

Mosaicing deep underwater imagery

Kuldeep Purohit, *, Subeesh Vasu, *, A.N. Rajagopalan*, V Bala Naga Jyothi†, Ramesh Raju †

*Department of Electrical Engineering, Indian Institute of Technology Madras.

† National Institute of Ocean Technology, Chennai

This is a draft version of the paper that got accepted in Indian Conference on Computer Vision, Graphics and Image Processing, (ICVGIP) DECEMBER 2016. The final copyrighted version is available at
<http://dl.acm.org/citation.cfm?id=3010029/>

Mosaicing Deep Underwater Imagery

Kuldeep Purohit

IIT Madras

kuldeephurohit3@gmail.com

Subeesh Vasu

IIT Madras

subeeshvasu@gmail.com

A. N. Rajagopalan

IIT Madras

raju@ee.iitm.ac.in

V Bala Naga Jyothi

National Institute of Ocean
Technology, Chennai
nagajyothi@niot.res.in

Ramesh Raju

National Institute of Ocean
Technology, Chennai
ramesh@niot.res.in

ABSTRACT

Numerous sources of distortions render mosaicing of underwater (UW) images an immensely challenging effort. Methods that can process conventional photographs (terrestrial/aerial) fail to deliver the desired results on UW images. Taking the sources of underwater degradations into account is central to ensuring quality performance. The work described in this paper specifically deals with the problem of mosaicing deep UW images captured by Remotely Operated Vehicles (ROVs). These images are mainly degraded by haze, color changes, and non-uniform illumination. We propose a framework that restores these images in accordance with a suitably derived degradation model. Furthermore, our scheme harnesses the scene geometry information present in each image to aid in constructing a mosaic that is free from artifacts such as local blurring, ghosting, double contouring and visible seams. Several experiments on real underwater images sequences have been carried out to demonstrate the performance of our mosaicing pipeline along with comparisons.

CCS Concepts

- Computing methodologies → Image processing; Reconstruction;

Keywords

Mosaic; dehazing; illumination correction; scattering, underwater.

1. INTRODUCTION

Underwater imaging has been an active area of research for the past few decades and has been quite successfully applied in shallow water scenarios.

Due to gradually increasing water pressure, it is not possible for humans to reach depths beyond a certain limit (60m) to capture images. Hence it is imperative to utilize ROVs

[21] to cover greater depths. ROVs (see Fig. 1) are commonly used for UW survey missions such as detection of obstructions that can be a hazard to navigation, surveying of oceans beds etc.

This is typically achieved by building large image mosaics. However, the uncontrolled imaging conditions introduced by these automated systems coupled with challenges posed due to physics of image formation in water necessitates sophisticated image processing for mosaicing to succeed.

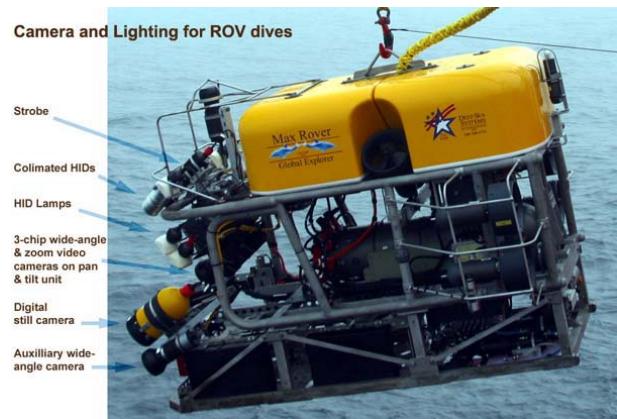


Figure 1: Light strobes and cameras on the Global Explorer ROV, a highly-mobile ROV designed and developed by Deep Sea Systems Inc. (Image courtesy of Arctic Exploration 2002, Ian MacDonald, Texas A & M, NOAA/OER.)

UW imaging is quite challenging due to the presence of numerous degradation sources which are very different from the terrestrial scenario. Often, there is need for a customized imaging model to retrieve useful information. Different colors of visible light spectra encounter different levels of attenuation while travelling through water [25]. The decay coefficients increase as we go deeper into sea. In fact, after a certain depth, the only visible light appearing in the medium is the blue color. For many applications, this effect warrants channel-wise treatment of UW images, unlike in conventional photography.

In ROV imaging, to acquire details, the camera is kept quite close ($< 5m$) to the seabed. The backscatter is minimized by reducing the overlap between the areas common to the camera's field of view and the light source's field of effect (typically from LEDs) using directional lighting. This

ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of a national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

ICVGIP '16 Guwahati, Assam India

© 2016 ACM. ISBN 978-1-4503-4753-2/16/12...\$15.00

DOI: <http://dx.doi.org/10.1145/3009977.3010029>

in turn causes non uniform illumination effects in the captured images. This degrading effect must be accounted for, especially in applications involving multiple images of the same scene.

Light scattering caused by suspended particles can lead to haziness in UW images which limits visibility and results in lower contrast. Visibility cannot be improved by additional artificial Strobe lighting since it increases backscattering, filling the image with more noise [13]. The small distance between the camera and scene leads to significantly less coverage of area in a single image. To extract meaningful information, one must build a mosaic from multiple images. However the proximity of the 3D scene and the camera can result in unavoidable parallax effects in the images. While traditional image mosaicing algorithms ([28]) align images using a single projective warp, such a model is not applicable when there are significant occlusion effects and depth variations ([30], [17], [31]). Without the knowledge of the 3D structure of the scene, it becomes challenging to account for parallax effects. The misalignments caused by variations in depth lead to blurring and double-contouring in the overlapping regions of the rendered mosaics.

