HIGH-RESOLUTION DEPTH AND REFLECTANCE MAP ESTIMATION IN UNDERWATER IMAGING

László Neumann[†], *Rafael Garcia*[§], *Jordi Ferrer*[§]

Institute of Informatics and Applications
University of Girona, Spain
Email: {rafa,jferrer}@eia.udg.es

Attila Neumann

Institute of Computer Graphics and Algorithms
Technical University of Vienna, Austria
E-mail: aneumann@cg.tuwien.ac.at

ABSTRACT

A novel technique to simultaneously determining the three-dimensional geometry and reflectance map of a scene in underwater imaging will be presented . Underwater light is affected by an exponential wavelenght-dependent absorption and undesired scattering effects. These special conditions make image and video processing a difficult task in participating media. In this paper it is shown how to use this disturbing absorption effect to find a new way to accurately estimate the scene depth-map and multispectral reflectivities. The process requires the acquisition of a multispectral image of the underwater scenario with the appropriate camera and lightsource arrangement. The algorithm has been tested on several synthetic images for different noise levels. The promising results show that the technique provides accurate and stable results.

1. INTRODUCTION

Depth-map reconstruction at pixel level is an important problem in computer vision [1]. Many techniques have appeared in the past to recover three-dimensional scene structure from two-dimensional images. A possible classification of these techniques broadly divides them into passive and active methodologies []. The most popular passive techniques are stereo [2] or "structure from motion", which use multiple views to resolve structure ambiguities inherent in a single image. The main bottleneck of these methods is solving the correspondence problem [3], which is even more difficult in the case of underwater imaging. Another widely-used passive technique is "depth from defocus" [4], where the relationship between focused and defocused images is computed. As in the case of the previous techniques, depth from defocus requires scenes with enough texture frequency. With respect to the second category, among active techniques, light stripe methods appear as the most popular range sensing techniques. However, to achieve depth maps with sufficient spatial resolution, a large number of closely spaced stripes has to be used []. This implies projecting one stripe at a time, and acquiring a sequence of images to be able to associate a unique stripe with a given pixel. In this case, a large number of images is necessary to obtain dense reconstructions. Evolution of this technique gave rise to assignation of gray-codes to the stripes, reducing the number of required images [5], or the use of color-coded stripe patterns [6], which reduce the number of required images to 1. The inherent weakness of this approach is the difficulty in segmentation of the stripes in a real scenario which reflects differently many wavelengths of light.

Recovering dense depth maps is even more difficult if we consider an underwater scene. A major obstacle to processing images of the ocean floor is related to the special transmission properties of the light in the underwater medium [7]. Light suffers two different processes in the aquatic environment [8]: (*i*) absorption, where light "disappears" from the image-forming process, and (*ii*) scattering, unwanted multiple inter-reflection in the medium. These transmission properties of the medium causes blurring of image features and limited visual range [9, 10].

In this paper we propose a novel method for estimating the depth-map of an underwater scene by exploiting the attenuation coefficients of the water. At the same time, the mathematical derivation will allow also to recover the reflectance properties of the scene. An iterative and a closed form solution to the estimation of a dense depth map are provided.

The remainder of this paper is organised as follows: first, in Section 2, we present the theoretical derivation of our strategy. Then, Section 5 illustrates the performance of the technique over a set of synthetic images, demonstrating the validity of our proposal. Finally, conclusions are given in Section 6.

This work has been partially funded by a grant from the Spanish Ministry of Education and Science (CTM2004-04205)

[†]L. Neumann is an ICREA researcher in the Graphics Group of Girona (GGG)

[§]R. Garcia and J. Ferrer are with the Computer Vision and Robotics Group

2. CAMERA AND LIGHTING ARRANGEMENT

In undewater imaging, and especially in deep ocean applications, the "longest" visual range is obtained by means of using only a limited interval of the spectrum in the bluish range [7]. Unfortunately, also the use of the bluish range is limited to a few tens of meters. Thereby, for multispectral images narrow-band spectral intervals have to be used. Despite the fact that this is an energy-consuming approach, it is better than the efficiency of laser techniques. The proposed method does not need time consuming scanning, as required by laser, but taking single shots with 2 or 3 lamps a large area can be processed.

