Correspondence

Turbid Scene Enhancement Using Multi-Directional Illumination Fusion

Tali Treibitz and Yoav Y. Schechner, *Member, IEEE*

***Abstract*— Ambient light is strongly attenuated in turbid media. Moreover, natural light is often more highly attenuated in some spectral bands, relative to others. Hence, imaging in turbid media often relies heavily on artificial sources for illumination. Scenes irradiated by an off-axis single point source have enhanced local object shadow edges, which may increase object visibility. However, the images may suffer from severe nonuniformity, regions of low signal (being distant from the source), and regions of strong backscatter. On the other hand, simultaneously illuminating the scene from multiple directions increases the backscatter and fills-in shadows, both of which degrade local contrast. Some previous methods tackle backscatter by scanning the scene, either temporally or spatially, requiring a large number of frames. We suggest using a few frames, in each of which wide field scene irradiance originates from a different direction. This way, shadow contrast can be maintained and backscatter can be minimized in each frame, while the sequence at large has a wider, more spatially uniform illumination. The frames are then fused by post processing to a single, clearer image. We demonstrate significant visibility enhancement underwater using as little as two frames.**

***Index Terms*— Computer vision, image recovery, modeling and recovery of physical attributes, physics-based vision, vision in scattering media.**

I. Introduction

Images taken under artificial illumination in turbid media suffer from backscatter and non-uniform illumination. These phenomena degrade the contrast and visibility in the image, and lower the useable dynamic range. Backscattered light is scattered back from the medium to the camera. It is not reflected from the object. It is an additive component in the image, often veiling the objects. Moreover, non-uniformity is apparent when the field of view is wide.

Under ambient illumination, dynamic range problems are less severe than under artificial lighting and recent works suggest overcoming backscatter using a single hazy [1]–[3] or underwater [4] image. Previous approaches for visibility enhancement under artificial illumination in turbid scenes have relied heavily on scanning. Such current methods require acquisition of long image sequences by structured light [5]–[8] or time-gating [9]–[13]. Ref. [14] required many frames as well, to achieve quality results. Such sequences may lengthen the overall acquisition time. Moreover, such systems are often complex and expensive.

Backscatter can be eliminated if it is modulated between frames. Refs. [15], [16] suggested modulation using polarizers. This approach is fast and simple to implement, but requires mounting of polarizers, which attenuate the signal. In this brief we also suggest post-processing of several frames containing modulated backscatter. However, here lighting modulation is achieved by changing the location of the light source or switching between fixed sources at different locations. This is easy to implement, since available light sources can be used, without a need for polarizers. We show that even two frames can yield results having high visibility.

There are several key ideas in this method. First, by modulating the illumination between frames, the backscatter changes. Thus, each area appears in different levels of contrast in the acquired frames. Second, distinct illumination directions [17] create sharp shadows in different places. Different frames correspond to different lighting directions, hence different shadows and shadings. If all the sources are lit simultaneously, then shadows are filled, reducing object contrast. However, by fusing the raw frames corresponding to distinct light sources, we maintain the contrast created by shadows, making the object more visible. For each scene area, there is a frame (lighting and shadow direction) in which this area has the best contrast. Combining information from multiple frames using an appropriate criterion yields the best contrast for each area in a single output image.

This idea has some relation to the approaches considered in [5], [8]. There, a specialized structured light source illuminates the scene. The resulting frames have clearer areas (stripes or a two dimensional grid) of the scanned scene. The frames are then merged to a single image using knowledge of the structured light. Our approach uses few frames and does not require a structured light source. Our method is also related to other enhancement methods suggested in computer vision that use multiple images: flash/no-flash pairs [18], and multiple exposures [19]. The rest of this brief describes the model for the image acquisition and the image processing methods we apply.

II. Artificial Illumination

In this section we detail the image formation model, following [16]. Consider a perspective underwater camera (Fig. 1). Let $\mathbf{X} = (X, Y, Z)$ be the world coordinates of a point in the water. The world system's axes (X, Y) are set to be parallel to the (x, y) coordinates at the image plane, while Z aligns with the camera's optical axis. The system's origin is at the camera's center of projection. The projection[1] of $\mathbf{X}$ on the image plane is $\mathbf{x} = (x, y)$. Specifcally, an object point at $\mathbf{X}_{\text{obj}}$ corresponds to an image point $\mathbf{x}_{\text{obj}}$. The measured image is

$$I(\mathbf{x}_{\text{obj}}) = L_{\text{obj}}(\mathbf{x}_{\text{obj}})F(\mathbf{x}_{\text{obj}}) + B(\mathbf{x}_{\text{obj}}) \quad (1)$$

[1]For a perspective camera sealed in an underwater housing, the mapping between world and image coordinates depends on the shape of the housing's port. For an elaborate analysis of this mapping, see [20].

Manuscript received August 3, 2011; revised June 26, 2012; accepted June 26, 2012. Date of publication July 16, 2012; date of current version October 12, 2012. This work was supported in part by the Israel Science Foundation under Grant 1031/08, and was conducted in the Ollendorff Minerva Center for Vision and Image Sciences, funded through the BMBF. T. Treibitz is an Awardee of the Weizmann Institute of Science - National Postdoctoral Award Program for Advancing Women in Science and was supported in part by NSF Grant ATM-0941760. The work of Y. Schechner was supported in part by the Department of the Navy Grant N62909-10-1-4056 issued by the Office of Naval Research Global. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Wai-Kuen Cham.