Photomosaicing softwares (such as [4]) have become commonplace in the last decade. While they offer a wide angle view, their applicability is restricted mostly to terrestrial imaging. The problems encountered in mosaicing terrestrial, aerial and space images while accounting for motion and exposure variation are widely covered in the literature [12]. But porting these algorithms to UW imagery is a non-trivial task because UW offers a multitude of challenges.

In this paper, we address the problem of deep UW image mosaicing in the presence of degradations due to color cast, non-uniform illumination and haze. We derive image formation model for such a scenario and perform dehazing to obtain an enhanced set of input images for mosaicing. Since dehazing by itself cannot solve the problem of visibility in practical scenarios, we propose to supplement it with illumination compensation and color restoration by solving for different illumination maps in a channel-wise fashion. We estimate the transmission map from the green-blue channels of illumination-compensated images, and use it as a cue for scene depth to perform a depth-dependent spatially varying projective warp to align input images. Finally, we estimate optimal seams to combine aligned images with minimum intensity transitions across the borders using α -blending.

The main contribution of our work is in proposing a new mosaicing pipeline specially tailored for typical deep UW image sequences captured using an ROV. It involves modifying the classical underwater imaging model to reflect deep underwater image scenario. In the process, we advocate a depth-based stitching algorithm, which can be applied in general for mosaicing hazy images of 3D scenes.

2. RELATED WORKS

In this section, we give a brief overview of existing works that deal with UW degradation as well as UW image stitching. Since the degradation process in UW scenes is both multiplicative and additive [25], traditional enhancing techniques such as white balancing, histogram equalization etc. have shown strong limitations for such a task. [26] employed local histogram equalization to compensate for nonuniform illumination. The work in [27] used an adaptation of homomorphic filtering to remove illumination effects. [29] used

the sparseness of the gradients of image as a prior for estimating illumination map. Assuming a well-lit circular region in the middle of the image with a poorly illuminated area surrounding it, [22] used the illumination estimates obtained by selectively averaging multiple images, to compensate for nonuniform illumination. A survey on existing methodologies which handle low contrast and nonuniform illumination effects in UW images is presented in [11].

Among the works on dehazing and color correction, an algorithm for removing the effects of light scattering in UW images is presented in [5]. They perform simultaneous color correction and dehazing since both are inter-related. Chiang et al. [6] proposed an algorithm for wavelength compensation and dehazing in order to remove distortions caused by light scattering and color change, respectively. A single image based dehazing method which uses the depth estimate obtained by exploiting the differences in the attenuation of the color channels was presented in [5]. In [16], color and illuminations artifacts are removed using the attenuation coefficients derived from a sequence of overlapping UW images. A methodology to estimate transmission map from UW images was proposed by Drews et al. [7]. They have adapted the dark channel prior (DCP) for UW scenario by giving importance to blue and green channels. Galdran et al. [10] have proposed a red channel method for color correction and contrast improvement of UW images.

Few works also exists that have addressed stitching and/or alignment of UW images. Singh et al. [26] have presented a general framework for stitching large areas in UW by handling the presence of moving non-uniform lighting source, scattering of light waves etc. They have given emphasis to global consistency of the result. The work in [23] focuses on real time performance of mosaicing rather than dealing with distortion problems in UW imaging. The work in [9] produces a globally aligned photo-mosaic from multiple UW images by making use of camera position estimates obtained from navigation sensors. [3] uses bundle adjustment on visual features along with trajectory estimates from navigation data to reconstruct UW structures with good accuracy. In the UW giga-mosaicing work by Prados et al. [22], a novel blending approach was presented to remove photometric and geometric disparities in overlapping regions. They have also performed enhancement of input images through a number of preprocessing steps to reduce the impact of different degradations on the input images.

3. IMAGE FORMATION MODEL

The image formation model proposed in [15] and its simplified versions have been used in the last decade to develop practical restoration algorithms for UW images. The model incorporates two main sources present in the water, that contribute to the image. The first is the radiance of the scene object which reaches the camera after partial absorption and scattering. The second source of light is the medium which scatters a multitude of light rays travelling through it and directs them towards the camera. We express the captured image I as a superposition of a direct component E_d and a back scattering component E_b [6] as

$$I(x) = E_d(x) + E_b(x) \quad (1)$$

where x represents the spatial location of a pixel. For every pixel, both components are dependent on scene depth

s and the color channel λ . According to the Beer-Lambert law [8], transmission exponentially decreases with respect to distance from camera. Hence, the model for representing direct component E_d becomes

$$E_d(s, \lambda) = J(\lambda) \exp(-\alpha(\lambda)s) \quad (2)$$

where λ is the wavelength, and α is the attenuation coefficient given by $\alpha = a + b$, with a being the absorption coefficient and b the total scattering coefficient of the water. $J(\lambda)$ is the original image that would have been captured if there was no participating medium (water) present between the camera and the scene. These variables are all functions of the wavelength λ .

While the above model assumes a homogeneous light source present in the water, our considered scenario is limited by artificial light source. Accounting for the fact that the artificial light source and the camera are mounted on the same vehicle and hence are equidistant from the object (as depicted in Fig. 3), a light ray originating from the source and reflected from the object reaches the camera after traveling a distance of $2s$. Thus, the direct component E_d becomes

$$E_d(s, \lambda) = J(\lambda) \exp(-\alpha(\lambda)2s) \quad (3)$$

The back scattering component E_b can come from several sources including, direct sunlight, diffused skylight and light reflected by the earth surface. We now build upon the derivation of this component as a function of path-length [24]. When a light source with intensity I^S illuminates particles along the line of sight (LOS), its contribution to backscatter E_b is given by

$$E_b(s) \approx kb \int_0^s I^S(l) \exp(-\alpha l) dl \quad (4)$$

where k is a constant parametrised by the focal length of the camera. This integral accounts for scattering into the LOS at some distance l , followed by attenuation until it reaches the camera. The above equation says that we can recover the backscatter by integrating the attenuated field along a camera ray. Most of the existing works proposed to model scenes with presence of natural light distributed uniformly along the LOS i.e. $I^S(l) = I_0^S$ is independent of l . This assumption simplifies the above equation to

$$E_b(s) \approx kbI_0^S \int_0^s \exp(-\alpha l) dl \quad (5)$$

which further reduces to

$$E_b(s) = A(\lambda)(1 - \exp(-\alpha s)) \quad (6)$$

where $A(\lambda) = kI_0^S b / \alpha(\lambda)$ is a constant vector conventionally referred to as air-light [20].

