduce the notion of virtual depth map which we assign to a planar scene when observed through a dynamic water surface and transform the deskewing problem into one of space-variant deblurring from a single image within an alternating minimization framework. Since the nature of the virtual depth map can change during exposure due to change in wind properties, we also propose a shot detection framework to identify segments of frames from the captured video which conform to a single virtual depth map. While the overall wave motion can be arbitrary, within each segment the nature of wave is modeled as an exponentially decaying periodic wave.

1 Introduction

The area of underwater imaging has been gaining a lot of importance in the last decade with special emphasis on improving image quality. While degradation of underwater (UW) images can be of different types [12], we specifically deal with skew and motion blur as these are most common. Skew in UW images is caused by the refraction of light rays which results in geometric distortions. During exposure time, each sensor in the camera captures light rays from multiple scene points due to time-varying refraction caused by the dynamic water surface, thus inducing motion blur in the captured image. We address the problem of latent image recovery in the presence of both skew and motion blur.

There are many works in the literature which deal with the problem of deskewing. Murase et al. [8] have used optical flow to reconstruct the water surface and to restore the scene degraded by dynamic water surface. Efros et al. [2] pose reconstruction of images distorted due to moving water waves as a manifold learning problem and pick the best patch at every pixel location from the captured video. Based on bispectral analysis, Wen et al. [10] have proposed a technique to recover UW images affected by skew. The same authors also propose a ‘lucky’ imaging approach in [19]. Tian et al. [11] propose a model-based tracking approach to simultaneously restore the images as well as to extract the water surface. Oreifej et al. [10] propose a two-stage approach. In the first stage, they register every frame with

the mean frame. This is followed by low rank minimization to recover from noise and other errors induced in the first stage. All of these methods (except [10]) assume the video to be captured with high frame rate to avoid motion blur. A very recent work by Seemakurthy et al. [12] can restore images from skew as well as motion blur. However, they cannot handle attenuation effects of water waves.

In this paper, we propose a technique to restore underwater images degraded by attenuated water waves. Following others [10] [11], we too assume that the static camera is looking vertically downwards and imaging a planar scene but through spatially decaying periodic water waves (such as waves due to breeze in shallow water bodies). Also, we allow for the attenuation factor as well as the direction of the water waves to undergo changes during the exposure time of the camera (a situation not handled by [12]). We propose a shot detection method that automatically segments a captured video into groups of frames wherein each group is governed by a dominant water wave direction (i.e. unidirectional) and a single exponential factor of attenuation. Within each segment, we average the frames and show that the blur induced at different pixel locations are scaled versions of each other and model these scale factors through a virtually constructed depth map. Note that despite the scene being planar, the blur induced due to attenuated water waves is space-variant in nature. We pose deskewing of the planar scene as equivalent to a space-variant deblurring problem corresponding to a 3D scene with depth profile being the same as that of the virtual depth map. We propose an alternating framework to solve for the latent image from a single blurred observation. The procedure can likewise be repeated for other segments too, if needed.

The key contributions of our work are summarized below

- As a first of its kind, we introduce the concept of a virtual depth map which enables us to pose the deskewing problem equivalently as a space-variant deblurring problem.
- We reveal that the blur induced by exponentially decaying unidirectional water waves at different spatial locations can be expressed as scaled versions of one another.
- Shot detection module to segment a video into conforming groups of frames.
- An optimization framework for space-variant deblurring to solve for the latent image.

In sections 2 and 3, we propose our methodology for deskewing and deblurring (pertaining to a single segment). Within this segment, we assume that the water waves can be modeled as unidirectional and periodic with exponential decay. The process of shot detection to acquire such segments from a captured video is described in section 4.

2 Relation between wave attenuation and blur

We first describe the nature of blur induced in the camera image plane while imaging a planar scene through spatially decaying periodic water waves. Let \mathbf{f} be the image of the scene captured under still water. Let $I(\mathbf{x}, \tau)$ denote the instantaneous image formed from the light rays which gets projected onto the image plane at time τ in the presence of flowing water. These two images can be related [17] by

$$I(\mathbf{x}, \tau) = f(\mathbf{x} + \mathbf{w}(\mathbf{x}, \tau)) \quad (1)$$

where $\mathbf{w}(\mathbf{x}, \tau)$ is the warp at pixel \mathbf{x} and at time τ , which can be calculated as

$$\mathbf{w}(\mathbf{x}, \tau) = \alpha \bigtriangledown h(\mathbf{x}, \tau) \quad (2)$$

where $\alpha = h_0 \left(1 - \frac{1}{n}\right)$, h_0 is the height of the water column, n is the refractive index of water, and $h(\mathbf{x}, \tau)$ denotes the mathematical representation of dynamic water surface. According to [2] [3], the spatial decaying envelope of water waves can be modeled as