T. Treibitz is with the Department of Computer Science and Engineering., University of California, San Diego, CA 92093 USA (e-mail: tali@cs.ucsd.edu).

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

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2012.2208978

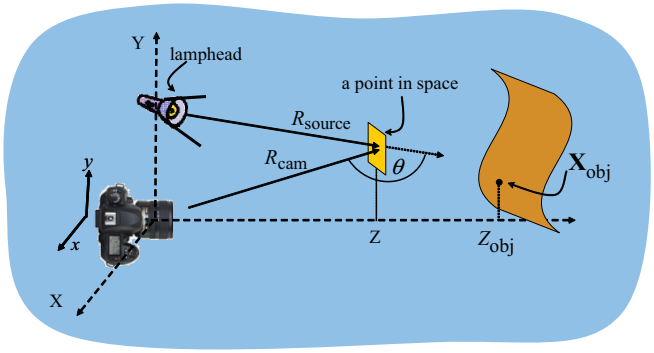

Fig. 1. Underwater camera imaging setup with an artificial light source [16]. The variables are detailed in the text.

where $B(\mathbf{x}_{obj})$ is the backscatter [21]–[23]. The variable $L_{obj}(\mathbf{x}_{obj})$ is the object radiance we would have sensed had no disturbances been caused by the medium along the line of sight (LOS), and under uniform illumination. Propagation of light to the object and then to the camera via the medium yields an attenuated [21], [22] signal, where $F(\mathbf{x}_{obj})$ is a falloff function described below.[2]

Consider for a moment a single illumination point source. From this source, light propagates a distance R_{source} to $\mathbf{X}_{obj}$, which is at distance R_{cam} from the sealed camera. The attenuation by the medium is characterized by an attenuation coefficient c. In addition, free space propagation creates a $1/R^2_{source}$ irradiance falloff. Hence

$$F(\mathbf{x}_{obj}) = \frac{\exp\left\{-c\left[R_{source}(\mathbf{X}_{obj}) + R_{cam}\right]\right\}}{R^2_{source}(\mathbf{X}_{obj})} Q(\mathbf{X}_{obj}) \qquad (2)$$

where $Q(\mathbf{X})$ expresses the non-uniformity of the scene irradiance, created solely by the anisotropy of the illumination (Fig. 2). Let I^{source} be the irradiance of a small region in the volume [21] by a small illumination source of radiance L^{source}:

$$I^{source}(\mathbf{X}) = L^{source}\left[1/R^2_{source}(\mathbf{X})\right] \exp[-cR_{source}(\mathbf{X})]Q(\mathbf{X}). \qquad (3)$$

The backscatter for a small illumination source is given [16], [21] by integration along the LOS

$$B(\mathbf{x}_{obj}) = \int_{R_{cam}=0}^{R_{cam}(\mathbf{X}_{obj})} b[\theta(\mathbf{X})] I^{source}(\mathbf{X}) \exp[-cR_{cam}(\mathbf{X})] dR_{cam}, \quad \mathbf{X} \in \text{LOS}. \qquad (4)$$

Here $\theta \in [0, \pi]$ is the scattering angle, and b is the scattering coefficient of the medium: it expresses the ability of an infinitesimal medium volume to scatter flux to θ. The integration range in Eq. (4) is bounded by the object distance on the LOS. Therefore, the backscatter accumulates (increases) with the object distance.

Changing the camera-illumination setup changes R_{source} and Q for each pixel and thus affects the falloff $F(\mathbf{x}_{obj})$ and backscatter $B(\mathbf{x}_{obj})$ in each pixel (Eqs. 1–4). Moreover, L_{obj} also changes, as shadows change with the direction of illumination, as in [17]. Fig. 3(a) illustrates an example of two different illumination setups employed to image a certain scene. Fig. 3(b) illustrates the image when both sources are used simultaneously. Fig. 3(c) illustrates the desired outcome of our method. The next sections detail the post processing we apply on the acquired frames and Sec. V analyzes how many frames are needed as an input.

[2]We consider here only single scattering. Multiple scattering blurs the results. It has been shown that blur is often only a secondary effect with regard to image deterioration [24] in turbid media, while contrast-loss is a prime effect. Contrast-loss under artificial illumination is mainly due to backscatter.

Fig. 2. Illustration [16] of an anisotropic illumination pattern $Q(\mathbf{X})$. Even in the same radial distance from the lamphead, the lighting changes laterally.

III. View Enhancement

Following Sec. II, the imaging process yields two (or more) frames of the same underlying scene, illuminated from different directions. In this section we detail how these frames are combined into a single image with higher contrast, visibility, and dynamic range. The frames are merged using image fusion. We used fusion based on pyramid decomposition [25], [26]. However, other techniques for image fusion exist in the literature and may be used instead [27]–[29].

