



IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

Improving Stereo Correspondence in Scattering Media by Incorporating Backscatter Cue

**A. Sarafraz, S. Negahdaripour
and Y. Y. Schechner**

CCIT Report #772
August 2010

 Electronics
Computers
Communications

DEPARTMENT OF ELECTRICAL ENGINEERING
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 32000, ISRAEL



Improving Stereo Correspondence in Scattering Media by Incorporating Backscatter Cue

Amin Sarafraz, *Member, IEEE*, Shahriar Negahdaripour, *Member, IEEE*, and Yoav Y. Schechner, *Member, IEEE*

Abstract—Image contrast can deteriorate significantly in scattering media, such as underwater, due to backscatter. This affects the performance of many computer vision techniques developed for open-air conditions, including the stereo matching algorithms, when applied to images acquired in these environments. It has been demonstrated that the backscatter field embodies depth information, thus can potentially provide an effective visual cue for 3-D reconstruction. In this paper, we address the estimation of the backscatter component in stereo images, in order to employ it as an additional cue for disparity estimation. More precisely, we decouple the stereo images into signal and backscatter components, and thus are able to make use of depth cues offered by both components in order to devise a more robust technique for disparity computation. Our method is invariant to illumination setup, and requires neither lighting calibration nor the knowledge of medium optical properties. Results of experiments with synthetic and real data are provided to demonstrate the performance of our new method.

Index Terms—Stereo, Scattering.

I. INTRODUCTION

FEATURE matching, the primary challenge in stereovision, has been studied over several decades; e.g. [1]. Despite extensive work dealing with open-air images, only a small number of studies has addressed the problem when the images are taken in scattering media, e.g., in fog, haze, and underwater [2], [3], [4], [5]. Most computer vision methods face significant difficulties if employed directly for underwater images [6], and those recorded in other scattering environments.

In such domains, the backscatter field can significantly corrupt the scene radiance, which comprise the signal component of the image. Additionally, the signal component undergoes attenuation due to the medium absorption and scattering as the light rays travel from the object to the camera. Since backscatter increases with the distance between the camera and the scene, this leads to a contrast decay that varies across the image [7]. In low-light environments, artificial sources are necessary, adding other complexities.

In this case, corresponding patches in stereo images might have different brightness levels with dissimilar backscatter

Amin Sarafraz is with the Department of Civil, Architectural and Environmental Engineering, University of Miami, Coral Gables, FL, 33124 USA e-mail: a.sarafraz@umiami.edu.

Shahriar Negahdaripour is with the Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL, 33124 USA e-mail: shahriar@miami.edu .

Yoav Y. Schechner is with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel e-mail: yoav@ee.technion.ac.il.

Manuscript received this date; revised that date.

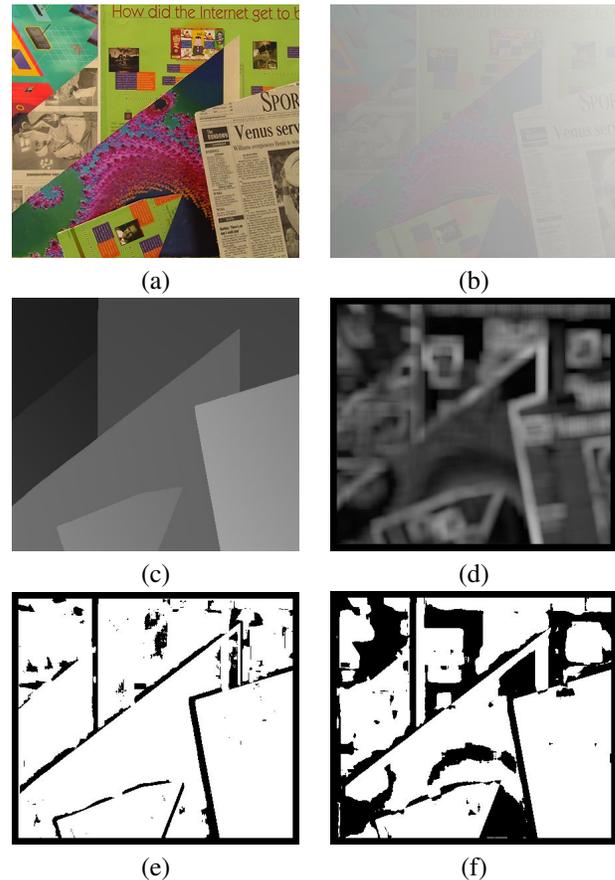


Fig. 1. Turbidity affects the performance of stereo matching methods. Left image of a stereo pair in (a) open air, (b) turbid water. (c) Ground truth disparity map: pixels with brighter gray value are closer to the camera. (d) Contrast map: pixels with higher contrast values are shown in brighter pixels. Validity map of the estimated depth map using normalized SSD in (e) open air, (f) turbid water. Pixels with correct estimated disparity are shown in white.

components. These effects make stereo matching even more challenging in scattering media. To demonstrate the negative impact of these effects on stereo matching results, we picked the standard normalized SSD method and run it on original Poster image from Middlebury dataset [1] and its corresponding synthetic stereo pair in turbid water. The validity map of estimated disparities are illustrated in Fig. 1. Comparing the validity maps of open air and turbid water images, one can readily see that the depth map estimation over the low-contrast areas is affected due to the signal attenuation and backscatter presence. Standard deviation of pixel values in a local window is used as the contrast measure and the contrast map is shown in Fig. 1 for reference. In scattering media, backscatter

component lowers the contrast of the image and wherever the backscatter dominates the signal, the stereo matching methods will have more difficulties in finding correspondences.

Some open-air methods that relax the brightness constancy assumption may be applied in clear water, where the scattering is negligible and the lighting is relatively uniform; e.g., [8], [9]. In other work targeted directly for underwater mapping, the images have been acquired under good visibility, and thus negligible backscatter [2], [3], [4], [5]. In [7], the authors propose the recovery of de-scattered stereo views to enable the application of a traditional stereo algorithm. Their method requires the estimation of the optical properties of the medium in order to compute the backscatter field from the depth map. Moreover, the depth cue in the backscatter field is totally ignored. In [10], the authors make use of both the polarization and stereo cues for image enhancement. In producing the de-scattering image, the scene depth map computed from the signal component is applied. This method also makes use of the known optical properties of the medium and does not exploit the depth cue of the backscatter component. The use of depth cue in the backscatter component has been demonstrated in [11], where two images acquired with different polarization filter settings from a single viewpoint are exploited for image enhancement by de-scattering.

Another possible solution includes recovering the object radiance using the methods of single image backscatter removal and hence increase the contrast of the image. Different methods are developed to address single image backscatter removal [12], [13], [14]. However, all these methods are designed for atmospheric airlight and assume uniform lighting and thus limits their applicability. Furthermore, the method in [12] is based on dark channel prior and might not work for images with invalid dark-channel prior.

In this paper, we make use of both the scattering field and binocular stereo in solving the correspondence problem, the first step towards the 3-D scene reconstruction. We show how to estimate the backscatter field without the knowledge of the medium optical properties. The method is invariant to illumination source, applying to both natural illumination and artificial lighting. Therefore, our method requires neither parametric lighting calibration nor knowledge of medium optical properties. We make use of representing the backscatter field in a local patch as a scaling of the backscatter at infinity, which can be readily estimated. By determining the scaling within the support region of each pixel, we exploit the depth cue in the backscatter field. Our results with both synthetic and real data verify the improved robustness and accuracy with respect to the cases where one or the other, but not both visual cues are employed.

II. STEREO MODELING IN SCATTERING MEDIA

Consider two calibrated cameras viewing the scene (Fig. 2). For simplicity, assume that their coordinate systems and optical axes are parallel, and that their baseline vector is $\mathbf{D} = (D, 0, 0)$ in the global coordinate system. Hence, the epipolar lines are parallel to the x axis. We denote the corresponding image coordinates of the same scene point in the left and right images by \mathbf{x}_{obj}^i , $i \in L, R$ where $\mathbf{x} = (x, y)$.

