# Relative Depth Estimation of Objects in Underwater Scenes

Ricardo Pérez-Alcocer, L. Abril Torres-Méndez and Ernesto Olguín-Díaz

Robotics and Advanced Manufacturing Group, CINVESTAV-Saltillo
Ramos Arizpe, Coahuila, MEXICO, 25900

**Abstract.** A method to estimate the relative depth of objects in an underwater environment is presented. The method uses as a depth cue the color of the objects without considering the degradation produced by the scattering and attenuation factors. To establish a relationship of the color intensity values, two images of the objects under different conditions are needed. In our case, one image is taken outside the water and the other underwater. For the preliminary experiments, we have generated underwater images from non-aquatic images with known range maps. We then apply our model to estimate the relative depth of the simulated underwater image. To evaluate the performance of our method, we first normalize both the resulting range map and ground truth range map and compared them.

## 1 Introduction

As underwater applications are becoming more feasible by current technology, the connection between depth perception and the actions to be taken in this kind of environments is fundamental. A great number of research works have focused on the development of underwater devices that autonomously can recollect, analyze and interpret information from this type of ecosystems. In this trend, the use of vision systems for aquatic robots ([1], [2], [3]) has become crucial for many underwater inspection and observation tasks. Unfortunately, the physical properties of underwater environments impose severe limitations on the quality of optical imaging due to the attenuation and scattering of light; and the living organisms in water. Furthermore, the time of the day and cloudiness of the sky have a great effect on the nature of the light available. Ambient light is practically nonexistent after few meters of depth due to the rapid attenuation of electromagnetic radiation underwater. The light undergoes scattering along the line of sight. The result is an image that is color depleted, blurry and out of focus. By 3m in depth there is almost no red light left from the sun. By 5m, orange light is gone, by 10m most yellow is also gone. By the time one reaches 25m only blue light remains [4]. Since many of the above factors are constantly changing, we cannot really know all the effects of water.

Some research works have proposed different ways to model the physical interaction of light in water [5], [4], [7]. However, if we want to have a more realistic

© J. A. Carrasco-Ochoa, J. F. Martínez-Trinidad and H. Sossa (Eds.)
Special Issue in Advances in Pattern Recognition
Research in Computing Science 44, 2009, pp. 29-36

model, we need to take into account other factors, such as all possible interactions of light sources with the objects; the analysis of the chemical components present in water depending on its source (e.g., lakes, rivers, ocean or even artificial pools). Clearly, we are in front of a very complex situation, as many different parameters (whose values are typically empirically assigned) need to be considered. In order to avoid this, it is common to assume aspects about the patterns present in the images to be analyzed and also about the environment on which they were captured. In this trend, some methods try to enhance visibility by decreasing the optical effects of scattering [6]. Another related application of such models is the color correction of images [7], [8]. In general, by understanding these models, we can be able to artificially generate images having an aquatic appearance from images taken on the surface. This is useful to perform analysis and recognition tasks on the patterns in the images and be able to estimate depth information, which is our main interest.

In this work, color information is used as a depth cue. First, we use the model proposed by Schechner and Karpel in [7] to artificially generate aquatic images. Once we have two images from the same scene, but under different conditions, we can estimate the relative depth information of each object in the underwater scene. To achieve this, we propose a depth estimation model based on the ideas presented in [6].

This paper is structured as follows. Section 2 presents the image formation model for aquatic images. In Section 3, we describe our method to estimate relative depth information from underwater images. Section 4 tests the proposed method on different generated underwater images. Finally, in Section 5 we give some conclusions and future directions.

## 2 The Image Formation Model for Underwater Images

The first step in the development of this work is to implement a simple model of underwater image formation. The understanding of certain properties observed in the aquatic images obtained from this model will help us to define an efficient method for depth estimation of underwater scenes.

In the formation of underwater images two main sources of light are present; the radiance of the objects in the environment and the ambient light. They both give as a result the characteristic tones existing in photos taken underwater.

In general, the signal is composed by two components denominated *direct transmission* and *forward scattering*. During the propagation of light rays coming from the objects to the camera, part of the energy is lost due to the scattering and the absortion. The component that represents these two factors is called direct transmission and is given by

$$D = L_{object} e^{-\beta z} \tag{1}$$

where $\beta$ is the attenuation coefficient and $L_{object}$ is the radiance of the object.

The component of forward scattering is similar to the direct transmission component. However, the forward scattering component generates blurred images given by the following convolution

$$F = D * g_z \tag{2}$$

where $D$ is given by Eq. (1) and $g_z$ is a point scattering function (PSF). The PSF is parametrized by the distance $z$. For this reason, objects located at far distances look blurrier than near objects. An example of a PSF is given by

$$g_z = (e^{-\gamma z} - e^{-\beta z})\mathcal{F}^{-1}\{G_z\} \quad \text{and} \quad \mathrm{G_z} = \mathrm{e}^{-\mathrm{Kz}\omega} \tag{3}$$

where $K > 0$ and $\gamma$ is a heuristic constant, $\mathcal{F}^{-1}$ is the Fourier inverse transform, and $\omega$ is the spatial frequency in the image plane.