A. Image Fusion Using Laplacian Pyramids

A popular fusion scheme [25], [26] is based on decomposing each frame into two pyramids. In the *Gaussian pyramid*, each level contains a lower resolution version of the previous level, i.e., higher levels have less high frequency content. The *Laplacian pyramid* is constructed from differences between Gaussian levels. Thus, the levels of the Laplacian pyramid contain band pass versions of the original image. The Laplacian pyramid together with the coarsest level of the Gaussian pyramid provide a complete representation of the image, i.e., the full resolution image can be reconstructed given them.

We denote images in the Gaussian and Laplacian pyramids by G_p and L_p respectively, where the index p denotes a level in the pyramid. Let N denote the total number of pyramid levels, which is a parameter for the algorithm. For G_p, $p \in [1, N]$, while for L_p, $p \in [0, N-1]$. The original image is G_0. The image at the coarsest level is G_N. The image at level p of the Gaussian pyramid is obtained by filtering level $(p-1)$ with a gaussian filter $\Psi(\mathbf{u})$, and then subsampling

$$\mathrm{G}_p(\mathbf{x}) = \sum_{\mathbf{u}} \Psi(\mathbf{u}) \mathrm{G}_{p-1}(2\mathbf{x} + \mathbf{u}). \qquad (5)$$

Here $\mathbf{u}$ spans the support of Ψ. Let $\hat{\mathrm{G}}_{p+1}$ be a magnification of a coarse image G_{p+1} to an image having the same size as G_p:

$$\hat{\mathrm{G}}_p(\mathbf{x}) = \sum_{\mathbf{u}} \Psi(\mathbf{u}) \mathrm{G}_p\left(\left\lfloor \frac{\mathbf{x}+\mathbf{u}}{2} \right\rfloor\right). \qquad (6)$$

Thus, G_p and $\hat{\mathrm{G}}_{p+1}$ are images of the same size, but $\hat{\mathrm{G}}_{p+1}$ is coarser. Then, the Laplacian pyramid levels are obtained by

$$\mathrm{L}_p = \mathrm{G}_p - \hat{\mathrm{G}}_{p+1}. \qquad (7)$$

For image fusion, the pyramid decompositions of the input frames are merged to a new pyramid, $\left\{\mathrm{L}^{new}_p\right\}_{p=1}^{N-1} \cup \mathrm{G}^{new}_N$. Then, this new pyramid is decoded to form the fusion result. We detail this process in the next section.

B. Merger Criterion

Let $k \in [1, M]$ be the index of an input frame, where M is the number of the input frames. Each level p in the new Laplacian pyramid L^{new} is processed distinctly from the other levels and is a function of the same level in the input Laplacian pyramids:

$$\mathrm{L}^{new}_p = \mathcal{R}\left(\left\{\mathrm{L}^{[k]}_p\right\}_{k=1}^{M}\right), \quad p \in [1, N-1]. \qquad (8)$$

Here $\mathcal{R}$ is a fusion criterion.

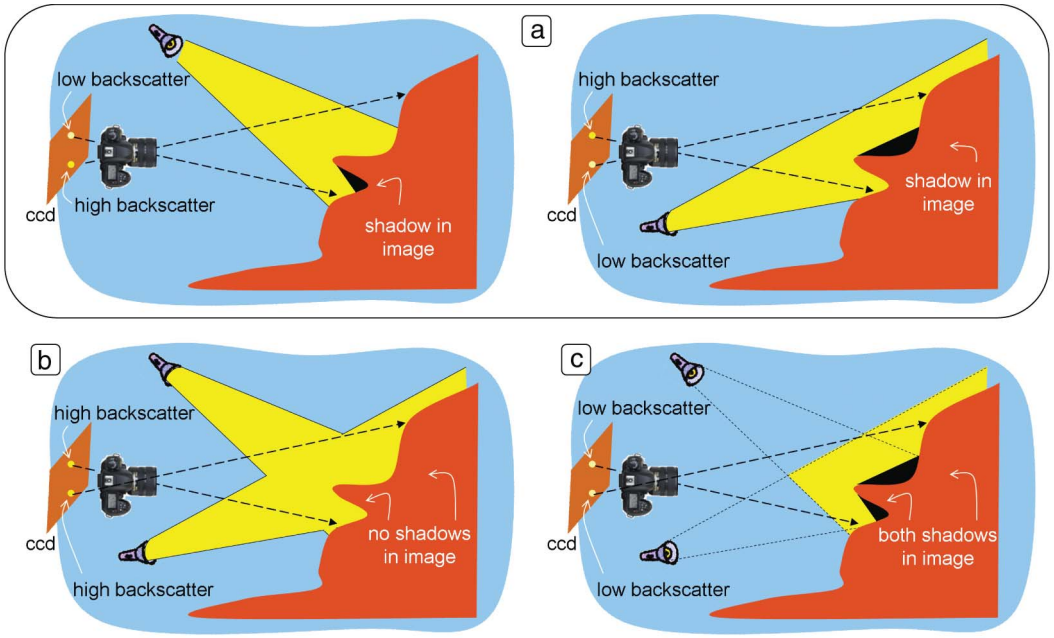

Fig. 3. (a) Underwater scene illuminated from two different directions. The shadows vary, as well as the intensity of the backscatter in each pixel. (b) Same scene, illuminated with both sources simultaneously. There are no deep shadows, and there is high backscatter. (c) Desired outcome of image fusion. Both shadows appear, while the backscatter component is lower.