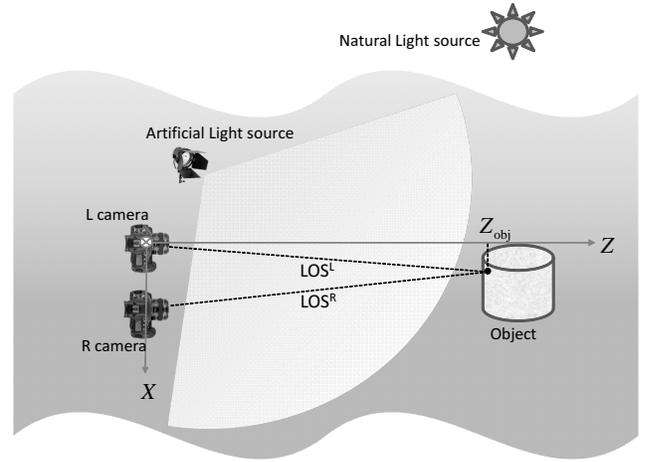


Fig. 2. Stereo setup in a scattering media.

In scattering media, the left I^L and right I^R images may be modeled by the attenuated signal S and the backscatter B components [10]:

$$I^L(\mathbf{x}_{obj}^L) = S^L(\mathbf{x}_{obj}^L) + B^L(\mathbf{x}_{obj}^L) \quad (1)$$

$$I^R(\mathbf{x}_{obj}^R) = S^R(\mathbf{x}_{obj}^R) + B^R(\mathbf{x}_{obj}^R) \quad (2)$$

Since the camera coordinates systems and their optical axes are parallel, we can write:

$$\mathbf{x}_{obj}^R = \mathbf{x}_{obj}^L + (d, 0) \quad (3)$$

where d is the disparity for two corresponding points in the left and right views. More details and formulation of the signal and the backscatter components can be found in [10].

III. ESTIMATION OF BACKSCATTER AND SIGNAL COMPONENTS

Our goal is to decompose each image in a stereo pair recorded in a scattering media into its backscatter and signal components. Consider local patches around the corresponding pixels in the left and the right views which we call them $\Omega_{\mathbf{x}_{obj}^L}$ and $\Omega_{\mathbf{x}_{obj}^R}$, respectively. We define the following fields in terms of the signal and backscatter components of the two views:

$$\mathbf{S}(\Omega_{\mathbf{x}_{obj}^L}, \Omega_{\mathbf{x}_{obj}^R}) = S^L(\Omega_{\mathbf{x}_{obj}^L}) + S^R(\Omega_{\mathbf{x}_{obj}^R}) \quad (4)$$

$$\mathbf{B}(\Omega_{\mathbf{x}_{obj}^L}, \Omega_{\mathbf{x}_{obj}^R}) = B^L(\Omega_{\mathbf{x}_{obj}^L}) + B^R(\Omega_{\mathbf{x}_{obj}^R}) \quad (5)$$

$$\rho(\Omega_{\mathbf{x}_{obj}^L}, \Omega_{\mathbf{x}_{obj}^R}) = \frac{B^L(\Omega_{\mathbf{x}_{obj}^L}) - B^R(\Omega_{\mathbf{x}_{obj}^R})}{B^L(\Omega_{\mathbf{x}_{obj}^L}) + B^R(\Omega_{\mathbf{x}_{obj}^R})} \quad (6)$$

It is noted that where the propagation path from target to the two cameras is roughly the same, the brightness constancy assumption may be applied to the attenuated signals:

$$S^L(\Omega_{\mathbf{x}_{obj}^L}) = S^R(\Omega_{\mathbf{x}_{obj}^R}) \quad (7)$$

For simplicity hereafter we omit the arguments $\Omega_{\mathbf{x}_{obj}^L}, \Omega_{\mathbf{x}_{obj}^R}$ from the equations, in the remainder.

Using Eqs.(1) and (2) and the definitions for \mathbf{B} , \mathbf{S} and ρ , the following equations can be derived for local patches around the corresponding points $\{\mathbf{x}_{\text{obj}}^L, \mathbf{x}_{\text{obj}}^R\}$ in the left and right images:

$$I^L + I^R = \mathbf{B} + \mathbf{S} \quad (8)$$

$$I^L - I^R = \rho \mathbf{B} \quad (9)$$

Suppose that we can estimate ρ and the backscatter at infinity in the left and right images, B_∞^L and B_∞^R , respectively. (We show how these are estimated in the next section.) These are denoted $\hat{\rho}$, \hat{B}_∞^L and \hat{B}_∞^R . Then, \mathbf{B} can be readily estimated from (9):

$$\hat{\mathbf{B}} = (I^L - I^R) / \hat{\rho} \quad (10)$$

Referring to Eq. (6), it is noted that ρ could be close to (or equal to) zero, where there is not enough difference between the left and right backscatter fields. Therefore, \mathbf{B} in Eq.(10) will be overestimated. To avoid that, we restrict the $\hat{\rho}$ to a lower bound $\hat{\rho}_0$. The final formulation of \mathbf{B} would be

$$\hat{\mathbf{B}} = \pm \frac{I^L - I^R}{\max(\hat{\rho}, \hat{\rho}_0)} \quad (11)$$

We used the value 0.1 for $\hat{\rho}_0$ in all results shown in this paper. The \pm sign will assure that the value of estimated backscatter is positive. Following [12], we assume backscatter is uniform across a local patch. Similarly, we assume the backscatter at infinity for a local patch is uniform across the entire local patch but it might have different intensity values in left and right views. Once we have an estimate for \mathbf{B} , it is distributed between B^L and B^R according to their corresponding backscatter at infinity:

$$\hat{B}^L = \hat{s}^L \hat{\mathbf{B}}; \quad \hat{s}^L = \frac{\hat{B}_\infty^L}{\hat{B}_\infty^L + \hat{B}_\infty^R} \quad (12)$$

$$\hat{B}^R = \hat{s}^R \hat{\mathbf{B}}; \quad \hat{s}^R = \frac{\hat{B}_\infty^R}{\hat{B}_\infty^L + \hat{B}_\infty^R} \quad (13)$$

By substituting the estimates of B^L and B^R into Eq. (1) and (2), the signal component of the left and right images can be estimated:

$$\hat{S}^L = I^L - \hat{B}^L \quad (14)$$

$$\hat{S}^R = I^R - \hat{B}^R \quad (15)$$

We emphasize that the above equations hold for the matching $\{\mathbf{x}_{\text{obj}}^L, \mathbf{x}_{\text{obj}}^R\}$ in the left and right images. This fact plays a key role in the development of our method. To illustrate this fact more explicitly, we demonstrate an example here. Suppose we are searching along an epipolar line for matches by incrementing disparity from d_{\min} to d_{\max} . At each disparity, we extract local patches around the pixels in the left and the right images, i.e. $\Omega_{\mathbf{x}_{\text{obj}}^L}$ and $\Omega_{\mathbf{x}_{\text{obj}}^R}$. Using Eqs (6), (10), (12), we estimated B^L for $\Omega_{\mathbf{x}_{\text{obj}}^L}$ and showed in second row of Fig.3. Comparing to B_∞^L shown in the first row of Fig.3, it can be seen that at correct disparity, the estimated B^L is smooth and similar to backscatter at infinity. Also, note that the estimated backscatter at correct disparity has lower intensity value than B_∞^L which is the upper limit for \hat{B}^L .

A. Estimation of ρ and Backscatter at Infinity

We can estimate B_∞^L and B_∞^R by taking an image with no object in the field of view of the camera. This means that $S^L = S^R = 0$. By substituting this into Eq.(1) and Eq.(2), one can conclude

$$\hat{B}_\infty^L = I_\infty^L \quad (16)$$

$$\hat{B}_\infty^R = I_\infty^R \quad (17)$$

Using equal-path-length assumption, in a homogenous medium, ρ for a local patch can be estimated as follows:

$$\hat{\rho} = \frac{\hat{B}_\infty^L(\Omega_{\mathbf{x}_{\text{obj}}^L}) - \hat{B}_\infty^R(\Omega_{\mathbf{x}_{\text{obj}}^R})}{\hat{B}_\infty^L(\Omega_{\mathbf{x}_{\text{obj}}^L}) + \hat{B}_\infty^R(\Omega_{\mathbf{x}_{\text{obj}}^R})} = \frac{I_\infty^L(\Omega_{\mathbf{x}_{\text{obj}}^L}) - I_\infty^R(\Omega_{\mathbf{x}_{\text{obj}}^R})}{I_\infty^L(\Omega_{\mathbf{x}_{\text{obj}}^L}) + I_\infty^R(\Omega_{\mathbf{x}_{\text{obj}}^R})} \quad (18)$$

IV. DISPARITY ESTIMATION

Having estimated the backscatter and the signal components for both images, we can formulate the stereo matching problem as a minimization of the following energy function:

$$E_T = E_S + E_B \quad (19)$$

Here E_S and E_B are the energy functions for the estimated signals and estimated backscatters respectively.