While surveying in deep oceans, the light existing in the scene $I^S(l)$ originates from an artificial light source and its magnitude decreases as the distance l from the light source increases. Hence the assumption of homogeneous I^S does not hold good any more. We replace I^S by $I_0^S \exp(-\alpha(l))$ to obtain the modified expression

$$E_b(s) = \int_0^s I_0^S \exp(-\alpha l) \exp(-\alpha l) dl \quad (7)$$

or equivalently

$$E_b(s, \lambda) = A(\lambda)(1 - \exp(-2\alpha(\lambda)s)) \quad (8)$$

From Eq. 1, Eq. 3, and Eq. 8 the total light that reaches the camera from a point at distance L from the light source can be written as

$$I(s, \lambda) = J(\lambda) \exp(-2\alpha(\lambda)s) + A(\lambda)(1 - \exp(-2\alpha(\lambda)s)) \quad (9)$$

The image degradation model in deep UW scenarios with artificial light can be considered equivalent to the haze model in atmospheric images $I(x) = J(x)t(x) + A(x)(1 - t(x))$, but with depth value that is twice the actual depth. Another difference is that the transmission map is different for each channel. This mathematical similarity to atmospheric-haze model enables us to utilize approaches based on the well-known dark-channel prior for ROV image restoration.

4. ILLUMINATION COMPENSATION

To obtain a photometrically consistent result, the images input to the mosaicing algorithm should be consistent in color and illumination. State-of-the-art dehazing approaches assume that the illumination is constant in the scene. When this assumption is violated, the airlight cannot be considered to be constant. The result is a transmission underestimate or overestimate in unevenly illuminated areas, and color distortions characterized by dark blue regions appearing in the restored images. This necessitates non-uniform illumination correction before de-hazing. Since errors in estimated illumination could seriously hamper mosaicing performance due to errors in feature matching and blending, we port the robust parametric approach of [29] to UW images. Their model describes the input image z as the product of the non-uniformity free intensity image i and an illumination map m , $z(x) = i(x)m(x)$. Instead of computing m directly, a common approach is to solve for its logarithm. If $Z = \ln(z)$, $I = \ln(i)$, and $M = \ln(m)$, we have $Z(x) = I(x) + M(x)$.

Let the gradients of Z , I , and M for each pixel x be $\psi^Z(x)$, $\psi^I(x)$, and $\psi^M(x)$, respectively. The following relation holds due to linearity of gradient operator.

$$\psi^Z(x) = \psi^I(x) + \psi^M(x) \quad (10)$$

From Eq. 10, M can be estimated as a maximum a posteriori (MAP) solution. Using Bayes' rule, this amounts to solving the optimization problem

$$M = \arg \max_M P(M|Z) \propto \arg \max_M P(Z|M)P(M) \quad (11)$$

To compute $P(Z|M)$, a sparsity prior on the image gradient distribution is imposed as below

$$P(Z|M) = P(\psi^I) = \exp -|\psi^I|^\alpha = \exp - \sum_{(x)} |\psi^{Z(x)} - \psi^{M(x)}|^\alpha \quad (12)$$

where $\alpha < 1$. To represent the non uniform field M , it is assumed that it can be well-approximated by a bivariate polynomial, for which the model of degree D is

$$M(x) = \sum_{t=0}^D \sum_{l=0}^t a_{t-l,l} p^{t-l}(x) q^l(x) \quad (13)$$

Here p and q are the row and column index of the pixel. Hence, maximizing the probability is the same as minimizing

$$O = \sum_{(i,j)} |\psi^Z(x) - \psi^M(x)|^\alpha + \sum_{t=0}^D \sum_{l=0}^t a_{t-l,l} \quad (14)$$