Various fusion criteria $\mathcal{R}$ exist in the literature [29]. Each criterion acts to maximize a different property of the image, e.g., intensity, gradients or contrast. We tried several criteria. Since backscatter degrades contrast, we found that contrast maximization [30] is a criterion which yields the best appearing results, for our turbid scenarios.

Weber's definition for contrast [31] is

$$\text{contrast} = \frac{\text{object brightness} - \text{background brightness}}{\text{background brightness}}. \quad (9)$$

Having a Gaussian pyramid decomposition, the background of an object in level p can be defined as the object area in a coarser level $p+1$. Thus, using the pyramid notation, Eq. (9) is expressed as

$$\mathrm{C}_p = \frac{\mathrm{G}_p}{\hat{\mathrm{G}}_{p+1}} - 1 \quad (10)$$

resulting in a *contrast pyramid* C_p, $p \in [0, N-1]$.

In each scale p, the values for pixel $\mathbf{x}$ are taken from the input frame $\hat{k}_p(\mathbf{x})$ that has the highest contrast in this location, at that scale:

$$\hat{k}_p(\mathbf{x}) = \underset{k \in [1,M]}{\operatorname{argmax}} \left\{ |\mathrm{C}_p^{[k]}(\mathbf{x})| \right\}. \quad (11)$$

Then, the new Laplacian pyramid is constructed based on the contrast pyramids of the input frames $\mathrm{C}_p^{[k]}$, $p \in [0, N-1]$:

$$\mathrm{L}_p^{\text{new}}(\mathbf{x}) = \mathrm{L}_p^{[\hat{k}_p(\mathbf{x})]}(\mathbf{x}). \quad (12)$$

C. Countering Nonuniform Illumination

Eq. (12) creates all levels of a new pyramid

$$\left\{ \mathrm{L}_p^{\text{new}} \right\}_{p=1}^{N-1}$$

except of the coarsest level, $\mathrm{G}_N^{\text{new}}$. In various image fusion applications, the coarsest level is usually combined by averaging the coarse levels of the input frames or taking their maximum [29]. The coarse level holds the lowest spatial frequencies in the decomposition, often associated with illumination [32]. We found that in our scenarios, using values from the coarse levels of the input frames maintains the non-uniform scene irradiance of each frame. This results in an image having a low dynamic range.

To make the illumination more uniform, we assign the coarse level a *constant* value, which is the spatial mean of all input coarse levels

$$\mathrm{G}_N^{\text{new}} = \frac{1}{M N_x} \sum_{\mathbf{x}} \sum_{k=1}^{M} \mathrm{G}_N^{[k]}(\mathbf{x}) \quad (13)$$

where N_x is the number of the pixels in the coarse level. This action nulls low frequency variations, leaving only a uniform level that aliases as uniform illumination. This idea is similar to homomorphic filtering [32]. Such assignment yields a somewhat arbitrary value to the coarse level. For this approach to work well, the parameter N should be chosen carefully, so that the coarsest level mainly represents the illumination spatial frequencies, rather than object frequencies. In our experiments, the input images were of size 1500×1000 pixels. We found empirically that when the coarse level is of size $\sim 8 \times 8$, the results appear to have more uniform illumination, while demonstrating the local object contrasts.

IV. Results

Our implementation is based on a code from [33], with modifications. The method above is described for grayscale images. In Eq. (12) each color channel (in RGB color space) is treated separately. In Eq. (13) we assign the same value to the coarse level for all three color channel (the average value). This gives an added functionality of color balancing.

Figs. 4 and 5 demonstrate the results of our method in two underwater experiments. Here, we used two frames per scene, in each of which a wide angle source illuminated the scene from a different direction. Images were taken with a Nikon D100 in a Sealux housing with a dome port, and underwater AquaVideo SuperNova

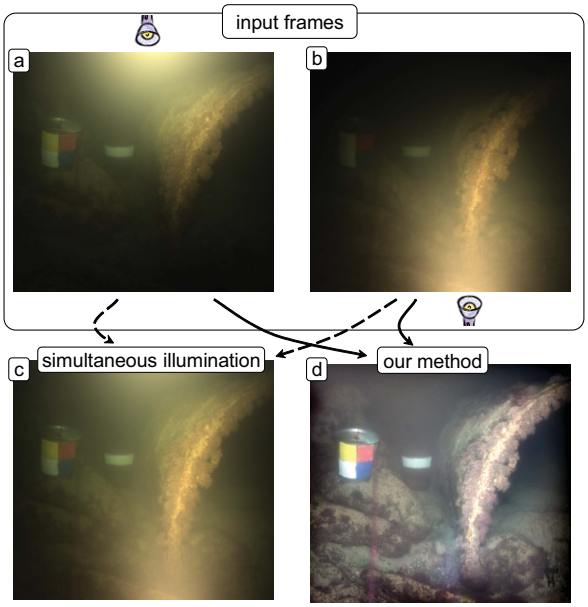

Fig. 4. Result of an underwater experiment using multidirectional illumination. (a) and (b) Raw frames. (c) Image as if the scene was illuminated by both sources simultaneously, generated by combining frames (a) and (b). (d) Result of our method.

light sources having 80W Halogen bulbs. Images were taken while diving in the Mediterranean, in the ancient Caesarea port, with a visibility of a few meters.