$$h(\mathbf{x}, t) = A(\mathbf{x}) \cos(\mathbf{x} - t) \quad (3)$$

where $A(\mathbf{x})$ denotes the functional form of the envelope. It should be noted that a decaying sinusoidal form is considered only for the sake of mathematical convenience. However, the following analysis is applicable for any spatially decaying periodic wave form. From Eqs. 2 and 3, the warps induced at a pixel \mathbf{x} can be derived as

$$w(\mathbf{x}, t) = \alpha \nabla h(\mathbf{x}, t) = \alpha \sqrt{\nabla A(\mathbf{x})^2 + A(\mathbf{x})^2} \cos(\mathbf{x} - t + \phi) = \alpha^1(\mathbf{x}) p(\mathbf{x}, t) \quad (4)$$

where $\phi(\mathbf{x}) = \tan^{-1} \left(\frac{A(\mathbf{x})}{\sqrt{\nabla A(\mathbf{x})^2 + A(\mathbf{x})^2}} \right)$ and $\alpha^1(\mathbf{x}) = \alpha \sqrt{\nabla A(\mathbf{x})^2 + A(\mathbf{x})^2}$. Since the warp induced at pixel \mathbf{x} is dependent on the nature of the flow gradient, intensities corresponding to different scene points will impinge on a single sensor. This leads to a blurred image \mathbf{g} , which is the average of different skewed versions of the focused observation \mathbf{f} , and can be expressed as

$$g(\mathbf{x}) = \frac{1}{T_e} \int_0^{T_e} f(\mathbf{x} + \mathbf{w}(\mathbf{x}, \tau)) d\tau \quad (5)$$

where T_e is the exposure time of the camera. During the exposure time T_e , let \mathbf{w}_1 and \mathbf{w}_2 be the set of warps at pixel locations \mathbf{x}_1 and \mathbf{x}_2 , respectively. Let \mathbf{p}_1 and \mathbf{p}_2 be the set of values contributed by the periodic component $p(\mathbf{x}, t)$ for one period at \mathbf{x}_1 and \mathbf{x}_2 , respectively, and α_1^1 and α_2^1 be the respective attenuation constants. Then, assuming that one cycle of the periodic component has passed through the camera's field-of-view (FOV), we can conclude that $\mathbf{p}_1 = \mathbf{p}_2$. This is because although the set of values due to a periodic component can be different at a particular instant of time, the sets \mathbf{p}_1 and \mathbf{p}_2 will be the same when one cycle of $p(\mathbf{x}, t)$ is considered [2]. If \mathbf{w}_1 and \mathbf{w}_2 are the set of warps induced at pixels \mathbf{x}_1 and \mathbf{x}_2 then

$$\mathbf{w}_1 = \alpha_1^1 \mathbf{p}_1, \mathbf{w}_2 = \alpha_2^1 \mathbf{p}_2 \implies \mathbf{w}_2 = k \cdot \mathbf{w}_1 \quad (6)$$

where $k = \frac{\alpha_2^1}{\alpha_1^1}$. Eq. 6 implies that the set of warps induced by unidirectional attenuated periodic water waves at any two pixels in the image plane can be related by a scale factor k . The value of k at a pixel location will depend on the form of the water wave attenuation.

2.1 Deskewing through deblurring

As is evident from the above analysis, the set of transformations at any two pixel locations can be related by a scale factor. Interestingly, in the literature corresponding to the restoration of blurred images of 3D scenes due to camera motion [2], it is known that the set of transformations at different pixels induced due to camera motion can be directly related to scene depth. The nature of blur is space-variant. Hence, if we consider the set of all scale factors k in our scenario as virtual depth (\mathbf{V}_d) (although the scene under consideration is planar), the blurred observation generated due to decaying unidirectional water waves can be equivalently generated by a virtual camera motion imaging the scene with depth profile \mathbf{V}_d . The nature of \mathbf{V}_d depends on the nature of the water waves causing the blur.

We choose the form of the attenuation to be exponential which is a reasonable assumption for most practical applications [3] [4]. An exponentially decaying water wave in 2D can be

completely defined using its amplitude, decay factor (d_f), and decay direction ($\mathbf{d}_d \in \mathbb{R}^2$). We model the virtual depth in terms of the virtual depth parameters ($\mathbf{V}_{dp} = (\mathbf{d}_d, d_f)$). The amplitude component of the water wave will act as a common scale factor for the entire depth map and we absorb it as part of the unknown camera motion, i.e. if the amplitude is high, the camera motion covers a wider range of poses and vice versa. We establish this analogy between the transformation space induced due to attenuated water waves and camera motion in order to pose the problem of deskewing as one of space-variant deblurring. The unknowns in our deblurring problem are the blur transformation weights (ω) corresponding to the virtual camera motion, the underlying virtual depth (\mathbf{d}_p), and the latent image (\mathbf{f}).