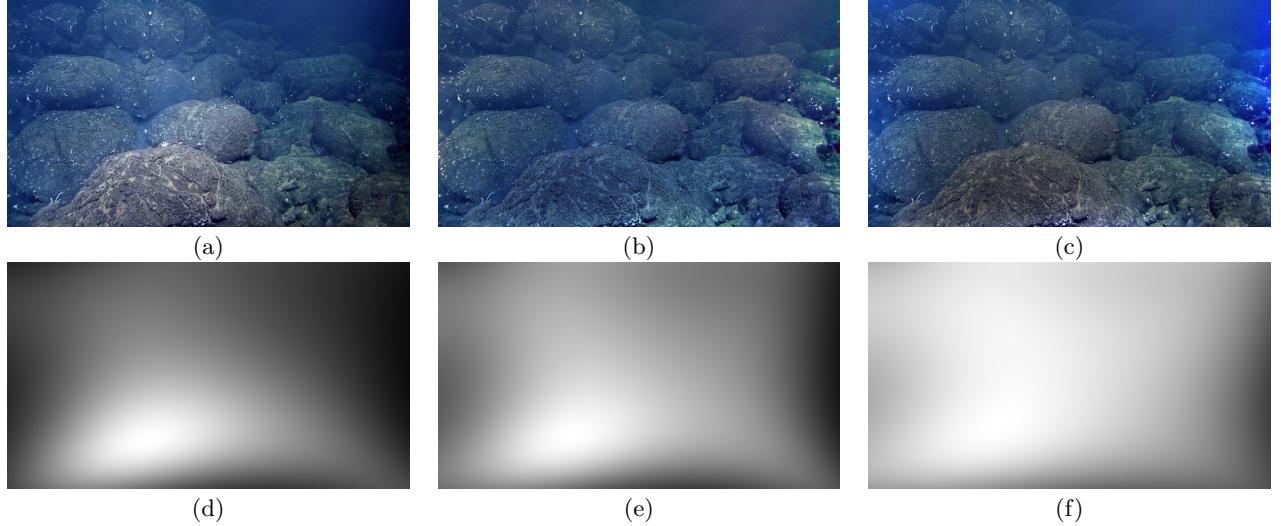


Figure 2: Results obtained after illumination compensation. (a) Shows a typical degraded image. It contains a blue-color haze effect along with difference in brightness between the center region and the corner regions. (b) Shows the uniformly illuminated image obtained after separately estimating the illumination maps for the 3 channels: red, green and blue (shown in Fig. (d), (e) and (f), respectively). It can be inferred from these maps that the blue light reaches most regions of the scene while red light reaches the least. (c) Shows an incorrectly restored image, which contains unwanted color changes across the image because the same illumination map is used to obtain all the three color channels.

In order to get a satisfactory solution in quick time, the iteratively re-weighted least squares (IRLS) [18] is employed. The objective function in each iteration of the IRLS is

$$O = \sum_{(x)} w_k(i; j)(\psi^Z(x) - \psi^M(x))^2 + \sum_{t=0}^D \sum_{l=0}^t a_{t-l, l} \quad (15)$$

where weight $w_k(x)$ is computed in terms of the optimal M_{k-1} from the last iteration as

$$w_k(x) = e^{-S_1}(1 - e^{-S_2}) \quad (16)$$

where $S_1 = \psi_Z(x) - \psi_{M_{k-1}}(x)$ and $S_2 = \alpha(S_1)^{\alpha-1}$. For a single image, if initialized with $M(x) = 0$ (i.e. $b(x) = 1$) for all pixels (x) , three or four iterations are sufficient to obtain perceptually correct illumination maps [29]

Estimating separate illumination maps for each channel is necessary because the attenuation effects are different for each color channel. Experimentally, we found that estimating the illumination map from a grey-scale image and using the same map for all the three channels leads to dominating appearance of different colors in darker and brighter regions of the restored image, as shown in Fig. 2.

For each color channel, the illumination distribution on the image frame m will remain the same for all the images taken with the same imaging set-up. Although the maps obtained with this method for each image are accurate, the approach is prone to error if there is vast differences in the scene depth within an image. The effect of depth on the illumination is prominent in UW scenes because regions beyond a few meters in the scene always appears dark irrespective of the illumination. Hence, we average the illumination maps obtained from each of the individual images to obtain a robust estimate m_{avg} . The inverse of this map ($1/m_{avg}$) is multiplied with each of the input image z to obtain a homogeneously illuminated image i as shown in Fig 2 (b).

5. DEHAZING

We adopt the red dark channel prior based de-hazing mechanism proposed in [10] since it has fewer free parameters than previously existing methods, and can cope efficiently with bright objects present in the scene. The method is based on the following rearrangement of the original model for the three channels i.e.

$$1 - I_R(x) = t(x)(1 - J_R(x)) + (1 - t(x))(1 - A_R(x)) \quad (17)$$

$$I_G(x) = t(x)J_G(x) + (1 - t(x))A_G(x) \quad (18)$$

$$I_B(x) = tJ_B + (1 - t(x))A_B(x) \quad (19)$$

The authors of [10] have proved that

$$\begin{aligned} J^{RED}(x) &= \min(\min_{y \in \omega(x)}(1 - J_R(y)), \min_{y \in \omega(x)}(J_G(y)), \\ &\quad \min_{y \in \omega(x)}(J_B(y))) = 0 \end{aligned} \quad (20)$$

holds true for a non-degraded UW image, where $\omega(x)$ is a small neighbourhood around pixel x . The minimum operation $\min_{y \in \omega(x)}$ finds the minimum over the neighbourhood $\omega(x)$.

This red channel J^{RED} is then used to estimate the airlight and transmission maps as follows: The waterlight is the intensity of pixel in the degraded image that corresponds to the brightest pixel in its red channel.

$$A = (I_R(x_0); I_G(x_0); I_B(x_0)) \quad (21)$$

such that $I^{RED}(x_0) \geq I^{RED}(x) \vee x$

$$\begin{aligned} t(x) &= 1 - \min \left(\frac{\min_{y \in \omega(x)}(1 - I_R(y))}{1 - A_R}, \frac{\min_{y \in \omega(x)}(I_G(y))}{A_G}, \right. \\ &\quad \left. \frac{\min_{y \in \omega(x)}(I_B(y))}{A_B} \right) \end{aligned} \quad (22)$$

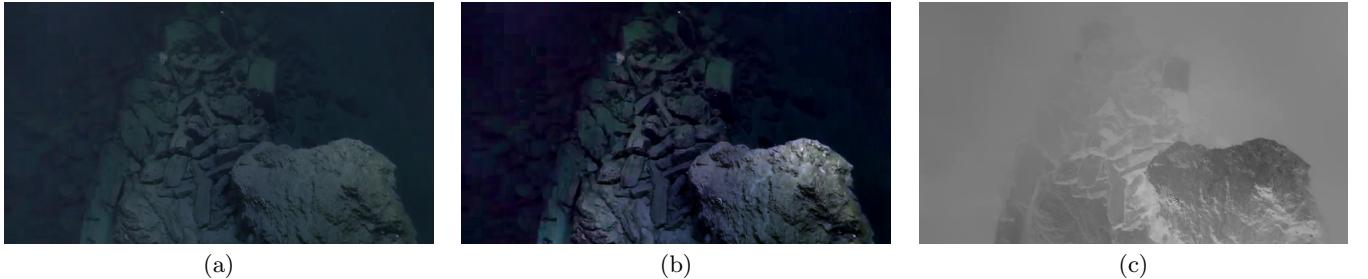


Figure 3: Results after removing haze from a uniformly-lit image. (a) Image containing blue-color haze. (b) and (c) show the restored image and the depth-map obtained after applying the algorithm of [10], respectively. The darker pixels in (c) depict regions closer to the camera.