In both experiments, the residual backscatter following our method is hardly visible, compared to significant backscatter in the original images. This improves the dynamic range of the resulting images, revealing objects that are veiled in the input frames.

To demonstrate the benefit of using multiple frames, Fig. 6 demonstrates the results of applying Eq. (13) on a single frame per scene. Then, Eq. (13) serves to homogenize the illumination and color balance. While the images are enhanced relative to their corresponding inputs in Fig. 4a-c, strong backscatter is still visible.

Fig. 7 shows results of fusing the input images from Fig. 4a,b, using two fusion methods, different from the one we proposed. In Fig. 7a,b, fusion is done by wavelet decomposition (using Daubechies Spline wavelet) and taking the *maximum* in each level. The coarse level of the result is obtained by averaging the coarse levels of the input images. As seen in Fig. 7a,b, the result is better than each of the input images separately, but has higher backscatter than in our method. Using this criterion with Laplacian decomposition yields a similar result. In Fig. 7c, the decomposition is done using pyramids. The merge criterion is a variant of Eq. (10) where the contrast is the ratio of Laplacian pyramids (without subtracting 1), $C_p = G_p/\hat{G}_{p+1}$. The coarse level is constructed using our method (Sec. III-C). This method is similar to the method we use and yields good results. However, note that in this case, less shadows are preserved, for example, in the top-left and bottom-right areas of the image. Thus, our result (Fig. 4d) demonstrated a better contrast.

V. How Many Sources are Needed?

Our method is general and works on any number of input frames. The number of needed frames depends on the extent of backscatter in the image and its spatial distribution in the scene. To achieve a result effectively clear of backscatter, each area of the scene has to appear clear in at least one of the input frames. As described in Sec. II, the backscatter intensity at each pixel depends on the water properties and setup geometry. The number of needed frames also depends on a definition of acceptable clarity in output images. This level depends on the end application [34] and the noise levels in the raw images [35].

Following Eq. (4), the amount of backscatter depends on the scattering coefficient $b(\theta)$. In this brief we consider cases where $b(\theta)$ is not low, and thus a single image is not satisfactory. Therefore, our method requires at least two frames. Obviously, adding more frames yields better results: with more lighting directions, chances increase for low backscatter values and more shadows across the entire image stack. However, adding more frames makes the acquisition more complex and increases the acquisition time. At some stage, adding more lighting directions yields diminishing returns on the result. Thus, it is up to the user to decide how much effort is worth, for getting more quality at the expense of more time.

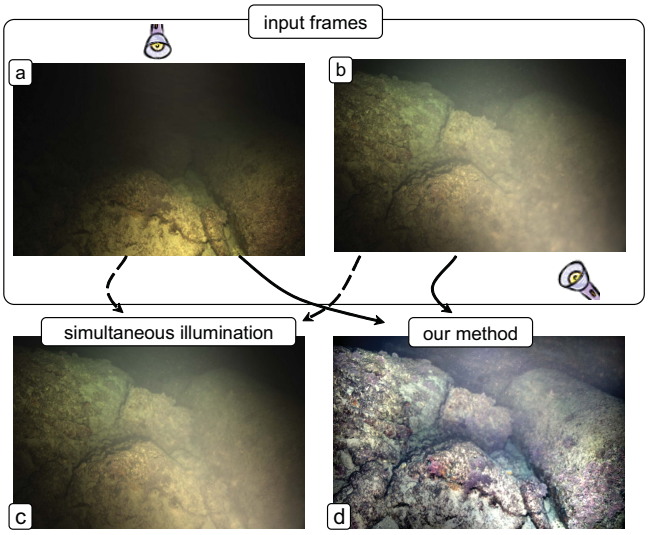

Fig. 5. Result of an underwater experiment using multidirectional illumination. (a) and (b) Raw frames. (c) Image as if the scene was illuminated by both sources simultaneously, generated by combining frames (a) and (b). (d) Result of our method.

Fig. 6. Results of applying (13) on any single frame. This yields correction of the nonuniform illumination, and also color balance. The backscatter remains in the images. (a)–(c) Enhancements of images (a)–(c) from Fig. 6.

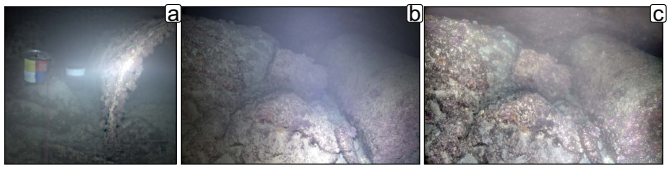

Fig. 7. Comparison of other fusion methods on the input images. (a) and (b) Fusion is done using wavelet decomposition and maximum criterion on the input images from Figs. 4(a) and (b) and 5(a) and (b), respectively. Both results are better than the input images, but have less contrast than using our method. (c) Laplacian pyramid fusion using a modified contrast criterion ($C_p = G_p/\hat{G}_{p+1}$) yields a good result, but with less shadows relative to our method. For example, this is seen in the top-left and bottom-right areas of the image. This lack of shadows decreases contrast relative to our method.