3 Proposed methodology

In this section, we describe our proposed approach to recover the latent image. The given blurred observation \mathbf{g} can be expressed in terms of the focused observation \mathbf{f} as

$$\mathbf{g} = \mathbf{K}\mathbf{f} \quad (7)$$

where \mathbf{K} is the blurring operator whose columns consist of the point spread function (PSF) corresponding to the pixel position to which it is associated with, while \mathbf{g} and \mathbf{f} denote the lexicographically ordered versions of the blurred and focused observations, respectively. In our case, the \mathbf{K} matrix can be generated using ω and \mathbf{d}_p . In Eq. 7, both \mathbf{K} and \mathbf{f} are unknowns, which makes the problem of deblurring ill-posed. We propose to solve for the unknowns by minimizing the following cost function which builds on the relation in Eq. 7 as

$$E(\hat{\mathbf{f}}, \hat{\mathbf{d}}_p, \hat{\omega}) = \sum_{\delta^* \in \Theta} \lambda_{\delta^*} \|(\delta^* f) o T_{\omega}^{\mathbf{d}_p} - \delta^* \mathbf{g}\|_2^2 + \text{prior}(\mathbf{f}, \mathbf{d}_p, \omega). \quad (8)$$

where $T_{\omega}^{\mathbf{d}_p}$ indicates the transformation function determined by \mathbf{d}_p and ω , the operator o represents space variant convolution by $T_{\omega}^{\mathbf{d}_p}$ on \mathbf{f} and its derivatives ($\Theta = \{\partial_0, \partial_x, \partial_y, \partial_{xx}, \partial_{xy}, \partial_{yy}\}$), and λ_{δ^*} is the associated weight for these terms. Because of the non-convexity of the above cost function, minimizing E is nontrivial. To address this issue, we split this problem into three regularized least square forms, one each for the estimation of \mathbf{f} , \mathbf{d}_p , and ω . These individual problems are then integrated to form an alternating minimization (AM) framework that solves for the unknowns from a single motion blurred image as explained next.

3.1 Estimation of transformation weights

Within the AM framework, this step yields a refined estimate for ω using the current estimates of \mathbf{d}_p and \mathbf{f} . We modify Eq. 8 to obtain the following optimization problem.

$$\hat{\omega} = \arg \min_{\omega} \left\{ \lambda_{\Theta_0}^{\omega} \| \mathbf{f} o T_{\omega}^{\mathbf{d}_p} - \mathbf{g} \|_2^2 + \sum_{\delta^* \in \Theta - \Theta_0} \lambda_{\delta^*}^{\omega} \| (P(\delta^* \mathbf{f})) o T_{\omega}^{\mathbf{d}_p} - \delta^* \mathbf{g} \|_2^2 + \lambda_{\omega} \|\omega\|_1 \right\} \quad (9)$$

Here the operator P refers to the prediction operator which predicts the salient edges of the latent image using bilateral filtering, shock filtering, and gradient magnitude thresholding [10]. This prediction step enables estimation of ω to be relatively stable against estimation errors in latent image thus facilitating convergence. $\lambda_{\delta^*}^{\omega}, \lambda_{\Theta_0}^{\omega}$ are the weights on the respective terms in the data cost. The transformation weight vector induced by unidirectional cyclic waves in the image plane is sparse in nature [10]. Hence, we impose an L_1 prior on ω .

By representing the camera motion to be a vector of weights (ω) corresponding to different poses of the camera, we reduce Eq. 9 as

$$\hat{\omega} = \arg \min_{\omega} \left\{ \|A_f^{\mathbf{d}_p} \cdot \omega - \mathbf{b}_{g_1}\|_2^2 + \lambda_{\omega} \|\omega\|_1 \right\} \quad (10)$$

where $A_f^{\mathbf{d}_p}$ is a matrix whose columns consist of lexicographically ordered warped versions of \mathbf{f} and its derivatives (weighted by $\lambda_{\delta^*}^{\omega}$) corresponding to the poses in ω and known depth \mathbf{d}_p , and \mathbf{b}_{g_1} is formed by stacking corresponding weighted versions of \mathbf{g} .