Following other DCP based approaches, the obtained transmission map is refined using the Guided filter [14], which captures the finer gradients in the degraded image and preserves them in the transmission map while filtering.

Earlier approaches based on DCP fail when bright coloured objects are present in a scene, or there is artificial lighting in the scene. When the dark channel is used to estimate t , nearby bright objects are classified as being far away. This then causes issues in the final dehazed image as these low values in the transmission map can lead to over-saturation in the output image. But the method considered here overcomes this issue by detecting artificial light using saturation.

$$J^{RED-SAT}(x) = \min\left(\min_{y \in \omega(x)}(1 - J_R(y)), \min_{y \in \omega(x)}(J_G(y)), \min_{y \in \omega(x)}(J_B(y)), \min_y \in Sat(y)\right) \quad (23)$$

Relative depth map of an image can be obtained by

$$d(x) = -\log(t(x)) \quad (24)$$

Finally, the restored image is obtained as

$$J_c(x) = \frac{(I_c(x) - A_c)}{\max(t(x), t_0)} + (1 - A_c)A_c \quad (25)$$

where $c \in \{R, G, B\}$ refers to the color-channel. At the end of this step, the restored images are free from color-cast, haze and are all uniformly illuminated. Fig. 3 illustrates the dehazed images and the estimated transmission maps thus obtained for an example UW image. The images obtained after dehazing and their transmission maps are the inputs for our mosaicing module.

6. DEPTH-AWARE STITCHING

A common assumption while performing image stitching is that a global homography can relate any pair of input images. This is clearly violated in the case of UW images, since the effect of depth is prominent due to the relatively small distance between the camera and the scene. This motivated us to use a locally varying homography model in the spirit of [30]. The work in [30] estimates local homographies for each patch in the image. The homography for a patch is estimated by giving higher weights to the feature points which are located closer. This generates a set of homographies that vary smoothly across the image.

In the case of depth-aware scenario, the variation of depth may not be smooth all over the image, and hence the above model is less likely to produce good quality alignment. We

propose to use the depth estimated from haze in UW images, to guide the homography estimates for each patch. We basically enforce the condition that the homography corresponding to patches from the same depth should be the same, and the homographies from nearby depths should be closer to each other. Ours is a feature-based mosaicing technique to construct an image mosaic that is free from parallax effects. The UW image sequence is captured using a monocular video camera. The transmission map obtained from Eq. (24) is used to obtain a rough estimate of the scene depth which is then used for depth-aware warping of images.

Let $\mathbf{x} = [x' \ y']^T$ and $\mathbf{x}' = [x' \ y']^T$ be the location of matching points across overlapping images I and I' . We relate these two positions using the local homography H_* as

$$\hat{\mathbf{x}}' \sim H_* \hat{\mathbf{x}} \quad (26)$$

where H_* is a 3×3 matrix, $\hat{\mathbf{x}} = [\mathbf{x}^T \ 1]^T$ represents \mathbf{x} in homogeneous coordinates, and \sim indicates equality up to a scale. Assume that we have N number of point correspondences across these two images. Let h be the lexicographically arranged form of homography H . Assuming a single global homography H between the two images, we can estimate h using direct linear transformation (DLT) [32] as

$$\hat{h} = \arg \min_h \sum_{i=1}^N \|a_i h\|^2 \quad s.t. \quad \|h\| = 1 \quad (27)$$

where a_i is a 2×9 matrix formed from the coordinates \mathbf{x}_i and \mathbf{x}'_i of i^{th} point correspondence [30]. In our formulation, we differently weight the feature points depending on the depth to which they belong. We use the following minimization to estimate a depth-aware local homography estimate.

$$\hat{h}_* = \arg \min_h \sum_{i=1}^N \|w_*^i a_i h\|^2 \quad s.t. \quad \|h\| = 1 \quad (28)$$

where the scalar weights w_* give higher weights to the feature matches that correspond to the same depth. We determine w_* as

$$w_*^i = \exp\left(-\frac{\|d_* - d_i\|^2}{\sigma^2}\right) \quad (29)$$

Here σ is a scalar which controls the degree of influence due to feature points from neighboring depth layers while estimating the homography at \mathbf{x}_* , while d_* and d_i represent the depth values at \mathbf{x}_* and \mathbf{x}_i . Intuitively, this corresponds to