A number of different frameworks, e.g., the weighted support windows, hierarchical techniques, occlusion handling methods, etc., can be utilized in defining our energy functions. Among these, the weighted support windows is the simplest implementation, where we can analyze the results from our method with those obtained by processing the raw images, or the signal components only.

For the signal part, we use the commonly adopted normalized Sum of Squared Distances (SSD) as the similarity measure:

$$E_S = \text{SSD}(\hat{S}^L, \hat{S}^R, d) \quad (20)$$

Next, we recall that we assumed the backscatter is uniform across a local patch. Thus, at correct disparity, the estimated backscatter in the left and right images for a local patch are the scaled versions of their corresponding backscatter measures at infinity. Therefore, a suitable energy function E_B is the square of the discrepancy between the estimated left/right back-scatter and the prediction by the scaled left/right back-scatter at infinity:

$$E_B = \sum_{\Omega_{\mathbf{x}_{\text{obj}}^L}} (|B_\infty^L - k^L \hat{B}^L|) + \sum_{\Omega_{\mathbf{x}_{\text{obj}}^R}} (|B_\infty^R - k^R \hat{B}^R|) \quad (21)$$

Here, any suitable norm can be applied within the support windows $\Omega_{\mathbf{x}_{\text{obj}}^L}$ and $\Omega_{\mathbf{x}_{\text{obj}}^R}$, centered at the left point $\mathbf{x}_{\text{obj}}^L$ and the assumed match $\mathbf{x}_{\text{obj}}^R$. The scales $k^{L,R}$ can be easily estimated by dividing estimated backscatter of local patch by the corresponding backscatter at infinity and averaging over the whole local patch; i.e.

$$k^{L,R} = \text{avg}\left(\frac{B_\infty^{L,R}}{\hat{B}^{L,R}}\right) \quad (22)$$

The solution is taken as the disparity gives the minimum total energy:

$$\hat{d}_T = \arg \min(E_T) \quad (23)$$

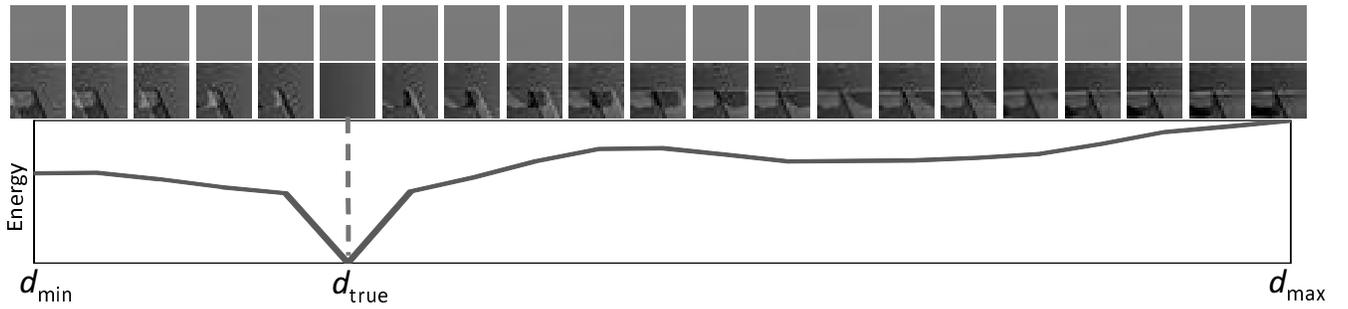


Fig. 3. Illustration of estimated backscatter in a local patch at different disparity. (First row) Corresponding backscatter at infinity for local patch in the left image, (Second row) estimated backscatter for local patch in the left image, (Bottom) energy function which shows the similarity between estimated backscatter and backscatter at infinity around the local patch.

A. Implementation

The performance of our method can be demonstrated through a simple implementation. For each pixel, we compute the energy function for the disparity range of interest, and determine the solution based on the minimum energy point. The computation steps are as follows:

- ```

for $d = d_{\min} : d_{\max}$ do
 - Compute Backscatter components using Eq. (10) followed by Eqs. (12) and (13).
 - Compute signal components using Eqs. (14) and (15).
 - Compute the scales k^L, k^R by dividing the estimated backscatter to the corresponding backscatter at infinity over the support window.
 - Compute the total energy from the energy for the signal and the backscatter components, as given by Eqs. (20) and (21).
end for
 - Select the disparity with minimum energy as the final solution

```

## V. EXPERIMENTS AND RESULTS

We contrast the performance in disparity computation by applying the normalized SSD solution to the raw stereo data and the signal component, as well as with our method that utilizes the depth cues in both the backscatter and signal components. To quantify the accuracy of the computed correspondences, we have determined the percentage of correctly matched points, utilizing the ground truth. We present the results for synthetic and real data sets.

### A. Synthetic Images

To evaluate the impact of backscattering and signal attenuation on disparity estimation, we have generated synthesized images for different ocean environments with varying degree of backscatter. We have applied the camera and lighting settings from our experiments with real data (see next section). The stereo baseline is assumed to be 17 [cm], with the light source placed between the two cameras at (8,-4,0) [cm] w.r.t the left camera (see Fig. 2). We have assumed a (non-uniform) Gaussian distribution for the illumination field.

We have chosen samples from the standard Middlebury stereo data set [1] for the scene radiance, namely the Poster,

Venus, Sawtooth, Teddy, and Cones. These images contain objects with different shapes and texture. The matching window size is set to 21 [pix]  $\times$  21 [pix] for the entire data set. With the ground truth disparity maps, we have utilized the models of the falloff and backscatter functions [10], and the values given for the optical properties of different ocean water [15], to compute the backscatter and signal components of the left and right images. The signal component is computed by the multiplication of the falloff function with the object radiance. The synthetic images are the sum of the signal and backscatter components, as indicated in Eqs. (1) and (2). Finally, we have applied additive random noise with variance of one gray level.

Fig. 4 shows the left and right Poster data, and the synthetic images in Turbid Harbor. The target scene is assumed to be at an average distance of 70 [cm] from the camera.

The disparity range,  $d_{\max} - d_{\min}$ , was set to 30 pixels for all synthetic data. The results of applying the normalized SSD on the raw image intensities and our method are shown in Fig. 4. The ground truth disparity map and validity map of the estimated disparity map using the original Poster images are shown for reference. As can be seen, the SSD-based method applied to the raw images cannot accurately recover the disparity in regions with weak texture. In contrast, our method successfully estimates the disparity in these regions, and its performance is close to the performance of SSD-based method when applied to open-air images. This fact shows that despite significant signal deterioration in scattering media, this is alleviated by the recovery of the signal and backscatter, and exploiting the depth cues in both components.