3.2 Estimation of virtual depth parameters

To estimate \mathbf{d}_p , given the current estimates of ω and \mathbf{f} , we resort to the following formulation.

$$\hat{\mathbf{d}}_p = \arg \min_{\mathbf{d}_p} \left\{ \lambda_{\Theta_0}^{\mathbf{d}_p} \|\mathbf{f} o T_{\omega}^{\mathbf{d}_p} - \mathbf{g}\|_2^2 + \left(\sum_{\delta^* \in \Theta_{-0}} \lambda_{\delta^*}^{\mathbf{d}_p} \|(P(\delta^* \mathbf{f})) o T_{\omega}^{\mathbf{d}_p} - \delta^* \mathbf{g}\|_2^2 \right) + \lambda_{\mathbf{d}_p} \|\mathbf{d}_p\|_1 \right\} \quad (11)$$

Similar to estimation of ω we have observed that the predicted edges can significantly improve the estimation of depth parameters. Here \mathbf{d}_p is a vector of weights corresponding to each virtual depth parameter in the search space. Among all possible values, a unique combination of decaying factor and decaying direction vector will describe the virtual depth. We impose a sparse prior on \mathbf{d}_p to arrive at

$$\hat{\mathbf{d}}_p = \arg \min_{\mathbf{d}_p} \left\{ \|C_f^{\omega} \cdot \mathbf{d}_p - \mathbf{b}_{g_2}\|_2^2 + \lambda_{d_p} \|\mathbf{d}_p\|_1 \right\} \quad (12)$$

Each column in C_f^{ω} is formed by stacking the weighted blurred forms of image f and its predicted gradients, blurred using the corresponding depth parameter from \mathbf{d}_p and known transformation weight ω , and \mathbf{b}_{g_2} is formed by stacking corresponding weighted versions of \mathbf{g} . Although only one single value in $\hat{\mathbf{d}}_p$ is supposed to be nonzero, because of the estimation errors in the known values of ω and \mathbf{f} and the discrete-space approximation for depth parameters, we will have more than one entry in $\hat{\mathbf{d}}_p$ to be nonzero. We treat the depth parameter corresponding to the centroid of these weights as our refined estimate for \mathbf{d}_p . We employ nnLeastR function in SLEP package [10] to solve Eqs. 10 and 12.

3.3 Estimation of latent image

To estimate the latent image from known estimates of ω and \mathbf{d}_p , we minimize

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \left\{ \|\mathbf{f} o T_{\omega}^{\mathbf{d}_p} - \mathbf{g}\|_2^2 + \lambda_{rs} \|M_{rs}(\partial \mathbf{f} - \partial \mathbf{g})\|_2^2 + \lambda_f \|\nabla \mathbf{f}\|_{0.8} \right\} \quad (13)$$

We exclude gradient terms for latent image estimation. Motivated by the works of [11], [12], we impose a sparsity prior on the gradients of the latent image ($L_{0.8}$). Along with a global sparsity prior on the latent image, we include an additional regularization term (with weight λ_{rs}) to suppress ringing artifacts. This local prior is used to enforce the condition that the gradients of both blurred and latent image should be the same in smooth regions. M_{rs} is the weight map highlighting the smooth regions in the image and is found by thresholding the standard deviation of intensity values in \mathbf{g} over a small neighborhood. More details on

this prior and its effectiveness in suppressing ringing artifacts can be found in [16]. The regularized least square form of the above problem can be expressed as follows

$$\hat{\mathbf{f}} = \arg \min_f \left\{ \|\mathbf{B}_\omega^{\mathbf{d}_p} \cdot \mathbf{f} - \mathbf{b}_{g_3}\|_2^2 + \lambda_f \|\nabla \mathbf{f}\|_{0.8} \right\} \quad (14)$$

where $\mathbf{B}_\omega^{\mathbf{d}_p}$ is formed by stacking the blur matrix from known values of \mathbf{d}_p and ω with the weighted gradient operator corresponding to the local prior, while \mathbf{b}_{g_3} is formed by stacking \mathbf{g} with its weighted gradients. We solve Eq. 14 using iteratively re-weighted least squares.

3.4 Alternating minimization for latent image estimation

Our main AM framework consist of two sub problems, one meant for the refinement of ω and the other meant for the refinement of \mathbf{d}_p . In our first subproblem, we use the current estimate of \mathbf{d}_p to estimate ω using Eq. 10. We use this estimated ω along with \mathbf{d}_p to obtain an estimate of latent image. We do this iterative refinement of weight function and latent image until convergence. In the second subproblem, we iterate between depth estimation using Eq. 12 and latent image estimation to obtain a refined estimate for depth. In our main AM framework, we alternate between these two subproblems until convergence to obtain final estimates of \mathbf{d}_p and ω . Eventually, we perform non-blind deblurring to yield the desired latent image. The main steps of our scheme are listed in Algorithm 1.

Algorithm 1 Algorithm for deskewing.

Input: Blurred image (\mathbf{g}), initial estimates of virtual depth parameters (\mathbf{d}_{p_0}) and latent image (\mathbf{f}_0).

Output: Deskewed image.