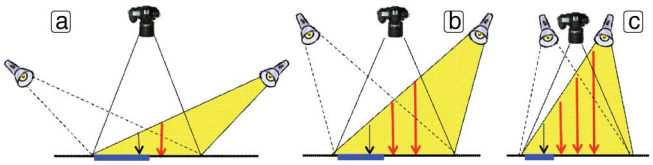

Fig. 8. Backscatter levels in an image change as a function of setup. Arrows indicate the backscatter intensity, where red is an "unacceptable" level, for this case. The blue line indicates an area in the image that has an "acceptable" level. (a) About half of the image has low backscatter. Two input images are required for our method. The dashed light cone demonstrates a possible second lighting direction. (b) and (c) Under the same conditions, the two sources do not suffice to get acceptable quality across the field. Hence, additional images are required.

How many frames are needed for an "acceptable" result? The imaging setup has a significant influence on the backscatter, as detailed in Sec. II. The main parameters are the object distance (closer is better) and the overlap between the light cone and the field of view (less is better). In Fig. 8 we illustrate three different camera-illumination setups, with intensity of backscatter increasing from (a) to (c). Generally, the wider the camera-illumination baseline, the weaker the backscatter. However, there is an effective limit to the distance of a object, as irradiance falls off with this distance. This decreases the SNR. Ref. [36] analyzes this trade-off when noise is mainly signal independent. The analysis shows that there is an optimal position for the light source, that depends on the water and imaging parameters. In addition, even if not limited by SNR, there is usually a technical limit to placing a light source far from the camera.

In Fig. 8, arrows indicate the backscatter intensity, where red is an "unacceptable" level. A blue line indicates an area in the image which has an "acceptable" level. For example, in Fig. 8a, approximately half of the image is clear. Thus, in Fig. 8a, our method requires only two input images, taken with the same baseline. In Fig. 8b,c, the clear portion is smaller, thus the method needs additional input images, to clear the full field of view.

VI. Discussion

We presented a simple method to enhance visibility of scenes imaged under artificial illumination in a turbid medium. As opposed to various methods that rely on scanning and specialized hardware, we demonstrate good results using only two frames with off-the-shelf fixed illumination sources. We believe that in addition to underwater environments, the method can be applied to fog. It is possible that other scattering media, such as some tissue and microscopic specimen may benefit from this approach as well. The method is limited in the sense that it does not remove backscatter per se. In any area, the apparent backscatter in the result corresponds to the minimum backscatter along the acquired frames. Polarizers [16] can be combined to further remove backscatter.

Here we show visibility enhancement without using multi-directional lighting to estimate the 3D scene structure. In principle, multi-directional lighting can be used to estimate shape, based on photometric stereo, shape from shadows or shading [37]. Refs. [8], [38] have demonstrated underwater photometric stereo for collimated distant light sources. However, this assumption is often not satisfied in underwater imaging. Ref. [39] simulated the possibility of photometric stereo using simulated near-by light sources. The formulation has to account for non-uniform backscatter as well. Hence, further work is needed so that data acquired as described above is used for *both* visibility enhancement and shape estimation.

Acknowledgment

This work relates to Department of the Navy Grant N62909-10-1-4056 issued by the Office of Naval Research Global. The United States Government has a royalty-free license throughout the world in all copyrightable material contained herein.