Fig. 5 shows four more synthetic images and corresponding results. The contrast map and disparity map of each image are shown for reference. Standard deviation of intensities of a local window of size 21[pix]  $\times$  21[pix] around each pixel is used as the contrast measure for that pixel. As expected, the backscatter becomes more dominant with increasing scene distance from the cameras. Moreover, it affects more those regions with low contrast. Not surprising, this leads to poor estimation performance in distant scene points with weak surface texture. Most of these faults has been compensated when we decouple signal and backscatter. For more detailed assessment of these results, we categorize the pixels into *Textured* and *Non-Textured* regions. We assign a pixel to the *Textured* region if the average gradient within its support

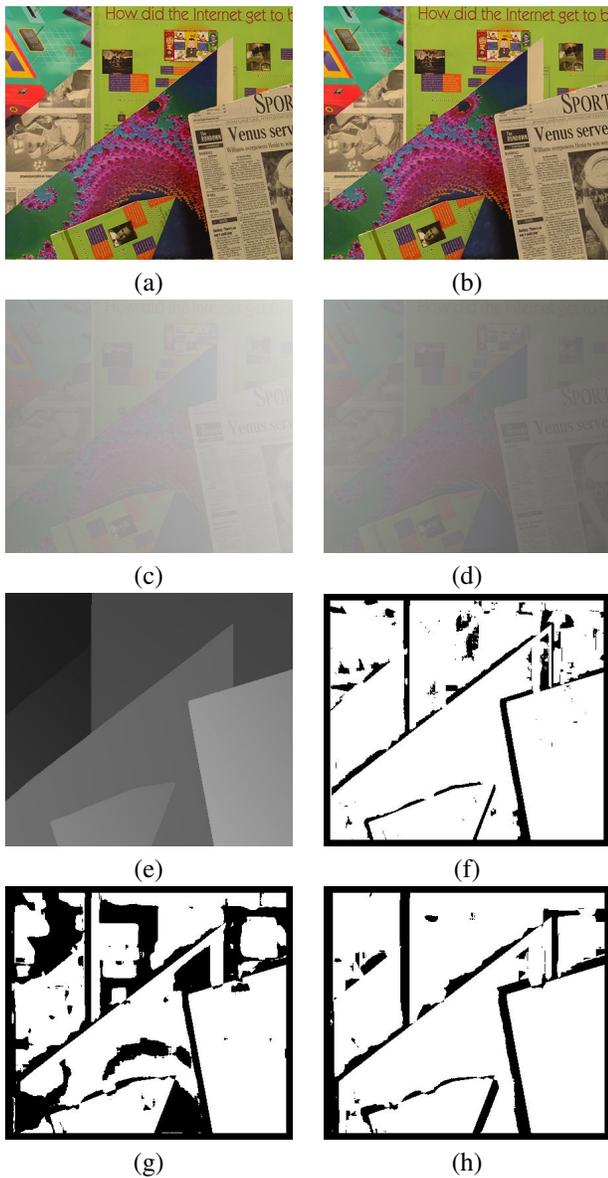
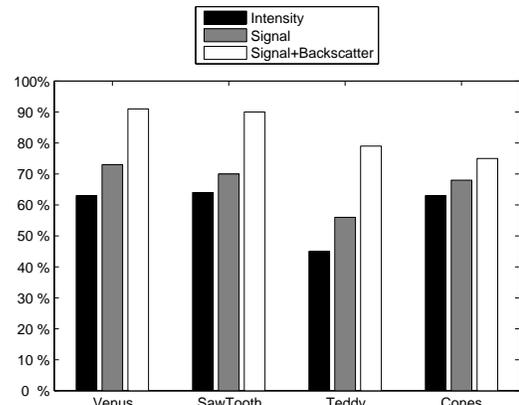


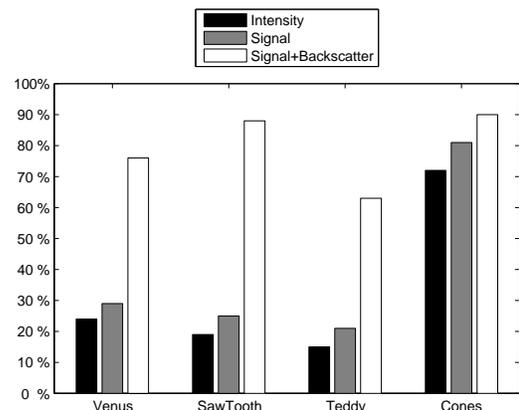
Fig. 4. Left (a) and right (b) open-air images. (c,d) Turbid-water stereo images synthetically generated from a-b. Results of disparity computation in the form of binary validity maps for good (white) and bad (black) pixels, with threshold set at one-pixel disparity error. Validity maps are given for normalized SSD applied to (f) open air images in a-b, (g) synthetic turbid-water images in c-d, and (h) for the proposed method to scattering and signal components.

window is above a given threshold. Otherwise, it belongs to a *Non-Textured* region. Since the labeling depends on the threshold, we evaluate the results for a range of thresholds (see Fig. 6). For *Textured* regions, increasing the threshold means including points with stronger surface markings. Conversely, we capture points with weakest texture within the *Non-Textured* regions, when we lower the threshold. The gradient magnitude computation has been carried out for both the original and synthetic images.

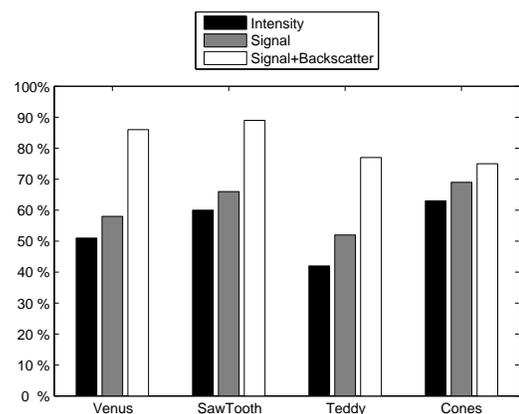
As we note in Figs. 6 (top row and second row), the performance of the competing methods converge in higher textured regions (greater thresholds), and the improvement from our method becomes negligible. It must be noted that the



(a)



(b)



(c)

Fig. 7. Percentage of Pixels with correct disparity for all synthetic dataset in (a) *Textured* region, (b) *Non-Textured* region, and (c) Whole image. The *Textured* and *Non-Textured* regions are provided in Middlebury dataset [1]

number of pixels in these regions decrease as the threshold is raised. Also, the gradient magnitude of the synthetic images is typically smaller in comparison to the original images.

In the *Non-Textured* regions, our method produces more pixels with correct disparity pixels as the threshold decreases. See Fig. 6 (third row and bottom row). This emphasizes the

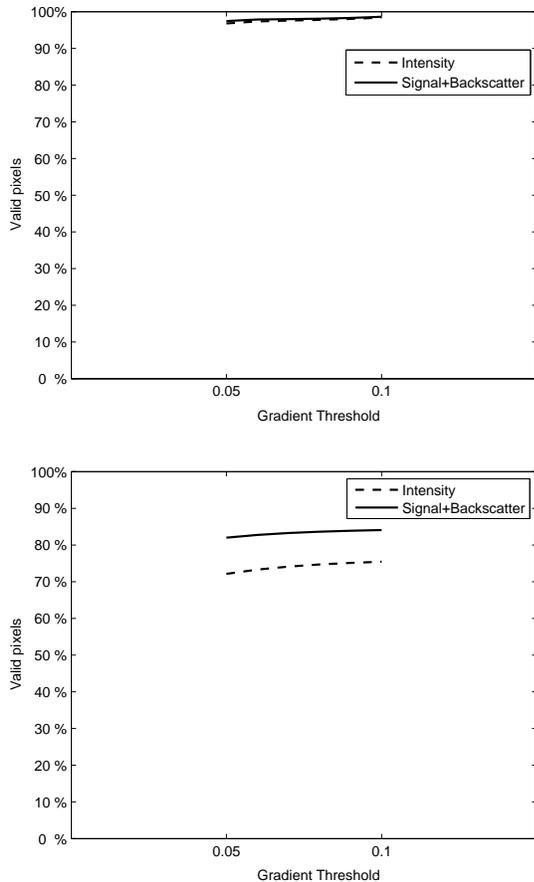


Fig. 8. Percentage of pixels with correct disparity as a function of gradient magnitude threshold for real data; *Textured* (top row) and *Non-Textured* (bottom row) regions. Image intensities are normalized and gradient magnitude of the image ranges from 0 to 0.2

role of the backscatter field in the disparity estimation since it dominates the signal component in the low textured areas. In this analysis, we have discarded the occluded areas.

Fig. 7 shows the improvement within Textured and Non-Textured regions, as well as the entire image for the Middlebury data set [1]. The results illustrate that although we gain by performing the stereo matching on the recovered signal component, the proposed method performs better by exploiting the depth cues by both the signal and backscatter components. More importantly, the improvement is notably more significant over the low textured regions, where the backscatter can dominate the signal component.