- 1: $\hat{\mathbf{d}}_p \leftarrow \mathbf{d}_{p_0}$, $\hat{\mathbf{f}} \leftarrow \mathbf{f}_0$
- 2: **repeat**
- 3: **repeat**
- 4: $\mathbf{d}_p \leftarrow \hat{\mathbf{d}}_p$, $\mathbf{f} \leftarrow \hat{\mathbf{f}}$, and estimate $\hat{\omega}$ by using \mathbf{d}_p and \mathbf{f} in Eq. 10.
- 5: $\omega \leftarrow \hat{\omega}$ and estimate $\hat{\mathbf{f}}$ by using ω , \mathbf{d}_p in Eq. 14.
- 6: **until** convergence of $\hat{\omega}$
- 7: **repeat**
- 8: $\mathbf{f} \leftarrow \hat{\mathbf{f}}$, $\omega \leftarrow \hat{\omega}$, and estimate $\hat{\mathbf{d}}_p$ by using \mathbf{f} , ω in Eq. 12.
- 9: $\mathbf{d}_p \leftarrow \hat{\mathbf{d}}_p$ and estimate $\hat{\mathbf{f}}$ using \mathbf{d}_p and ω in Eq. 14.
- 10: **until** convergence of $\hat{\mathbf{d}}_p$
- 11: **until** convergence of $\hat{\mathbf{d}}_p$ and $\hat{\omega}$
- 12: Non-blind deblurring using final estimates of \mathbf{d}_p and ω .

3.5 Initialization of virtual depth parameters and latent image

A good initialization scheme is essential for our AM framework. We first estimate the PSFs from overlapping local patches located all over the image using the state-of-the-art space-invariant deblurring approach of [21]. Although the image as a whole is space variantly blurred, we can treat local patches in the image to have space invariant blur since depth variations are typically smooth. Let \mathcal{P} be the set of all estimated PSFs. Because these estimates can be in error, we subject them all to multiple outlier rejection stages to enable us

to identify reliable PSFs. In our scenario, the variation of PSFs is smooth and hence in the first stage of rejection we calculate the entropy for each of the PSFs and compare them with the mean of entropies of the neighborhood PSFs. If the difference is more than a threshold (0.25 times the mean of all neighborhood entropies) we declare them as outliers. This scheme is used by the authors in [2] and we have found it to be effective in our case also. Let \mathcal{P}_1 be the set of all inlier PSFs after the first stage of rejection. Due to the unidirectional nature of water waves (within a single segment), the PSFs induced at different locations in the image plane are dominant in one direction. We compute the dominant direction for each of the PSFs in \mathcal{P}_1 using weighted PCA and build a histogram of orientations. We use the centroid of the bin corresponding to the peak value of the histogram as an initial estimate for \mathbf{d}_d . All the PSFs for which the dominant direction is away from \mathbf{d}_d (i.e greater than a particular threshold in the Euclidean sense) are removed from the set \mathcal{P}_1 to obtain a refined set of PSFs \mathcal{P}_2 . The kernels which are along a direction perpendicular to the water flow and at a particular distance from the origin of the water waves should be the same. We use this fact to filter the PSFs in \mathcal{P}_2 by comparing the entropy with the neighborhood defined by \mathbf{d}_d , resulting in set \mathcal{P}_3 . By using weighted PCA on all the PSFs, we project them onto their respective dominant directions to filter out estimation errors. In \mathcal{P}_3 , we consider the kernel whose length is closer to the median length with direction closest to the dominant direction as the reference kernel and find the scale of every other kernel. By employing RANSAC on these scale estimates, we fit an exponential curve to find an initial estimate of the decay factor (d_f). Assuming the reference PSF as the initial value of the transformation weights and by using the virtual depth parameters, we solve for an initial estimate of \mathbf{f} using Eq. 14. (Please refer to supplementary material for a visual illustration of refinement of PSFs).

4 Shot detection

It is important to note that the scaling relationship among PSFs at multiple locations holds true if there exists a single virtual depth map induced by the attenuation of water waves. In the presence of longer exposures, the parameters of the induced virtual depth map may vary because of changes in wind direction and/or speed. To overcome this challenge, we capture a video of long duration and propose a shot detection scheme to select a set of frames that can be modeled by a single virtual depth map with minimum model fitting error (E_r).

For shot detection, we analyze E_r for stacks of frames corresponding to different time windows from the video, and identify an optimal time window over which the water flow best fits our model. We perform this in a scale-space fashion, in which for each level we estimate E_r for time windows of different length. We start by analyzing small overlapping time windows all over the video (say level 1). We average the frames in each of these time windows to generate corresponding blurred images. Using the procedure described for initialization (section 3.5), we find E_r for each blurred image. The error in the exponential fitting of all inliers from RANSAC, averaged over the total number of inliers is used as E_r . A minimum value of E_r represents the best model fitting time window for each level. This process is repeated for larger time windows until the minimum fitting error for the current level starts increasing significantly as compared to the previous level. The time window which gives minimum error among all the levels is chosen as our optimal set of frames. In practice, we can have multiple non-overlapping stack of frames wherein each group is governed by a dominant water wave direction and a single exponential factor of attenuation. In that case, one can choose the non-overlapping stacks having E_r below a threshold to yield

multiple latent image estimates. Empirically we have observed that choosing the smallest time window to be 1 second gives reliable performance. In our real experiments, we found that after 2 to 3 levels, the modeling error starts to increase consistently indicating a change in the wave nature. More details on shot-detection can be found in the supplementary material.