References

[1] R. Fattal, "Single image dehazing," *J. ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–9, 2008.
[2] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," in *Proc. IEEE Comput. Vision Pattern Recogn. Conf.*, Jun. 2009, pp. 1956–1963.
[3] R. T. Tan, "Visibility in bad weather from a single image," in *Proc. IEEE Comput. Vision Pattern Recogn. Conf.*, Jun. 2008, pp. 1–8.
[4] N. Carlevaris-Bianco, A. Mohan, and R. Eustice, "Initial results in underwater single image dehazing," in *Proc. IEEE MTS OCEANS*, Sep. 2010, pp. 1–8.
[5] J. S. Jaffe, "Enhanced extended range underwater imaging via structured illumination," *Opt. Express*, vol. 18, no. 12, pp. 328–340, 2010.
[6] D. Kocak, F. Dalgleish, F. Caimi, and Y. Schechner, "A focus on recent developments and trends in underwater imaging," *Marine Technol. Soc. J.*, vol. 42, no. 1, pp. 52–67, 2008.
[7] M. Levoy, B. Chen, V. Vaish, M. Horowitz, I. McDowall, and M. Bolas, "Synthetic aperture confocal imaging," *J. ACM Trans. Graphics*, vol. 23, no. 3, pp. 825–834, 2004.
[8] S. G. Narasimhan, S. K. Nayar, B. Sun, and S. J. Koppal, "Structured light in scattering media," in *Proc. IEEE Comput. Vision 10th Int. Conf.*, Oct. 2005, pp. 420–427.
[9] S. G. Demos and R. R. Alfano, "Temporal gating in highly scattering media by the degree of optical polarization," *Opt. Lett.*, vol. 21, no. 2, pp. 161–163, 1996.
[10] G. R. Fournier, D. Bonnier, L. J. Forand, and P. W. Pace, "Range-gated underwater laser imaging system," *Opt. Eng.*, vol. 32, pp. 2185–2190, Sep. 1993.
[11] S. Harsdorf, R. Reuter, and S. Tönebön, "Contrast-enhanced optical imaging of submersible targets," *Proc. SPIE*, vol. 3821, pp. 378–383, Jun. 1999.
[12] M. P. Strand, "Imaging model for underwater range-gated imaging systems," *Proc. SPIE*, vol. 1537, pp. 151–160, Dec. 1991.
[13] B. A. Swartz and J. D. Cummings, "Laser range-gated underwater imaging including polarization discrimination," *Proc. SPIE*, vol. 1537, pp. 42–56, Dec. 1991.
[14] S. K. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar, "Fast separation of direct and global components of a scene using high frequency illumination," *J. ACM Trans. Graphics*, vol. 25, no. 3, pp. 935–944, 2006.
[15] T. Treibitz and Y. Y. Schechner, "Instant 3descatter," in *Proc. IEEE Comput. Vision Pattern Recogn.*, 2006, pp. 1861–1868.
[16] T. Treibitz and Y. Y. Schechner, "Active polarization descattering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 3, pp. 385–399, Mar. 2009.
[17] F. P. León, "Automated comparison of firearm bullets," *Forensic Sci. Int.*, vol. 156, no. 1, pp. 40–50, 2006.
[18] A. Agrawal, R. Raskar, S. K. Nayar, and Y. Li, "Removing photography artifacts using gradient projection and flash-exposure sampling," *J. ACM Trans. Graphics*, vol. 24, no. 3, pp. 828–835, 2005.
[19] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. 24th Ann. Conf. Comput. Graphics Interactive Tech.*, 1997, pp. 369–387.
[20] T. Treibitz, Y. Y. Schechner, C. Kunz, and H. Singh, "Flat refractive geometry," in *Proc. IEEE Comput. Vision Pattern Recogn. Conf.*, Jun. 2012, pp. 1–8.
[21] J. S. Jaffe, "Computer modelling and the design of optimal underwater imaging systems," *IEEE J. Oceanic Eng.*, vol. 15, no. 2, pp. 101–111, Apr. 1990.
[22] B. L. McGlamery, "A computer model for underwater camera system," *Proc. SPIE*, vol. 208, pp. 221–231, Sep. 1979.
[23] C. D. Mobley, *Light and Water: Radiative Transfer in Natural Waters*. New York: Academic, 1994.
[24] Y. Y. Schechner and N. Karpel, "Recovery of underwater visibility and structure by polarization analysis," *IEEE J. Oceanic Eng.*, vol. 30, no. 3, pp. 570–587, Jul. 2005.
[25] E. Adelson, C. Anderson, J. Bergen, P. Burt, and J. Ogden, "Pyramid methods in image processing," *RCA Eng.*, vol. 29, no. 6, pp. 33–41, 1984.

[26] P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, Apr. 1983.
[27] H. Li, B. Manjunath, and S. Mitra, "Multisensor image fusion using the wavelet transform," *Graph. Models Image process.*, vol. 57, no. 3, pp. 235–245, 1995.
[28] A. Waxman, A. Gove, D. Fay, J. Racamato, J. Carrick, M. Seibert, and E. Savoye, "Color night vision: Opponent processing in the fusion of visible and ir imagery," *Neural Netw.*, vol. 10, no. 1, pp. 1–6, 1997.
[29] M. Smith and J. Heather, "A review of image fusion technology in 2005," *Proc. SPIE*, vol. 5782, pp. 29–45, Mar. 2005.
[30] A. Toet, "Image fusion by a ratio of low pass pyramid," *Pattern Recogn. Lett.*, vol. 9, no. 4, pp. 245–253, 1989.
[31] E. Peli, "Contrast in complex images," *J. Opt. Soc. Amer.*, vol. 7, no. 10, pp. 2032–2040, 1990.
[32] R. Garcia, T. Nicosevici, and X. Cufí, "On the way to solve lighting problems in underwater imaging," in *Proc. IEEE MTS OCEANS*, Oct. 2002, pp. 1018–1024.
[33] O. Rockinger. *The Image Fusion Toolbox for* MATLAB. (1999) [Online]. Available: http://www.metapix.de/toolbox.htm
[34] T. Treibitz and Y. Y. Schechner, "Recovery limits in pointwise degradation," in *Proc. IEEE Comput. Photography Int. Conf.*, Apr. 2009, pp. 1–8.
[35] T. Treibitz and Y. Y. Schechner, "Polarization: Beneficial for visibility enhancement?" in *Proc. IEEE Comput. Vision Pattern Recogn. Conf.*, Jun. 2009, pp. 525–532.
[36] M. Gupta, S. Narasimhan, and Y. Y. Schechner, "On controlling light transport in poor visibility environments," in *Proc. IEEE Comput. Vision Pattern Recogn. Conf.*, Jun. 2008, pp. 1–8.
[37] R. Szeliski, *Computer Vision: Algorithms and Applications*. New York: Springer-Verlag, 2010.
[38] S. Negahdaripour, H. Zhang, and X. Han, "Investigation of photometric stereo method for 3-d shape recovery from underwater imagery," in *Proc. IEEE MTS OCEANS*, Oct. 2002, pp. 1010–1017.
[39] N. Kolagani, J. Fox, and D. Blidberg, "Photometric stereo using point light sources," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 1992, pp. 1759–1764.