1) *Comparison with Backscatter Removal Methods*: As mentioned in the introduction, one solution to achieve better stereo matching results in scattering media is to remove backscatter using single image backscatter removal methods and then apply the stereo matching on enhanced images. Amongst different existing methods for single image backscatter removal, we selected the method introduced in [12] which is one of the most promising methods. The method is developed for uniform lighting. It means that the backscatter at infinity would be uniform for the entire field of view. First, we tried to use a constant value for  $B_\infty$ . For that purpose,



Fig. 10. (top row) stereo images in clear water, (middle row) stereo images of the same scene in turbid water made by adding milk to water, (bottom row) Backscatter at infinity for the left and the right images.

we used both the the average of  $B_\infty$  and maximum of  $B_\infty$  as an approximation for entire field of view. The first row and second row in Fig. 9 shows the corresponding validity maps. Also, we used the backscatter at infinity for the entire filed of view without assuming that the  $B_\infty$  is constant. This way, each pixel has its corresponding  $B_\infty$ . The results are shown in third row in Fig. 9. The corresponding percentage of valid pixels are shown below each validity map. In addition, we show the results of our method in bottom row of Fig. 9. As can be seen, overall the results of our method is better than all other cases especially in images with planar objects, i.e. Venus and Sawtooth. Also, if we look at the contrast maps of these two images we can see that Venus and Sawtooth contain more low contrast areas than the other two images, i.e Teddy and Cones.

## B. Real Images

We next evaluate the performance of our method with real data, comprising the stereo images of known targets in a water tank. We have recorded images in both clear water (for reference) and with the addition of a known volume of highly scattering low-fat milk (see Fig. 10 for comparison). Referring to bottom row of Fig. 10, we also recorded backscatter at infinity, i.e.  $B_\infty^L$  and  $B_\infty^R$ , by imaging a void area where there is no object in front of the camera setup except water column. All real images are recorded at the  $768 \times 1024$  pixel resolution.

The water tank data set comprises three objects, namely a planar mat, a cylindrical object, and a custom-made chart of various patterns and images that are often used to demonstrate/assess the performance of various image processing methods; e.g., B&W cameraman image, alphabet letters at different sizes and distances, bar codes at different thickness

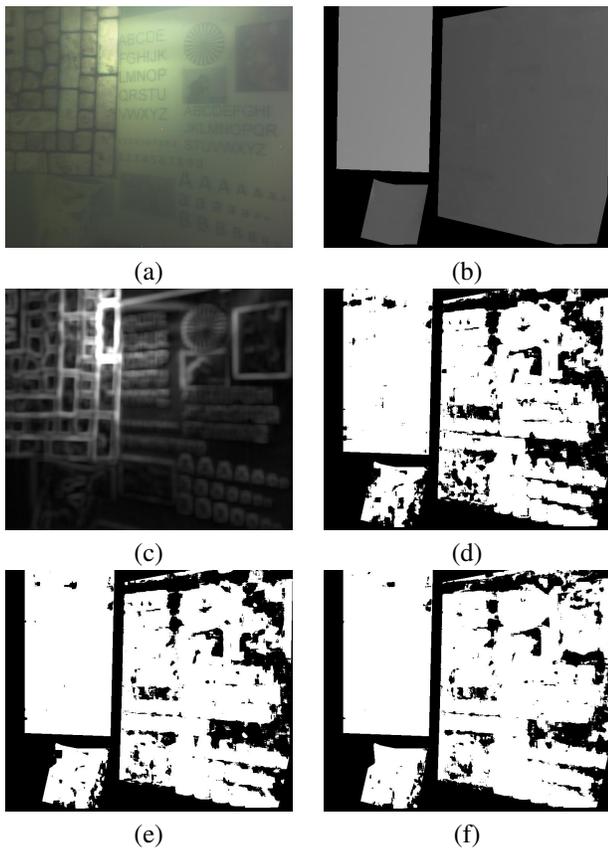


Fig. 11. (a) Real image, (b) ground truth disparity map, (c) Contrast map: pixels with higher contrast value are brighter. The binary validity maps show good (white) and bad (black) pixels (with threshold set at one-pixel error) for the disparity computed from (d) raw data, (e) signal component, and (f) our method.

and separation distances, etc. Placed at a slanted orientation, it covers a distance of 1-1.5 [m] relative to the cameras. The mat has a stronger texture and is placed closest to the stereo setup at about 70 [cm]. The cylinder is positioned at a distance of roughly 125 [cm] from the two cameras. A spotlight at about 2-3 [m] over the scene above the water surface provides a relatively uniform natural lighting, while the main light source positioned midway between the two cameras provides non-uniform illumination. We have computed the disparity map for the stereo images recorded in clear water with the known scene geometry, namely, target shapes and configurations. This serves as the ground truth in this experiment.

In our performance assessment for real data, we have set the threshold at 5 pixels in labeling the estimated disparity as accurate/inaccurate; i.e. a discrepancy disparity of 5 or more pixels relative to the ground truth is labeled as inaccurate estimate (see later results below). This is to account for various uncertainties in the estimation of the ground truth disparity map.

Knowing the minimum and the maximum distance of the objects from the camera setup, we set the disparity range,  $d_{\max} - d_{\min}$ , to 200 pixels for Tank data set. The results for our method and those obtained by utilizing solely the signal component exhibit comparable performance with low error in areas where the signal is dominant, e.g., the close-range tex-

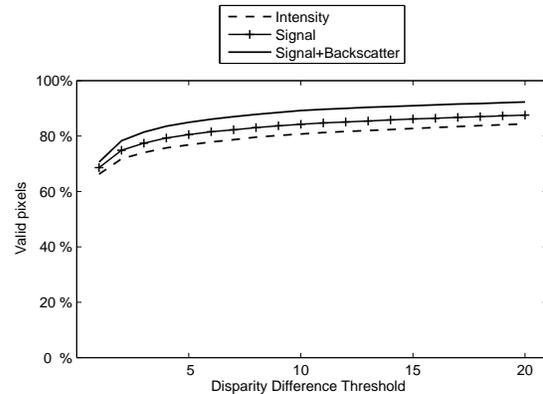


Fig. 13. Percentage of pixels with correct disparity over the entire image for different thresholds for disparity difference

tured mat. Where the backscatter is negligible, both methods have difficulties over texture-less areas, although our method performs marginally better. The main difference lies in the areas where the signal is veiled and the backscatter dominates. Some examples are the regions include the cameraman, some textured regions of the poster, the cylindrical object. In these region, the superiority of our method is significant.

For a more detailed examination of these results, we again concentrate on the performance within *Textured* and *Non-Textured* regions, as we described for the synthetic data sets. Referring to Fig. 8, we note that the results are consistent with our conclusions drawn from the experiments with synthesized data; in *Non-Textured* areas our method has a better performance while we obtain comparable accuracy in highly-textured regions. Here, as the threshold on the gradient magnitude is increased, the number of pixels within the *Textured* region decreases significantly.

As can be seen in Fig. 10, different regions have different amount of backscatter due to the non-uniformity in illumination. This results in a different amount of contrast degradation. For more exploration of the results shown in Fig. 11, we extracted different regions in the image which represent low, medium and high contrast areas. We picked 5 regions which is shown in Fig. 12. We name them as Fruits, Cylinder, Camera Man, Letters and Mat. The Fruits and Cylinder represent low contrast, Camera Man represents medium contrast and finally, Letters and Mat represent high contrast areas. The corresponding contrast map of each area along with the results of applying SSD on raw data and the results of our method are shown in Fig. 12. As can be seen our method over performs other methods in low contrast areas. The percentage of improvement over each area are shown below its column.

As stated above, these earlier results are based on an error threshold of 5 pixels to assess the accuracy of the estimated disparity. We depict in Fig. 13 how the performance varies with different thresholds. As illustrated, a larger threshold leads to a slightly greater improvement over the competing methods.

Finally, we have also performed experiments with stereo

images in a Marina and Outdoor pool, under natural lighting. By knowing the shape of the objects in the scene, we computed ground truth disparity maps by manually picking some corresponding points and interpolating for the rest of the object. Left images along with ground truth disparity maps and the validity maps for different methods are shown in Fig. 14. We compare the results using raw image data, by removing backscatter from images using haze removal method and our method. The percentage of valid points are shown below each validity map. As can be seen, our method improves the results of disparity map estimation by 6% in Marina and 10% in Outdoor pool data comparing to using raw data. Moreover, the haze removal method improves the result by 2% in Outdoor pool while it worsens the results by 9% in Marina image. A possible reason for this behavior in haze removal method is that since the left and the right images enhance separately, this might result in un-even enhancement in corresponding patches and consequently amplify the difference between them.