5 Experimental Results

We compare our latent image estimation results with the state-of-the-art space-variant deblurring approach [23] along with state-of-the-art deskewing methods [10], [11]. The input to [10] [11] should be in the form of video, while a single blurred image is used for [23] and our proposed algorithm. For quantitative analysis, we use peak signal to noise ratio (PSNR), structure similarity index measure (SSIM) [18], and normalised mutual information (NMI) [10]. The parameter values used in our experiments can be found in the supplementary material. Although, we can have multiple latent image estimates from a single video, we only show the estimate corresponding to minimum E_r .

5.1 Synthetic Examples

We use the model described in [8] to simulate a decaying envelope of the water waves which is given as $h(\mathbf{x}, t) = A(\mathbf{x}) \sin(\omega_x \mathbf{x} + \omega_y \mathbf{y} - t)$, where $A(\mathbf{x})$ denotes the decaying amplitude of the sinusoid, ω_x and ω_y are the spatial frequencies. In all our experiments, the decaying function is given by $A(\mathbf{x}) = A_0 \exp(d_f \cdot (\mathbf{d}_d^T \mathbf{x}))$. The first column of Fig. 1 shows the ground truth observations. For the synthetic examples in the first two rows of Fig. 1, we generated 512 skewed observations. In order to synthesize a practical scenario consisting of change in wind properties, we either changed d_f , A_0 or \mathbf{d}_d after the 256th frame. For the *wall* image, we changed \mathbf{d}_d from $[1 \ 0]^T$ to $[0 \ 1]^T$, while for the *text* example we increased the decay factor from -0.004 to -0.002 after the 256th frame. The value of $A_0 = 9$ was kept constant for these examples. We use the shot detection procedure described in section 4 to segment the video. The blurred image produced from one such segment is given as input to our algorithm (Fig. 1 (b)). It can be clearly seen in the highlighted red and green boxes of Fig. 1 (c,d), that both the state-of-the-art deskewing methods [10] [11] could not completely restore the observations from skew. The main reason behind the failure of [10] is that the water bases may not have been able to model the waves adequately, while the failure of [11] can be attributed to poor performance of their non-rigid registration framework. Fig. 1 (e) shows the failure of the space-variant deblurring approach of [23], mainly because due to the fact that it does not handle blur changes induced by depth variations. The last column gives the output of our method. As can be seen from the highlighted regions, the restored observations are sharper and are almost free of skew.

5.2 Real Examples

We used laminated textured sheets kept at the bottom of an aquarium. The field-of-view was kept large enough to witness the effect of wave attenuation in the captured video. The source of waves is a fan that acts as a wind blower. The shell example in Fig. 1 (third row) was captured with change in blower speeds to introduce a change in wind amplitude and spatial frequency. The stones example (Fig. 1, fourth row) was captured by a change in the origin of waves to synthesize change in wind direction. Both the videos were given as input to our

shot detection framework to segregate groups of frames that conform to a single virtual depth map. A group of frames which conform to a single virtual depth map with least fitting error is given as input to [10], while the average of these frames was given to [23] and our own method. Since both [10] and [23] do not account for the blur in individual frames, their deskewed outputs are not sharp (Fig. 1 (c,d)). Fig. 1 (e) shows that [23] also fails to recover a sharp estimate of the latent image. The results of our proposed approach (Fig. 1 (f)) are closer to the ground truth (captured in still water) as compared to competing algorithms. Table 1 gives a quantitative analysis for both synthetic and real examples. Clearly, our proposed approach outperforms other methods in objective evaluation too.