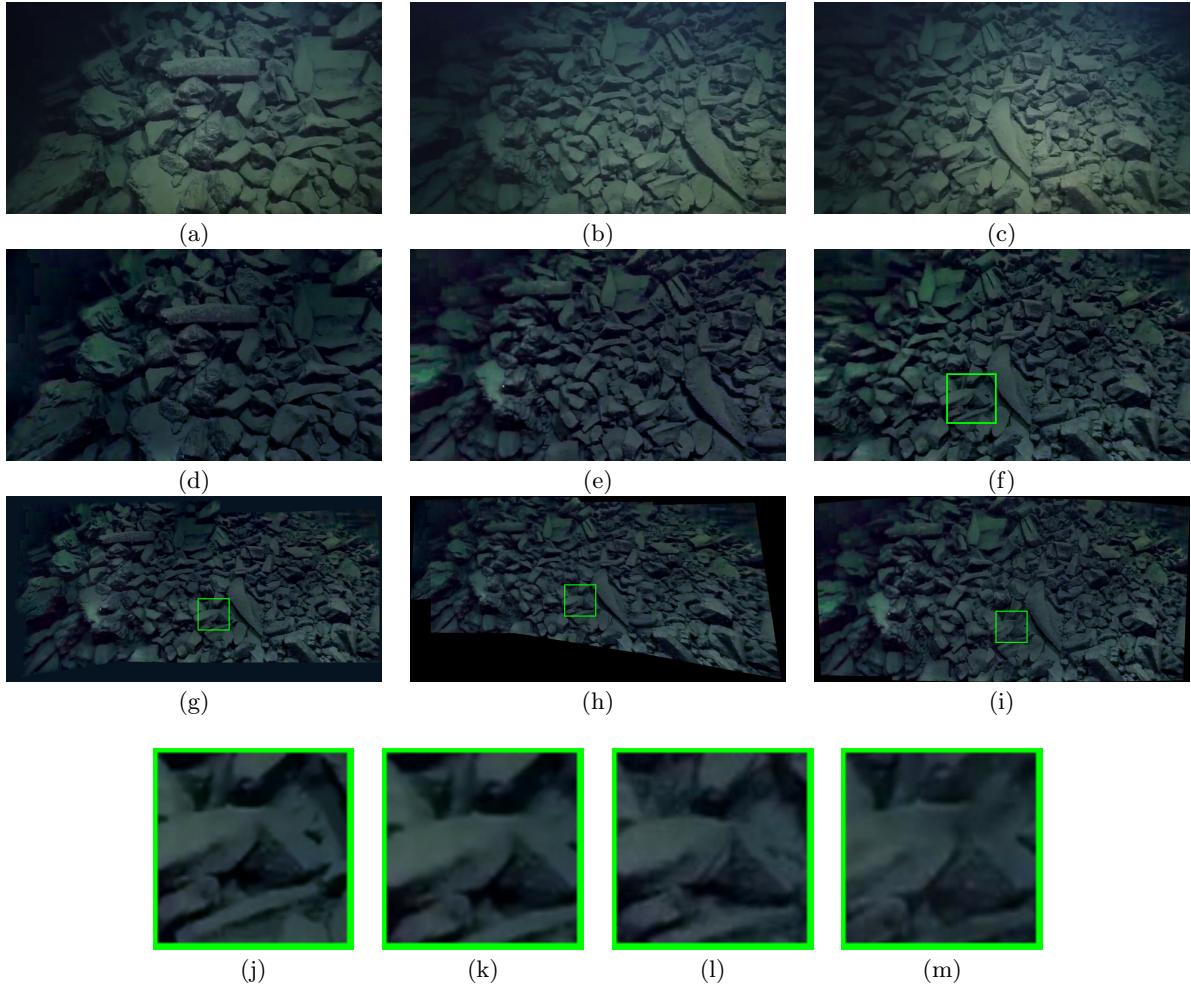


Figure 4: Results obtained after warping the images from experiment 1. The first row (a-c) shows the three input images captured from 3 geometrically-distant viewpoints. These images are restored using our method to obtain the images in second row (d-f). Resultant mosaics using (g) the proposed method, (h) using [30], and (i) AutoStitch [4]. (j-m) Magnified parts from (f, g, h, and i).

finding a local homography h_* that better respects the feature points from neighboring depth layers. The formulation in Eq. 28 can be written in matrix form as

$$\hat{h}_* = \arg \min_h \sum_{i=1}^N \|W_* Ah\|^2 \quad s.t. \quad \|h\| = 1 \quad (30)$$

where W_* is a diagonal weight matrix of size $\mathbb{R}^{2N \times 2N}$ with the weights w_*^i as entries along the diagonal [30]. This is a typical weighted singular vector decomposition (SVD) problem for which the solution is the least significant right singular vector of $W_* A$. In this manner, we estimate homographies corresponding to local patches all over the image and use them to align the images on a common canvas.

Finally, we use α -blending of the warped images to obtain a seamless mosaic. There will be parts in the overlapping region which are different because of the parallax effect (mainly due to significant translational motion of the camera between viewpoints). To address this issue, we make use of the fact that for image mosaicing, input images need not be aligned perfectly over the entire overlapping region [31]. Hence, while warping onto a common canvas, we use infor-

mation only from a single image in the overlapping regions. This helps significantly to reduce blur and artifacts caused by misalignment in parallax affected regions.

7. EXPERIMENTS

We demonstrate our approach on real images that are freely available online [1]. Since the corresponding ground-truth mosaics or restored images are not available, we depict only qualitative comparisons with existing methods. For matching across images, speeded up robust features (SURF) features [2] were extracted from the restored images. For mosaicing, given a set of input images, we restore them and feed them into our depth-based mosaicing algorithm.

We compare our results with the stitching method of [30]. We also compare our results with a commercially available stitching software Autostitch [4].

We considered two sets of images. The first set contained small depth variations within an image (Fig. 4) while the second set contained large differences in depth values (Fig. 5). For both of these examples, we used 3 input images. The input images for the first case are shown in Fig. 4 (a-c).