Let us consider 3 spectral channels to illustrate our new method for depth-map reconstruction. These channels are not pure spectral colors but tight spectral windows according to band-pass filters of 5-20 nm. In a small spectral window we can assume a homogenous, nearly constant behavior of the most important optical parameters, like scattering and absorption.

For simplicity, let us call the 3 channels r, g and b, although they can represent *e.g.* 460-480, 440-460, 420-440 nm, three bluish intervals in deep ocean water. Consider we have several point light sources. In practice we can use flashes with lenses and optical filters to create a pinhole-like lightsource. The narrow-band filters could be either (*i*) built in a multispectral camera, (*ii*) placed in front the lens of a standard grayscale camera, or (*iii*) applied directly to the light sources. Let k be the index of lightsources (L_k). Two lamps are required with known spatial energy distribution, and known extrinsic (position and orientation) parameters. The use of a higher number of lights provides redundancy, thus, increasing accuracy.

Prior to the experiments, both the intrinsic and extrinsic parameters of the camera have to be modeled through calibration, together with the geometry of the lightsources with respect to the camera. In addition, the absorption factors of the 3 channels (a_R, a_G, a_B) have to be obtained also through calibration. Without loss of generality, we assume that the absorption factors are constant in the field of view. Other central assumption is that the reflectance models, *i.e.*, the Bidirectional Reflectance Distribution Functions (BRDFs), have the following properties: the ratio of reflected rgb radiance values is practically constant for the used light and view directions. That is true for a diffuse material, but also for a wider class of BRDFs where the shape of the distribution is similar at every wavelength, differing only by a linear multiplicative factor. Replacing the 3 "pseudo" rgb channels with "real" rgb colors a pseudo color image can be created, typically with low-saturated colors. Namely, the differences of 3 neighboring bluish narrow-bands in the reflectance curve are small. In order to obtain enhanced colorful images with increased contrast [11] or [12] can be applied.

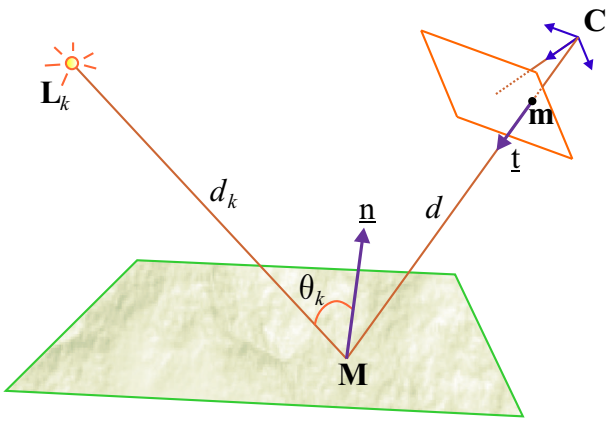


Fig. 1. Projection of a 3D world point **M** onto the image plane (**m**); d defines the unknown distance from scene point **M** the focal point of the camera, while d_k is the distance from lamp k to **M**, and Θ_k defines the angle between the incident light from lightsource k and the surface normal at point **M**.

3. THE BASIC EQUATIONS

In order to determine the depth map it would be enough to use 2 of the above mentioned "pseudo" rgb channels to derive the equations. Thereby, let us assume we work only with 2 wavelengths. For simplicity let us call the used λ_1 and λ_2 wavelengths r and g, independently of its real value.

We will demonstrate the method for diffuse materials. In this case the BRDF is a simple constant value [13]. Strictly, the BRDFs are:

$$BRDF_r = \frac{r}{\pi}[sr^{-1}], \qquad BRDF_g = \frac{r}{\pi}[sr^{-1}] \qquad (1)$$

where r and g are the 'reflectivity' values in the interval $[0, 1]$.