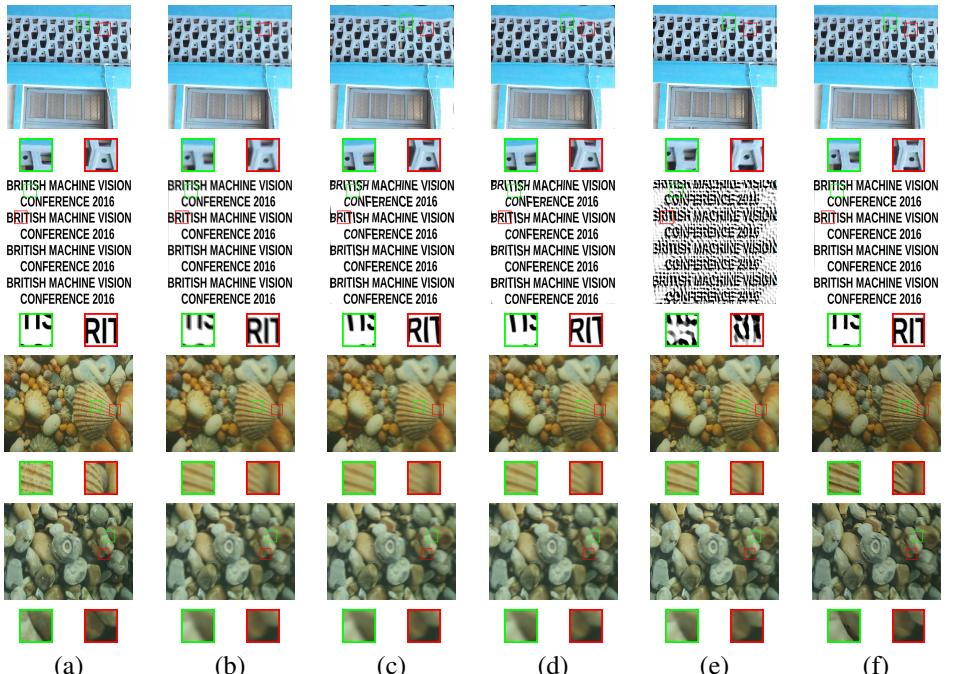


Figure 1: Synthetic and real experiments. (a) Focused observation, (b) blurred observation, (c) restored by [10], (d) output of [10], (e) output from [23] and (f) proposed method. First row: wall. Second row: text. Third row: shell. Fourth row: stones.

Table 1: Quantitative analysis for synthetic and real experiments.

	PSNR (in dB)				SSIM				NMI			
	Tian [10]	Oreifej [10]	[10]	Algorithm I	Tian [10]	Oreifej [10]	[10]	Algorithm I	Tian [10]	Oreifej [10]	[10]	Algorithm I
Wall	20.9638	22.4273	13.1212	22.8546	0.9994	0.9996	0.9943	0.9997	1.0626	1.0148	0.8994	1.0713
Text	13.5681	13.5679	11.1212	20.8162	0.9975	0.9975	0.9978	0.9997	0.5770	0.5766	0.5290	0.6107
Shell	26.3070	26.6811	23.6328	26.7811	0.9984	0.9983	0.9986	0.9999	0.9760	0.9544	0.9544	0.9820
Stones	23.4497	28.2615	27.0457	28.2815	0.9995	0.9994	0.9993	0.9998	0.9727	1.0248	1.0045	1.0250

We additionally performed outdoor experiments in swimming pool. Here the source of waves was the breeze as well the flow due to water circulation. Since we do not have control over natural breeze, we restrict ourselves to qualitative comparisons. It can be seen from Fig. 2 that the results of our proposed method for both shell as well as band examples are sharper when compared to the state-of-the-art (see highlighted red and green boxes).

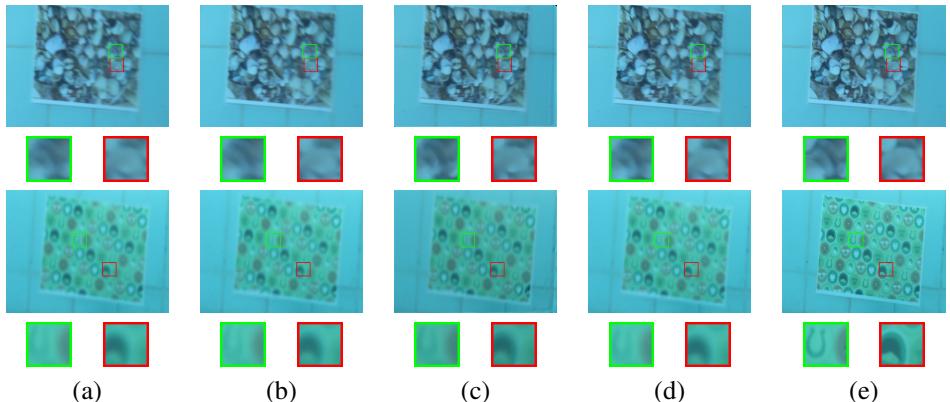


Figure 2: Real outdoor experiment. (a) Blurred image. Output of: (b) [17], (c) [10], (d) [24], and (e) proposed method. First row: shell. Second row: band.