Taking into account the direct transmission and the forward scattering, the signal can be defined as $S = D + F$. And defining $L_{object}^{effective}$ as

$$L_{object}^{effective} = L_{object} + L_{object} * g_z \tag{4}$$

we obtain that the signal is given by

$$S = e^{-\beta z} L_{object}^{effective} \tag{5}$$

On the other hand, the second component of the model is given by the backscattering that comes from the ambient illumination, from [7], the total backscattering is given by

$$B_\infty(1 - e^{-\beta z}) \tag{6}$$

Therefore, taking into account the signal and the backscattering, the total irradiance of the image is given by

$$I^{total} = S + B = e^{-\beta z} L_{object}^{effective} + B_\infty(1 - e^{-\beta z}) \tag{7}$$

Using the model described by Eq. (7) we can obtain simulated aquatic images by using an image taken on the surface and its associated range map. We have implemented this model and show some examples. We use the Middlebury[1] dataset in all our experiments. The original intensity images are depicted in the left column of Figure 1 and in their right are the associated range images. Some examples of the resulting aquatic images are shown in Figure 2 after applying the model explained above with the following parameters: $B_\infty = [48, 250, 255]$ and $\beta = [0.4, 0.0991, 0.0348]$. It can be noted that the images take a blue tone which is characteristic of underwater images. Also, it can be noted the effect that far objects look blurrier than near objects.

We will use this image formation model to simulate aquatic images and be able to validate our depth estimation model, which is described next.

[1] http://cat.middlebury.edu/stereo/data.html

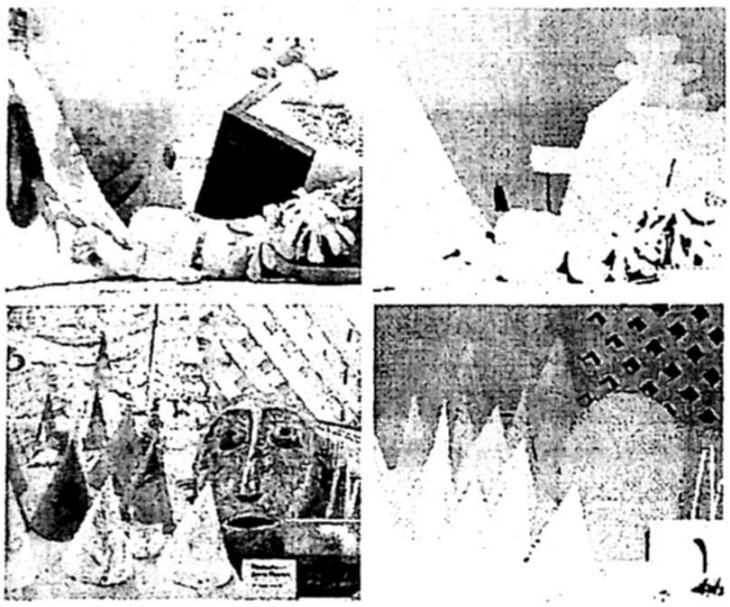

**Fig. 1.** Two examples of the original intensity images with their associate range images.

## 3 The Proposed Depth Estimation Model

Based on the comparative analysis carried out between the generated aquatic images and real aquatic images, some simplifications can be made to define our depth estimation model. The first simplification is that the forward scattering does not affect in a considerable manner. This is reasonable when the underwater illumination conditions also do not vary considerably in a given period of time. Thus, the Eq. 7 can be simplified for each of the RGB channels and is given by

$$\begin{aligned} E_R &= I_{OR}e^{-\beta_R d} + A_R(1 - e^{-\beta_R d}) \\ E_G &= I_{OG}e^{-\beta_G d} + A_G(1 - e^{-\beta_G d}) \\ E_B &= I_{OB}e^{-\beta_B d} + A_B(1 - e^{-\beta_B d}) \end{aligned} \tag{8}$$

where $I = [I_{OR}, I_{OG}, I_{OB}]$ represents the pixel values in the image taken outside the water, $\beta = [\beta_R, \beta_G, \beta_B]$ contains the degradation coefficients for each of the color channels. An important aspect to note here is that the three degradation coefficients are different, but they remain constant through all pixels in the image. On the other hand, $A = [A_R, A_G, A_B]$ contains the color values of the pixels that have an infinite distance to the camera. Therefore, considering the total radiance of the image in the red channel, we have

$$\begin{aligned} E_R &= A_R - (I_{OR} - A_R)e^{-\beta_R d} \\ E_R - A_R &= (I_{OR} - A_R)e^{-\beta_R d} \\ e^{-\beta_R d} &= \frac{E_R - A_R}{I_{OR} - A_R} \end{aligned} \tag{9}$$