Consider a scene point **M**, which is projected into the image plane at coordinates $\mathbf{m} = (x_i, y_i)$ (see figure 1). For a given lamp k, every pixel **m** of the image defines the following equations:

$$R_k = r \cdot \cos\Theta_k \cdot P_{Rk} \cdot d_k^{-2} \cdot e^{(a_R \cdot d)} \cdot e^{(a_R \cdot d_k)} \qquad (2)$$

$$G_k = g \cdot \cos\Theta_k \cdot P_{Gk} \cdot d_k^{-2} \cdot e^{(a_G \cdot d)} \cdot e^{(a_G \cdot d_k)} \qquad (3)$$

where d is the unknown distance from scene point **M** to the focal point **C** of the camera; d_k is the distance from lamp k to **M**; (r, g) define the reflectivity of scene point **M**, according to eq. 1; Θ_k is the light incident angle; (P_{Rk}, P_{Gk}) defines the power of lamp k at wavelengths r and g; and

(a_R, a_G) are the absorption factors. Finally, (R_k, G_k) are the values measured by the camera at point **m**. It should be noted that we have ignored in eqs. (2–3) some multiplicative constants: the factor $\frac{1}{\pi}$ in eq. (1) and a factor $\frac{1}{4\pi}$ describing the irradiance at point **M**. Namely, the mentioned irradiance value is $P_k/(4\pi d_k^2)$. This approach is allowed here since we will only use in the following equations the ratio between eqs. (2–3).

The intrinsic parameters of the camera can be modeled through calibration. Thereby, view direction is known for every pixel. We aim to determine the camera-scene distance at pixel resolution (d), to obtain a dense depth-map of the scene.

The unknowns in eqs. (2–3) are d, d_k, r, g and Θ_k. Therefore, for every pixel, 5 unknowns and 2 equations are obtained for every lamp k. Let us consider the minimal required number of lamps, i.e. 2, and assume that both light-sources reach point **M**. An additional constraint arises from the dependence of d_k as a function of d, according to the spatial geometry of the camera-lightsources. Thus, we have a partly overdetermined problem, but determining Θ_k is impossible using the equations derived for a single point **M**, since we have no information about surface normal $\mathbf{n}$ (see Fig. 1).

In order to solve for the unknowns we use logarithmic scales. The difference of logarithm of eq. (2) and eq. (3) is the logarithm of the ratio of eq. (2) and eq. (3). First, we compute the difference between eqs. (2) and (3) for $k = 1$, as shown in eq. (4). Then, the difference between eqs. (3) and (2) for the second light source $(k = 2)$ (see eq. (5)).

$$\begin{aligned}\log R_1 - \log G_1 = \log r - \log g + \log P_{R1} - \\ - \log P_{G1} + d(a_R - a_G) + d_1(a_R - a_G)\end{aligned} \tag{4}$$

$$\begin{aligned}\log G_2 - \log R_2 = \log g - \log r + \log P_{G2} - \\ - \log P_{R2} + d(a_G - a_R) + d_2(a_G - a_R)\end{aligned} \tag{5}$$

Fortunately, proceeding in this way Θ_k disappears from the equations. The sum of eqs. (4-5) results in the expression of eq. (6).

$$\rho + P = d_1(a_R - a_G) + d_2(a_G - a_R) \tag{6}$$

where ρ is a constant which depends on the ratio of R_k and G_k; and the constant P depends on the power of the light sources.

$$\rho = \log R_1 - \log R_2 - \log G_1 + \log G_2 \tag{7}$$

$$P = -\log P_{R1} + \log P_{G1} - \log P_{G2} + \log P_{R2}, \tag{8}$$

thus, eqs. (7) and (8) can be expressed as:

$$\rho = \log\left(\frac{R_1 \cdot G_2}{R_2 \cdot G_1}\right) \tag{9}$$

$$P = \log\left(\frac{P_{R2} \cdot P_{G1}}{P_{R1} \cdot P_{G2}}\right) \tag{10}$$

With this approach, the terms containing r or g eliminate each other. Eq. (6) is a non-linear function of the unknowns d_1 and d_2, which can be solved by iteration. We can write eq. (6) as:

$$\rho + P = (a_R - a_G)(d_1 - d_2) \tag{11}$$

Writing d_1 and d_2 as a function of the depth d of image point **m** (see Fig. 1):

$$d_k(d) = \|\mathbf{M} - \mathbf{L}_k\| \tag{12}$$

Since $\mathbf{M} = \mathbf{C} + d \cdot \mathbf{t}$, being **M** the 3D coordinates of the imaged scene point, **C** the 3D position of the camera focal point and $\mathbf{t}$ the unit vector of the viewing direction, which can be computed from the intrinsic parameters of the camera. Therefore, eq. (12) can be written as:

$$d_k(d) = \|\mathbf{C} + d \cdot \mathbf{t} - \mathbf{L}_k\| \tag{13}$$

$$d_k(d) = \left[\sum_j (C_j + d \cdot t_j - L_{kj})^2\right]^{\frac{1}{2}} \tag{14}$$

with j being the x, y and z coordinates.