Multiscale Distance Matrix for Fast Plant Leaf Recognition

Rongxiang Hu, Wei Jia, Haibin Ling, and Deshuang Huang

***Abstract*— In this brief, we propose a novel contour-based shape descriptor, called the multiscale distance matrix, to capture the shape geometry while being invariant to translation, rotation, scaling, and bilateral symmetry. The descriptor is further combined with a dimensionality reduction to improve its discriminative power. The proposed method avoids the time-consuming pointwise matching encountered in most of the previously used shape recognition algorithms. It is therefore fast and suitable for real-time applications. We applied the proposed method to the task of plan leaf recognition with experiments on two data sets, the Swedish Leaf data set and the ICL Leaf data set. The experimental results clearly demonstrate the effectiveness and efficiency of the proposed descriptor.**

***Index Terms*— Cost matrix, inner distance, multiscale distance matrix (MDM), plant leaf, shape recognition.**

I. Introduction

Shape is one of the most important features of an object. It plays a key role in many object recognition tasks, in which objects are easily distinguished by shape rather than other features such as edge, corner, color, and texture. There are usually two critical parts in a shape recognition approach, shape representation and shape matching. According to choices of shape representation, shape recognition approaches can be generally divided into two classes, i.e., contour-based and region-based, respectively [1].

In the past decade, research on contour-based shape recognition [2]–[17] is more active than that on region-based due to the following reasons [1]: Firstly, human beings are thought to discriminate shapes mainly by contour features. Secondly, in many shape applications only the shape contour is of interest, while the interior content is less important. Similarly, in this brief, we focus on contour-based shape recognition. Several important contour-based approaches have recently been proposed. Petrakis *et al*. [3] presented an effective contour-based approach using Dynamic Programming (DP), which is invariant to translation, scaling and rotation. Belongie *et al*. [2] proposed a shape feature called Shape Context (SC), which describes a shape by a set of 2-D histograms capturing landmark distributions. Ling *et al*. [10] extended SC to the Inner-Distance SC (IDSC) by replacing the Euclidean distance with the articulation insensitive inner-distance. McNeill *et al*. [6] introduced a multiscale shape matching algorithm named Hierarchical Procrustes Matching (HPM), which investigates shape matching at a variety of different positions. Felzenszwalb *et al*. [9] described a hierarchical shape representation called Shape Tree (ST) to capture shape geometry at different levels of resolution. Xu *et al*. [13] proposed a shape descriptor called Contour Flexibility (CF), which represents the deformable potential at each point on the contour. From these approaches, we conclude that the relative positions between the contour points contain rich information about the structure of objects, and a multiscale representation can better capture the geometric propensities of a shape.

Although the aforementioned contour-based approaches have reported promising recognition performances, they have to face a *crucial* problem, i.e., how to solve the correspondence between two shapes in the matching stage. The solution often requires computing the distance between the two shapes as a sum of matching errors between corresponding points or segments. Many existing contour-based approaches have applied DP procedures to address this problem, which is very time consuming [2]–[4]. As a result, alternative solutions that are computationally more efficient are desired for real-time applications [1].

In this brief, we propose a novel contour-based shape descriptor named Multiscale Distance Matrix (MDM) to capture the geometric structure of a shape while being invariant to translation, rotation, scaling, and bilateral symmetry. When applying MDM to shape recognition, there is no need to use DP to build point wise correspondence, which makes MDM an efficient shape descriptor. In addition, MDM is flexible in the underlying building distances: either the Euclidean distance or other metrics can be utilized in MDM to compute the dissimilarity of two shapes. Furthermore, we apply

Manuscript received June 16, 2011; revised February 17, 2012; accepted June 22, 2012. Date of publication August 2, 2012; date of current version October 12, 2012. This work is supported in part by the National Science Foundation of China, under Grant 61175022, and a grant from the Knowledge Innovation Program, Chinese Academy of Sciences. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ton Kalker.

R. Hu is with the Hefei Institutes of Physical Science, Chinese Academy of Science, Hefei 230031, China, and also with the Department of Automation, University of Science and Technology of China, Hefei 230027, China (e-mail: hurongxiang2008@gmail.com).

W. Jia and D. Huang are with the Hefei Institutes of Physical Science, Chinese Academy of Science, Hefei 230031, China (e-mail: icg.jiawei@gmail.com; dshuang@iim.ac.cn).

H. Ling is with the Department of Computer and Information Science, Temple University, Philadelphia, PA 19122 USA (e-mail: hbling@temple.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2012.2207391