### VI. CONCLUSIONS

We have proposed a new method to estimate the backscatter in stereo images acquired in scattering media. Based on this, we have devised a new technique to estimate binocular disparity that exploit the depth cues in both the backscatter and the signal components of the raw stereo pair. By taking advantage of these two independent cues, instead of treating the former as noise which has been done in previous stereovision work for scattering media, we are able to improve the performance in disparity map estimation. The improved accuracy is significant in the areas where signal is dominated by the backscatter component. Quantitative assessment of our method has been demonstrated in experiments with synthetic and real data, and knowledge of the ground truth. Future work involves a number of algorithmic enhancements, including the treatment of image regions where the proposed decomposition is ill-conditioned, and incorporating occlusion detection. In particular, the new paradigm can be utilized within other stereo matching techniques.

### ACKNOWLEDGMENT

The authors would like to thank...

### REFERENCES

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, pp. 7–42, 2001.

[2] G. Dudek, P. Giguere, C. Prahacs, S. Saunderson, J. Sattar, L.-A. Torres-Mendez, M. Jenkin, A. German, A. Hogue, A. Ripsman, J. Zacher, E. Miliotis, H. Liu, P. Zhang, M. Buehler, and C. Georgiades, "Aqua: An amphibious autonomous robot," *Computer*, vol. 40, no. 1, pp. 46–53, Jan. 2007.

[3] M. Johnson-Roberson, S. Kumar, O. Pizarro, and S. Willams, "Stereoscopic imaging for coral segmentation and classification," in *MTS/IEEE OCEANS conference*, Sept. 2006, pp. 1–6.

[4] S. Negahdaripour and H. Madjidi, "Stereovision imaging on submersible platforms for 3-d mapping of benthic habitats and sea-floor structures," *IEEE Journal of Oceanic Engineering*, vol. 28, no. 4, pp. 625–650, Oct. 2003.

[5] Y. Swirski, Y. Y. Schechner, B. Herzberg, and S. Negahdaripour, "Stereo from flickering caustics," in *IEEE International Conference on Computer Vision*, 2009.

[6] G. Foresti, "Visual inspection of sea bottom structures by an autonomous underwater vehicle," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 31, no. 5, pp. 691–705, Oct 2001.

[7] J. P. Queiroz-Neto, R. Carceroni, W. Barros, and M. Campos, "Underwater stereo," in *Brazilian Symposium on Computer Graphics and Image Processing*, 2004, pp. 170–177.

[8] Y. S. Heo, K. M. Lee, and S. U. Lee, "Illumination and camera invariant stereo matching," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, June 2008.

[9] L. Wang, R. Yang, and J. Davis, "Brdf invariant stereo using light transport constancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1616–1626, Sept. 2007.

[10] A. Sarafraz, S. Negahdaripour, and Y. Y. Schechner, "Enhancing images in scattering media utilizing stereovision and polarization," in *IEEE Workshop on Applications of Computer Vision (WACV)*, December 2009.

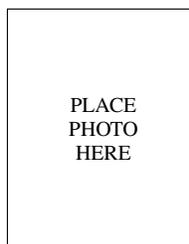
[11] T. Treibitz and Y. Y. Schechner, "Active polarization descattering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 3, pp. 385–399, 2009.

[12] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 20–25 2009, pp. 1956–1963.

[13] R. Fattal, "Single image dehazing," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–9, 2008.

[14] R. T. Tan, "Visibility in bad weather from a single image," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 0, pp. 1–8, 2008.

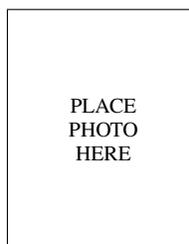
[15] C. D. Mobley, "light and water: radiative transfer in natural waters," 1994, academic Press.



**Amin Sarafraz** Biography text here.



**Shahriar Negahdaripour** Biography text here.



**Yoav Y. Schechner** Biography text here.

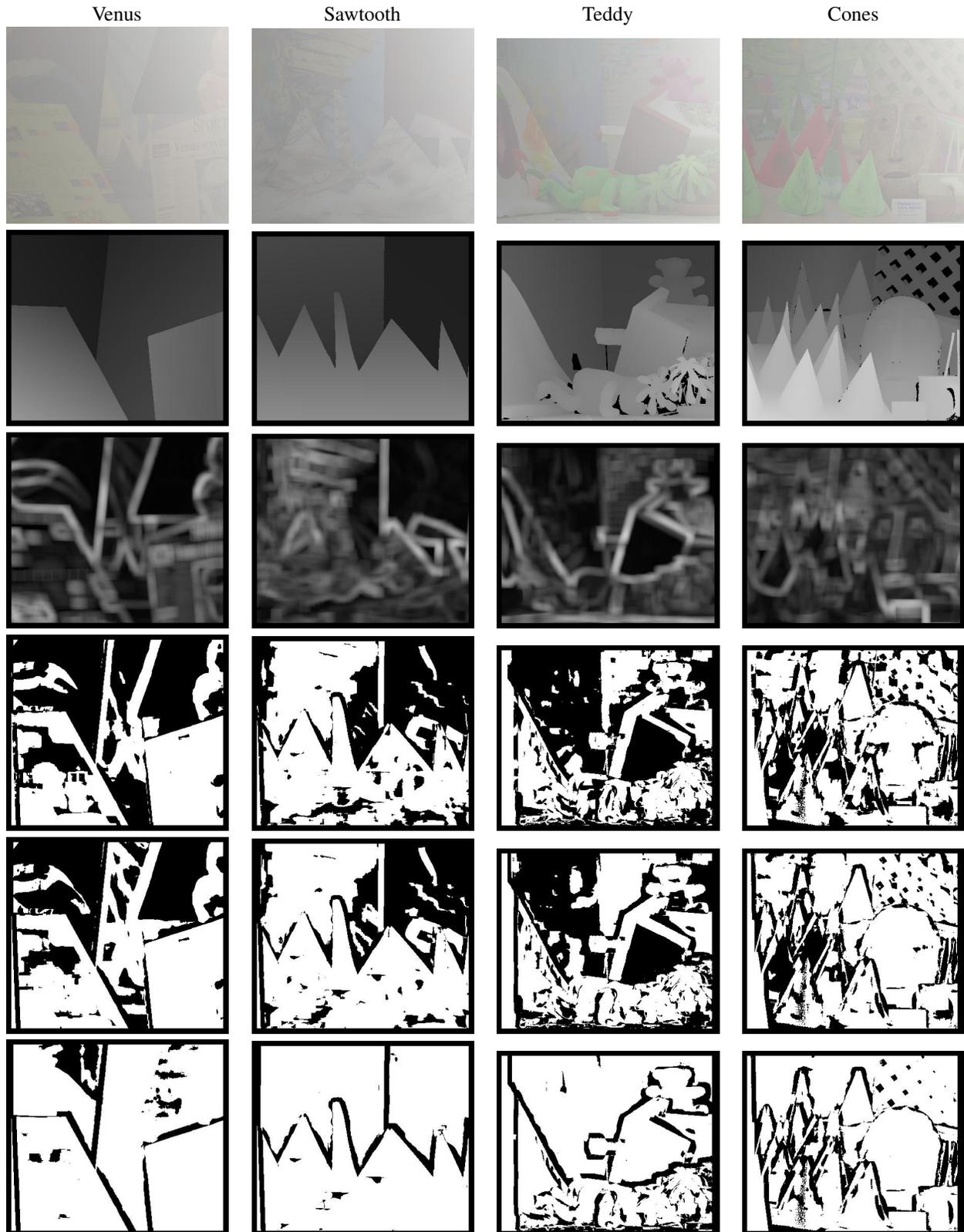


Fig. 5. Synthetic Turbid harbor images (Top row). Ground truth disparity map (Second row). Binary validity maps with good (white) and bad (black) pixels (with threshold set at one-pixel error) are given for the disparity computed from raw data (Third row), signal component (Fourth row), and our method (Bottom row).