4. SOLVING THE EQUATIONS

Knowing the geometry of the system, if the distances from a scene point to the lamps has been computed, any two values d_1 and d_2 allow the computation of d for that given pixel. We can use as initial value to accelerate the computations the distance values of a neighboring pixel, or the values of a lower resolution picture.

$$d_k(d) = \left[\sum_j (A_{kj} + d \cdot t_j)^2\right]^{\frac{1}{2}} \tag{15}$$

with j being the x, y and z coordinates. Therefore, expanding eq. (15), the distance d_k from every pixel to lamp k can be expressed as a function of the form $d_k = \sqrt{d^2 + b_k \cdot d + c_k}$, where only d is unknown and b_k and c_k are constants. Retaking eq. (11), we can write:

$$\frac{\rho + P}{a_R - a_G} = d_1 - d_2 \tag{16}$$

Now, we can define the constant B as

$$B = \frac{\rho + P}{a_R - a_G}$$

and considering d_1 and d_2 as expressed above, eq. (17) is obtained:

$$B = \sqrt{d^2 + b_1 \cdot d + c_1} - \sqrt{d^2 + b_2 \cdot d + c_2} \tag{17}$$

Solving eq. (17), two possible solutions for d can be obtained

$$d = \frac{-2b_1B^2 - 2b_2B^2 + 2c_1b_1 - 2b_1c_2 + 2c_2b_2 - 2c_1b_2 \pm 4\psi}{2 \cdot (4B^2 - b_2^2 - b_1^2 + 2b_1b_2)} \tag{18}$$

with

$$\psi = \sqrt{B^2 \cdot (b_2^2c_1 + b_1^2c_2 + b_1B^2b_2 + B^4 + c_1^2 + \Delta)}$$

$$\Delta = c_2^2 - 2B^2c_1 - 2B^2c_2 - b_1c_1b_2 - b_1c_2b_2 - 2c_2c_1$$

4.1. Removing the absorption effect

Knowing the distances we have to determine the r, g and $\cos\Theta_k$ values using eqs. (2–3). To approximate the $\cos\Theta_k$ values, the known depth map and finite element gradient estimation for the local gradient of the underwater terrain can be used.

The final step is to compute the corrected r and g reflectivities using the absorption and distance values. Having more than 2 lamps, we can build –for every selected pair of lamps– the eqs. (2-11). This redundancy increases the accuracy or enables the computation of the surface points which may be in shadow for some of the lamps.

The average or median value of the reflectivities for the active lamps gives the desired reflectivity values, be them r and g or rgb pseudo color triplets.

For non-diffuse, but appropriate class of BRDFs the cosine term will be multiplied by a term which depends on the geometry. The basic assumption: the constancy of rgb ratio is deeply used in the derivation above.

5. EXPERIMENTAL RESULTS

6. CONCLUSIONS

A new strategy for dense depth estimation has been described. This scene reconstruction is not only valid for diffuse materials (with constant BRDF), but for a wide class of materials, having "non- color-changing" BRDFs.

The algorithm takes into account BRDF functions both to compute the depth map of the scene, and to perform a radiometric correction of the pixels of the scene.

The up-to-date high-speed CMOS camera sensors and high power flashes ensure a viable implementation of our technique.

7. REFERENCES

[1] L. Robert and R. Deriche, "Dense depth map reconstruction: A minimization and regularization approach which preserves discontinuities," in *ECCV '96: Proceedings of the 4th European Conference on Computer Vision-Volume I*. 1996, pp. 439–451, Springer-Verlag.

[2] O.D. Faugeras and G. Toscani, "The calibration problem for stereo," in *IEEE Conference on Computer Vision and Pattern Recognition*, 1986, pp. 15–20.

[3] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *IJCV*, vol. 47, no. 1, pp. 7–42, 2002.

[4] S.K. Nayar, M. Watanabe, and M. Noguchi, "Real-time focus range sensor," *PAMI*, vol. 18, no. 12, pp. 1186–1198, 1996.