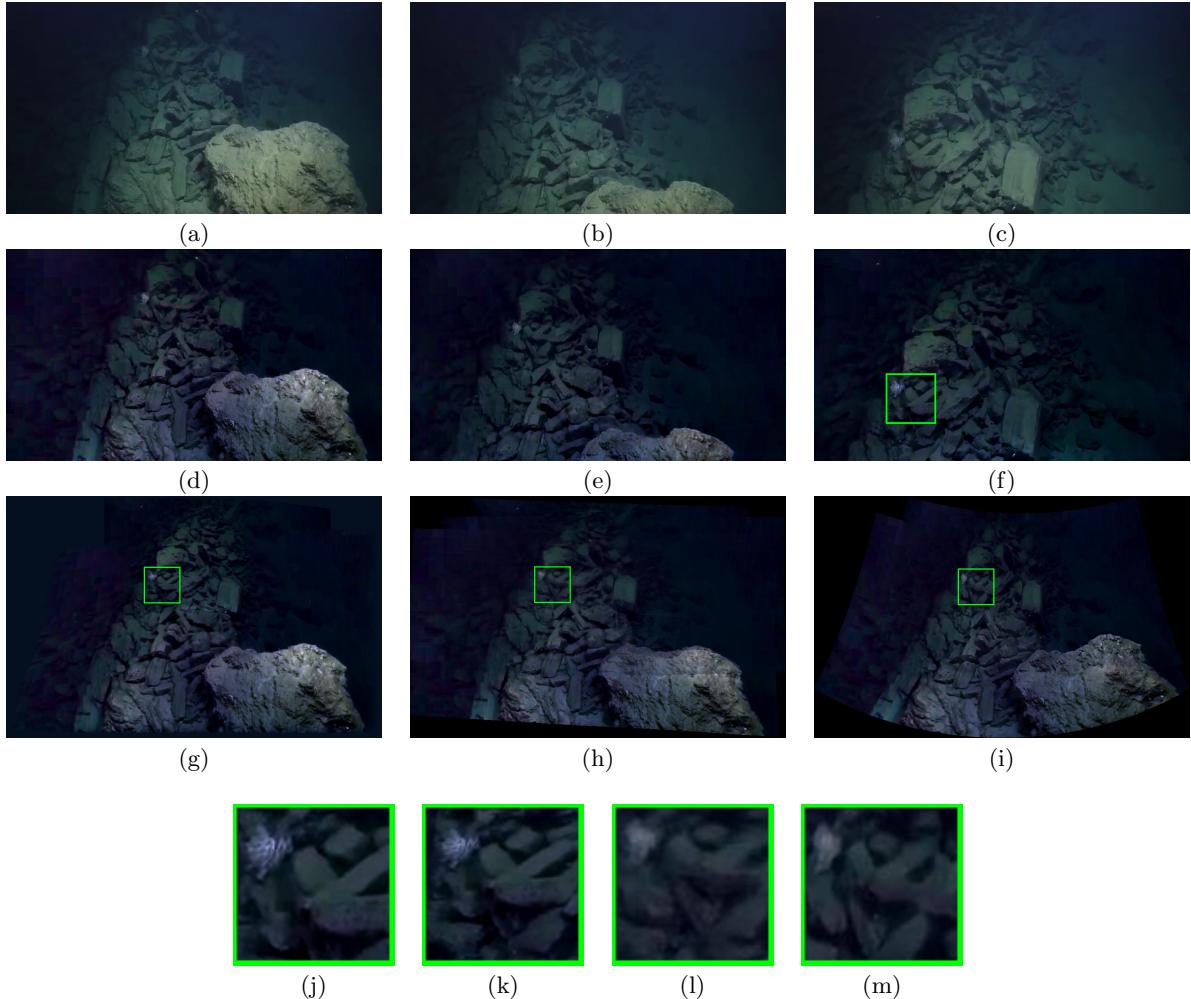


Figure 5: Results obtained after warping the images in experiment 2. (a-c) Input images, (d-f) Restored forms of (a-c) using our method. Mosaics obtained from (g) proposed method, (h) [30], and (i) AutoStitch [4]. (j - m) Zoomed-in portions from (f - i).

We obtain the corresponding restored images shown in Figs. 4 (d-f) by performing non-uniform illumination correction and dehazing on the inputs. The warped images obtained using our approach are used to obtain the result in Fig. 4 (g). As can be observed from the results (g-i) as well as the zoomed-in portions (Figs. 4 (j-0)), competing methods incur significant loss of local information due to blurring as well as double contouring due to sudden changes in depth. In contrast, our method preserves details quite effectively.

For the second experiment, Figs. 5 (a-c) are the input images and Figs. 5 (d-f) shows the corresponding restored images. As can be observed, the image mosaic generated using the proposed method (Fig. 5 (g)) surpasses the outputs from competing methods (Figs. 5 (h,i)). The magnified portions from the input image (Fig. 5 (j)) and mosaics (Figs. 5 (l-m)) reveals the presence of blurring in the overlapping regions for [30] and AutoStitch. However the proposed method (Fig. 5 (k)) is able to generate the mosaic without any deblurring artifacts. Intermediate images, additional results and quantitative comparisons using two no-reference quality metrics (Cumulative Probability of Blur Detection (CPBD) [19] and Entropy) are provided in the supplementary material.

8. CONCLUSIONS

In this paper, we proposed a unified framework for mosaicing of deep UW images captured using ROVs. We suggested a sequence of image correction steps and proposed a depth based stitching method for hazy images. The performance of our approach was demonstrated on challenging real images. Possible extensions to our work include detection of moving objects (organisms and fishes) and/or large particles scatter removal to broaden the scope and utility of the proposed framework.

Acknowledgment: Support provided by National Institute of Ocean Technology, Chennai, India through project sanction No. RB1516ELE011NIOTANRA is gratefully acknowledged.