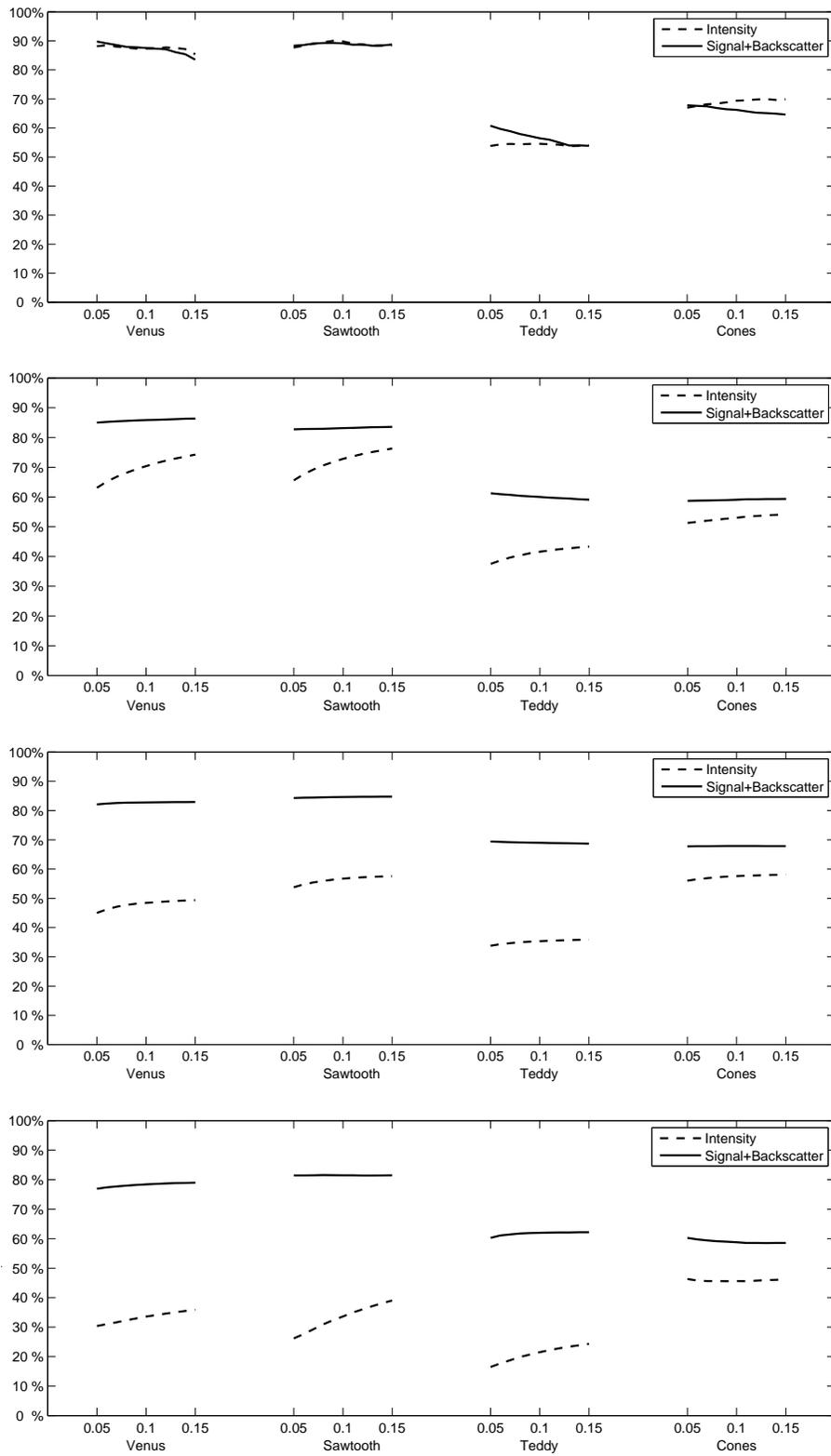


Fig. 6. Percentage of pixels with correct disparity as a function of gradient magnitude threshold for the Synthetic dataset in *Textured* and *Non-Textured* regions. Results for *Textured* regions by thresholding gradient magnitude of (Top row) synthetic images and (Second row) Original images. Results for *Non-Textured* regions by thresholding gradient magnitude of (Third row) synthetic images and (Bottom row) Original images. Gradient magnitude in all images ranges from 0 to 0.3.

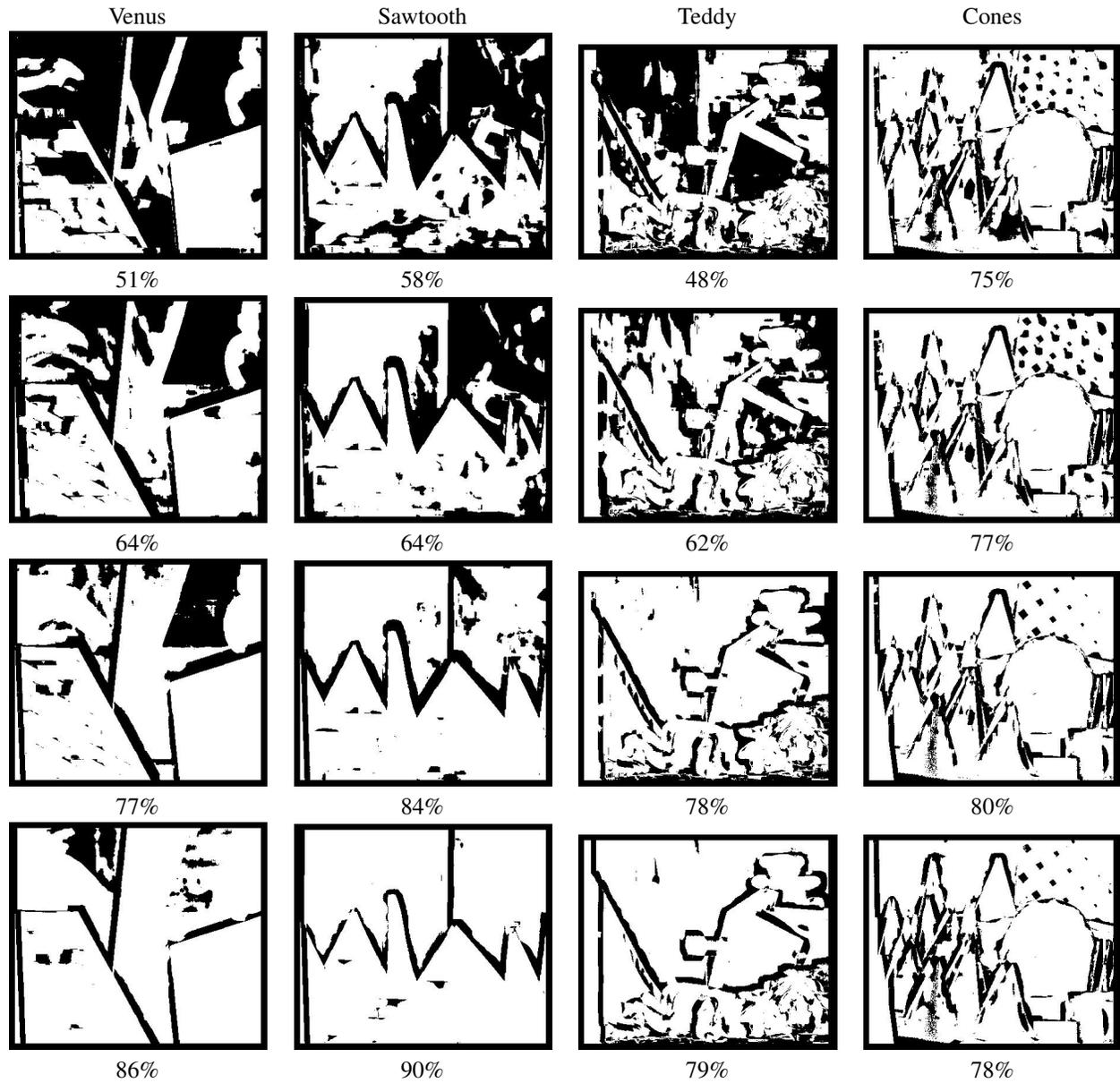


Fig. 9. Comparison of estimated disparity maps by removing backscatter using single image haze removal method and Our method. Validity map of Haze removal method for different images using (top row) the mean of the backscatter at infinity image as a constant value for backscatter at infinity, (second row) the maximum of the backscatter at infinity image as a constant value for backscatter at infinity, (third row) original backscatter at infinity image. (fourth row) Validity maps of different images using our method