[5] T. Sato, "Multispectral pattern projection range finder," in *SPIE Conference on Three-Dimensional Image Capture and Applications II, vol. 3640*, San Jose, California, 1999, p. 28–37.

[6] J. Batlle, E. Mouaddib, and J. Salvi, "A survey: Recent progress in coded structured light as a technique to solve the correspondence problem," *Pattern Recognition*, vol. 31, no. 7, pp. 963–982, 1998.

[7] J.S. Jaffe, "The domains of underwater visibility," in *SPIE Ocean Optics VIII*, 1986, pp. 287–293.

[8] R. Garcia, T. Nicosevici, and X. Cufi, "On the way to solve lighting problems in underwater imaging," in *IEEE OCEANS Conference*, Biloxi, Mississipi, 2002, pp. 1018–1024.

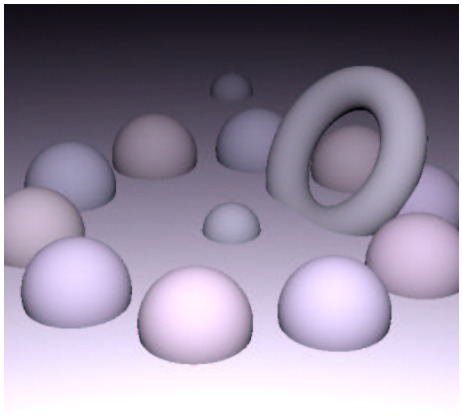
[9] H.R. Gordon, "Absorption and scattering estimates from irradiance measurements: Monte carlo simulations," *Limnology and Oceanography*, vol. 36, pp. 769–777, 1991.

[10] H. Loisel and D. Stramski, "Estimation of the inherent optical properties of natural waters from irradiance attenuation coefficient and reflectance in the presence of raman scattering," *Applied Optics*, vol. 39, pp. 3001–3011, 2000.

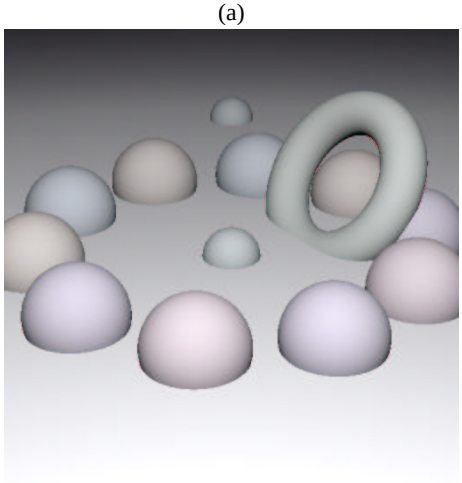
[11] E. Pichon, M. Niethammer, and G. Sapiro, "Color histogram equalization through mesh deformation," in *ICIP*, 2003, vol. 2, pp. 117–120.

[12] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Comput. Graph. Appl.*, vol. 21, no. 5, pp. 34–41, 2001.

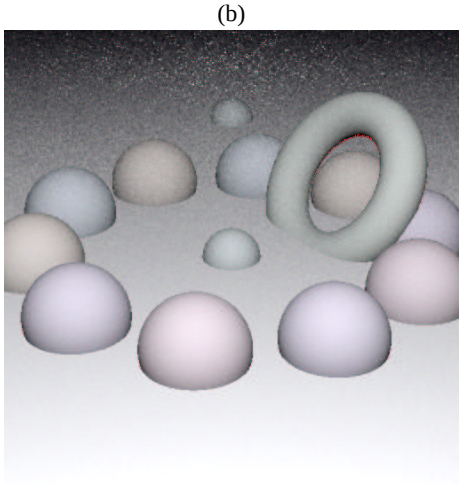
[13] L. Neumann and A. Neumann, "Radiosity and hybrid methods.," *ACM Trans. Graph.*, vol. 14, no. 3, pp. 233–265, 1995.



(a)



(b)



(c)

Fig. 2. Average percentage of correct correspondences after eliminating all the outliers using the proposed texture based methods $(1-6)$. (a) Original synthetic image. (b) Reconstructed image (continuous case). (c) Reconstructed image (discrete case with 16 bits per pixel).