9. REFERENCES

- [1] Nautilus live.
<http://www.nautiluslive.org/photos-videos>. [Online; accessed April-2016].
- [2] H. Bay, T.uytelaars, and L. Van Gool. Surf: Speeded up robust features. In *European conference*

- on computer vision*, pages 404–417. Springer, 2006.
- [3] C. Beall, F. Dellaert, I. Mahon, and S. B. Williams. Bundle adjustment in large-scale 3d reconstructions based on underwater robotic surveys. In *OCEANS 2011 IEEE-Spain*, pages 1–6. IEEE, 2011.
 - [4] M. Brown and D. G. Lowe. Automatic panoramic image stitching using invariant features. *International journal of computer vision*, 74(1):59–73, 2007.
 - [5] N. Carlevaris-Bianco, A. Mohan, and R. M. Eustice. Initial results in underwater single image dehazing. In *OCEANS 2010 MTS/IEEE SEATTLE*, pages 1–8. IEEE, 2010.
 - [6] J. Y. Chiang and Y.-C. Chen. Underwater image enhancement by wavelength compensation and dehazing. *IEEE Transactions on Image Processing*, 21(4):1756–1769, 2012.
 - [7] P. Drews, E. Nascimento, F. Moraes, S. Botelho, and M. Campos. Transmission estimation in underwater single images. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 825–830, 2013.
 - [8] S. Q. Duntley, A. R. Boileau, and R. W. Preisendorfer. Image transmission by the troposphere i. *JOSA*, 47(6):499–506, 1957.
 - [9] J. Ferrer, A. Elibol, O. Delaunoy, N. Gracias, and R. Garcia. Large-area photo-mosaics using global alignment and navigation data. In *MTS/IEEE OCEANS Conference, Vancouver, Canada*, pages 1–9, 2007.
 - [10] A. Galdran, D. Pardo, A. Picón, and A. Alvarez-Gila. Automatic red-channel underwater image restoration. *Journal of Visual Communication and Image Representation*, 26:132–145, 2015.
 - [11] R. Garcia, T. Nicosevici, and X. Cufí. On the way to solve lighting problems in underwater imaging. In *OCEANS'02 MTS/IEEE*, volume 2, pages 1018–1024. IEEE, 2002.
 - [12] D. Ghosh and N. Kaabouch. A survey on image mosaicing techniques. *Journal of Visual Communication and Image Representation*, 34:1–11, 2016.
 - [13] M. Gupta, S. G. Narasimhan, and Y. Y. Schechner. On controlling light transport in poor visibility environments. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.
 - [14] K. He, J. Sun, and X. Tang. Guided image filtering. In *European conference on computer vision*, pages 1–14. Springer, 2010.
 - [15] J. S. Jaffe. Computer modeling and the design of optimal underwater imaging systems. *IEEE Journal of Oceanic Engineering*, 15(2):101–111, 1990.
 - [16] J. W. Kaeli, H. Singh, C. Murphy, and C. Kunz. Improving color correction for underwater image surveys. *Proceedings IEEE/MTS OceansâŽ 11, Kona, Hawaii, 19–22 September 2011*, pages 805–810, 2011.
 - [17] C.-C. Lin, S. U. Pankanti, K. N. Ramamurthy, and A. Y. Aravkin. Adaptive as-natural-as-possible image stitching. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1155–1163. IEEE, 2015.
 - [18] P. Meer. Robust techniques for computer vision. *Emerging topics in computer vision*, pages 107–190, 2004.
 - [19] N. D. Narvekar and L. J. Karam. A no-reference image blur metric based on the cumulative probability of blur detection (cpbd). *IEEE Transactions on Image Processing*, 20(9):2678–2683, 2011.
 - [20] S. K. Nayar and S. G. Narasimhan. Vision in bad weather. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 2, pages 820–827. IEEE, 1999.
 - [21] A. Ortiz and J. Antich. Bayesian visual tracking for inspection of undersea power and telecommunication cables. *Journal of Maritime Research*, 6(2):83–94, 2014.
 - [22] R. Prados Gutiérrez, R. García Campos, N. R. E. Gráciás, J. Escartín, and L. Neumann. A novel blending technique for underwater gigamosaicing. © *IEEE Journal of Oceanic Engineering*, 2012, vol. 33, núm. 4, p. 626-644, 2012.
 - [23] K. Richmond and S. M. Rock. An operational real-time large-scale visual mosaicking and navigation system. In *OCEANS 2006*, pages 1–6. IEEE, 2006.
 - [24] Y. Y. Schechner and N. Karpel. Clear underwater vision. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages 1–536. IEEE, 2004.
 - [25] R. Schettini and S. Corchs. Underwater image processing: state of the art of restoration and image enhancement methods. *EURASIP Journal on Advances in Signal Processing*, 2010(1):1, 2010.
 - [26] H. Singh, J. Howland, and O. Pizarro. Advances in large-area photomosaicing underwater. *IEEE Journal of Oceanic Engineering*, 29(3):872–886, 2004.
 - [27] H. Singh, C. Roman, O. Pizarro, R. Eustice, and A. Can. Towards high-resolution imaging from underwater vehicles. *The International journal of robotics research*, 26(1):55–74, 2007.
 - [28] R. Szeliski. Image alignment and stitching: A tutorial. *Foundations and Trends® in Computer Graphics and Vision*, 2(1):1–104, 2006.
 - [29] S. A. J. G. Y. Zheng, M. Grossman. Automatic correction of intensity nonuniformity from sparseness of gradient distribution in medical images. In *the 12th International Conference on Medical Image Computing and Computer Assisted Intervention*, September 2009.
 - [30] J. Zaragoza, T.-J. Chin, M. S. Brown, and D. Suter. As-projective-as-possible image stitching with moving dlt. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2339–2346, 2013.
 - [31] F. Zhang and F. Liu. Parallax-tolerant image stitching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3262–3269, 2014.
 - [32] Z. Zhang. Parameter estimation techniques: A tutorial with application to conic fitting. *Image and vision Computing*, 15(1):59–76, 1997.