6 Conclusions

We proposed a deskewing technique to restore planar scenes distorted by water waves. Given a video, we proposed a shot detection framework to segment it into groups of frames that conform to a virtual depth map. The frames within a segment are governed by unidirectional attenuated waves. We established that, within a segment, the blur induced at different spatial locations of the image, can be expressed as scaled versions of each other. The notion of a virtual depth map enabled us to pose the problem of deskewing as one of space-variant deblurring. An alternating minimization framework was proposed to solve for the latent image wherein the virtual depth parameters and transformations induced in the image plane were obtained as by-products. Our method outperforms the state-of-the-art as is evident from our comparisons with competing methods.

Acknowledgment

Financial assistance provided by BRNS, Dept. of Atomic Energy, India through project sanction No.2012/21/14/BRNS/1681 is gratefully acknowledged.

References

- [1] Sunghyun Cho and Seungyong Lee. Fast motion deblurring. In *ACM Transactions on Graphics (TOG)*, volume 28, pages 145:1–145:8. ACM, 2009.
- [2] Alexei Efros, Volkan Isler, Jianbo Shi, and Mirkó Visontai. Seeing through water. In *Advances in Neural Information Processing Systems*, pages 393–400, 2005.
- [3] Mark D Groves. Steady water waves. *Journal of Nonlinear Mathematical Physics*, 11(4):435–460, 2004.

- [4] Anat Levin and Yair Weiss. User assisted separation of reflections from a single image using a sparsity prior. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (9):1647–1654, 2007.
- [5] Anat Levin, Rob Fergus, Frédéric Durand, and William T Freeman. Image and depth from a conventional camera with a coded aperture. In *ACM Transactions on Graphics (TOG)*, volume 26, pages 70:1–70:10. ACM, 2007.
- [6] An-Kuo Liu and Stephen H Davis. Viscous attenuation of mean drift in water waves. *Journal of Fluid Mechanics*, 81(01):63–84, 1977.
- [7] J. Liu, S. Ji, and J. Ye. SLEP: Sparse learning with efficient projections, 2009. URL <http://www.public.asu.edu/~jye02/Software/SLEP>.
- [8] B.L. Méhauté, B. LeMéhauté, and S. Wang. *Water Waves Generated by Underwater Explosion*. Advanced series on ocean engineering. World Scientific, 1996. ISBN 9789810220839. URL <https://books.google.co.in/books?id=6eivLpwrvtQC>.
- [9] Hiroshi Murase. Surface shape reconstruction of a nonrigid transport object using refraction and motion. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (10):1045–1052, 1992.
- [10] Omar Oreifej, Guang Shu, Teresa Pace, and Mubarak Shah. A two-stage reconstruction approach for seeing through water. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1153–1160. IEEE, 2011.
- [11] Chandramouli Paramanand and Ambasamudram Rajagopalan. Non-uniform motion deblurring for bilayer scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1115–1122, 2013.
- [12] O. M. Phillip. *The Dynamics of the Upper Ocean*. CUP Archive. URL <https://books.google.co.in/books?id=fYk6AAAAIAAAJ>.
- [13] Daniel Rueckert, Luke I Sonoda, Carmes Hayes, Derek L. G. Hill, Martin O. Leach, and David J. Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *Medical Imaging, IEEE Transactions on*, 18(8):712–721, 1999.
- [14] Raimondo Schettini and Silvia Corchs. Underwater image processing: state of the art of restoration and image enhancement methods. *EURASIP Journal on Advances in Signal Processing*, 2010:1–14, 2010.
- [15] Karthik Seemakurthy and Ambasamudram Narayanan Rajagopalan. Deskewing of underwater images. *Image Processing, IEEE Transactions on*, 24(3):1046–1059, 2015.
- [16] Qi Shan, Jiaya Jia, and Aseem Agarwala. High-quality motion deblurring from a single image. In *ACM Transactions on Graphics (TOG)*, volume 27, pages 73:1–73:10. ACM, 2008.
- [17] Yuandong Tian and Srinivasa G Narasimhan. Seeing through water: Image restoration using model-based tracking. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2303–2310. IEEE, 2009.

- [18] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004.
- [19] Zhiying Wen, Donald Fraser, and Andrew Lambert. Bicoherence: a new lucky region technique in anisoplanatic image restoration. *Applied optics*, 48(32):6111–6119, 2009.
- [20] Zhiying Wen, Andrew Lambert, Donald Fraser, and Hongdong Li. Bispectral analysis and recovery of images distorted by a moving water surface. *Applied optics*, 49(33):6376–6384, 2010.
- [21] Li Xu and Jiaya Jia. Two-phase kernel estimation for robust motion deblurring. In *Computer Vision–ECCV 2010*, pages 157–170. Springer, 2010.
- [22] Li Xu and Jiaya Jia. Depth-aware motion deblurring. In *Computational Photography (ICCP), 2012 IEEE International Conference on*, pages 1–8. IEEE, 2012.
- [23] Li Xu, Shicheng Zheng, and Jiaya Jia. Unnatural l0 sparse representation for natural image deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1107–1114, 2013.