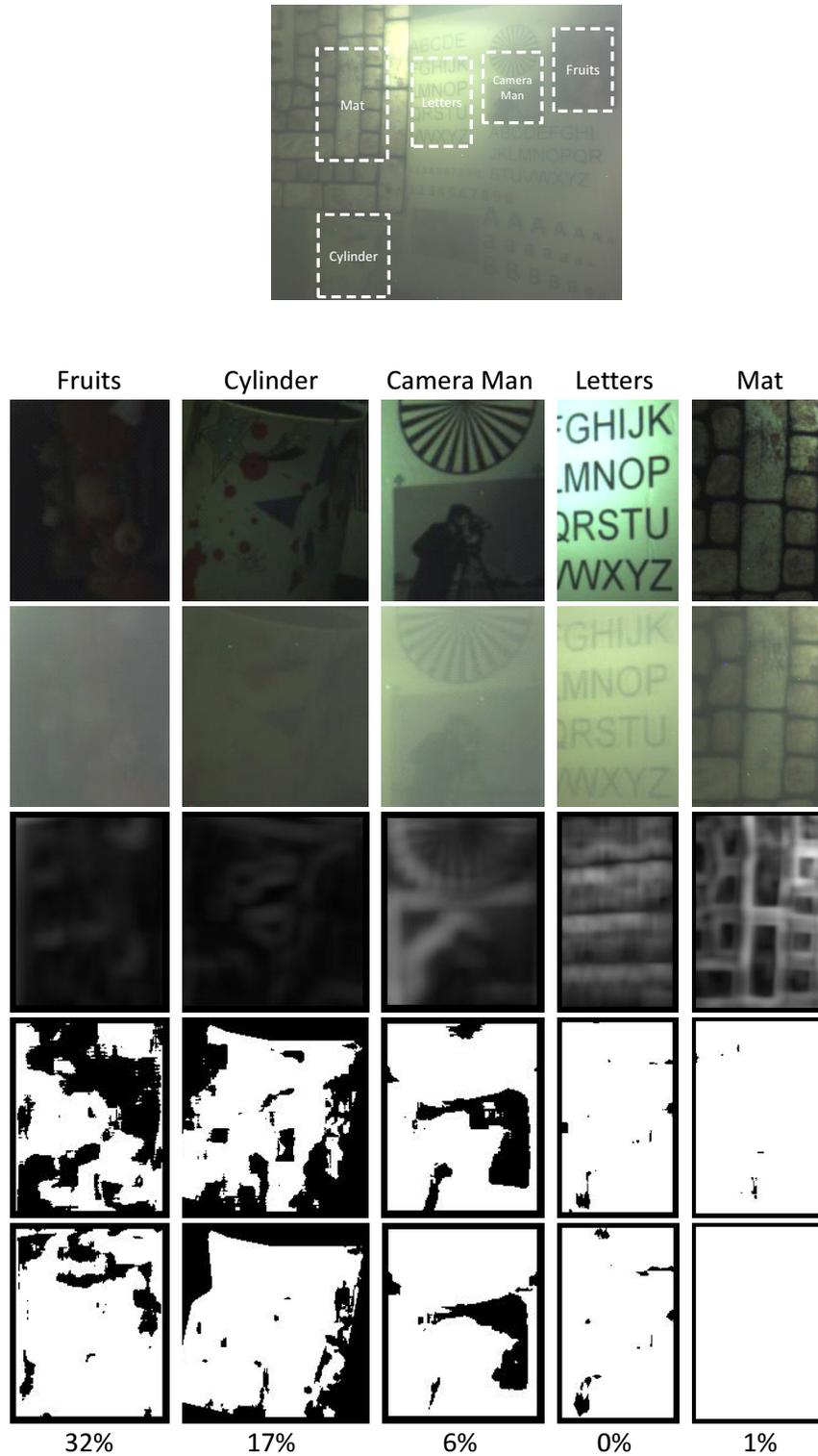


Fig. 12. (first row) parts of the Tank image in clear water, (second row) parts of the Tank image in turbid water, (third row) Contrast map, Binary validity maps with good (white) and bad (black) pixels are given for the disparity computed from (fourth row) raw data and (fifth row) our method. Percentage of improvement in disparity map estimation over each area is shown below its column.

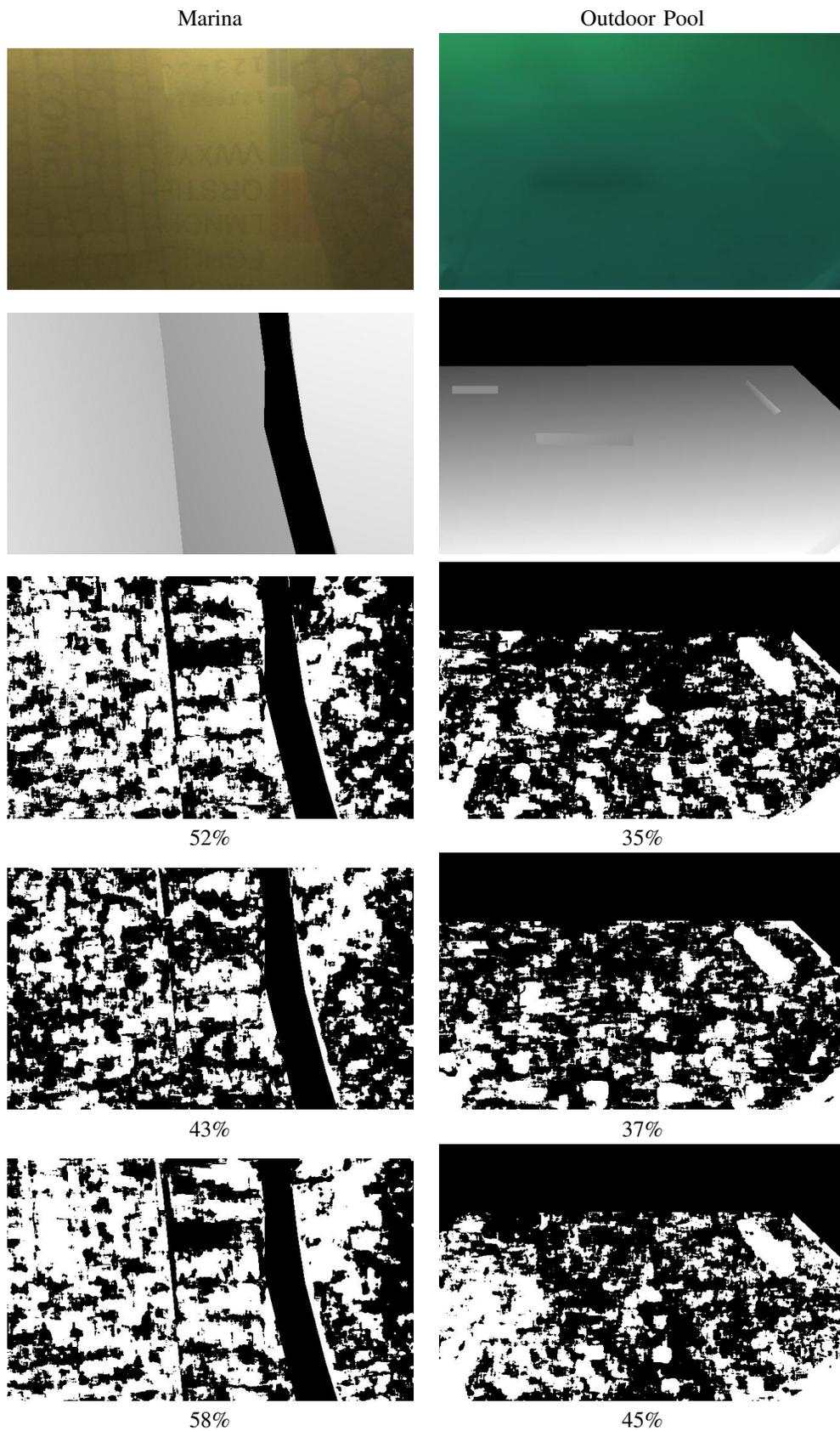


Fig. 14. Results of different methods using images taken in Marina and Outdoor pool: (top row) Left images, (second row) ground truth disparity maps, (third row) validity maps using normalized SSD on raw data, (fourth row) validity maps using normalized SSD on enhanced images after removing backscatter using haze removal method in [12], (bottom row) validity maps using our method