**Fig. 2.** Simulated aquatic images after applying the image formation model to the images in Figure 1.

In similar way we can obtain the formulas for the green and blue channels:

$$e^{-\beta_G d} = \frac{E_G - A_G}{I_{OG} - A_G} \tag{10}$$

$$e^{-\beta_B d} = \frac{E_B - A_B}{I_{OB} - A_B} \tag{11}$$

By dividing Eq. 9 and Eq. 10 and, considering that the value of $\beta$ for each color channel is different, we have that

$$R_1 = e^{-(\beta_R - \beta_G)d} = \frac{(E_R - A_R)(I_{OG} - A_G)}{(I_{OR} - A_R)(E_G - A_G)} \tag{12}$$

And applying the natural logarithm to Eq. 12:

$$\ln(R_1) = -(\beta_R - \beta_G)d = \ln\left[\frac{(E_R - A_R)(I_{OG} - A_G)}{(I_{OR} - A_R)(E_G - A_G)}\right] \tag{13}$$

Now, if we calculate $\ln(R_1)$ for two pixel points $i, j$ of the image and compute their ratio, we can obtain the relative depth for the two points in the scene.

$$\frac{\ln(R_1^i)}{\ln(R_1^j)} = \frac{z_i}{z_j} \tag{14}$$

Therefore, the relative depths for all pixels in the aquatic image can be calculated by knowing the intensity values (in the three color channels) of the non-aquatic image. Given that the computation of Eq. 14 for one pixel can contain certain level of perturbation, we made the model more robust by calculating

$$\frac{\ln(R_1^i)}{\sum_{j=0}^{j=N} \ln(R_1^j)} = \frac{z_i}{\sum_{j=0}^{j=N} z_j} \tag{15}$$

where $\sum_{j=0}^{j=N} z_j$ is an arbitrary constant that represents the sum of the distances of all pixels in the image.

An important aspect of the depth estimation method is that it obtains three estimates for each pixel by calculating $R_1$ as the ratio between (9) and (10), between (9) and (11), and also between (10) and (11), thus we have that

$$R_1 = e^{-(\beta_R - \beta_B)d} = \frac{(E_R - A_R)(I_{OB} - A_B)}{(I_{OR} - A_R)(E_B - A_B)} \tag{16}$$

$$R_1 = e^{-(\beta_G - \beta_B)d} = \frac{(E_G - A_G)(I_{OB} - A_B)}{(I_{OG} - A_G)(E_B - A_B)} \tag{17}$$

## 4 Experimental Results

In this section, we present some of the experimental results obtained after using our depth estimation model. We carried out several tests on different images by assigning different values to the degradation coefficient and compare the results.

In Fig. 3 we show two different scenes with two different degradation coefficients each. The first and third images in the left column are the generated aquatic images after applying the image formation model with degradation coefficient $\beta = [0.4, 0.0991, 0.0348]$ to the original images in Fig. 1. These images correspond to Case I. Here, the maximum distance to an object in the image is 0.8m. The second and last images in the left column are the simulated images when using $\beta = [0.1, 0.05, 0.005]$ and corresponds to Case II. The middle column shows the estimated range maps for each of the image at the left, after applying our model. In order to evaluate the performance of our method, we compute the histograms of the residual errors of the estimated depth values and ground truth range after normalization. The corresponding error histograms are shown in the third column, the $x$-axis represents the residual errors while the $y$-axis are the number of pixels with same residual error. For Case I, the highest frequencies are accumulated in the smallest residual errors, thus indicating a good result. For Case II, it is more noticeable the color loss in the objects due to the high degradation coefficient used. This is reflected in the histogram as there exist a considerable amount of pixels whose residual errors are bigger than the previous case. These pixels are located mainly in the background regions of the scenes.

From the results presented several important aspects can be noted. In principle, we can see that in all cases, the estimated depth for images with lower level of degradation is very close to the original value. Second, we can observe that our model does not perform well when dealing with objects with texture in their surface. On the other hand, we see that the positional relationship between objects in the estimate scene, is quite similar to that of the original range image.

We can find some advantages by using our method. First, we do not require the typically stereo vision system to estimate depth. This is particularly difficult in underwater images as corresponding features are difficult to match. Our method only needs to know *a priori* the color appearance of objects outside the water. Then, together with its underwater version we can estimate the relative depth information. While this restriction can be seen as a disadvantage, it is important to mention that it actually depends on the final task to accomplish.

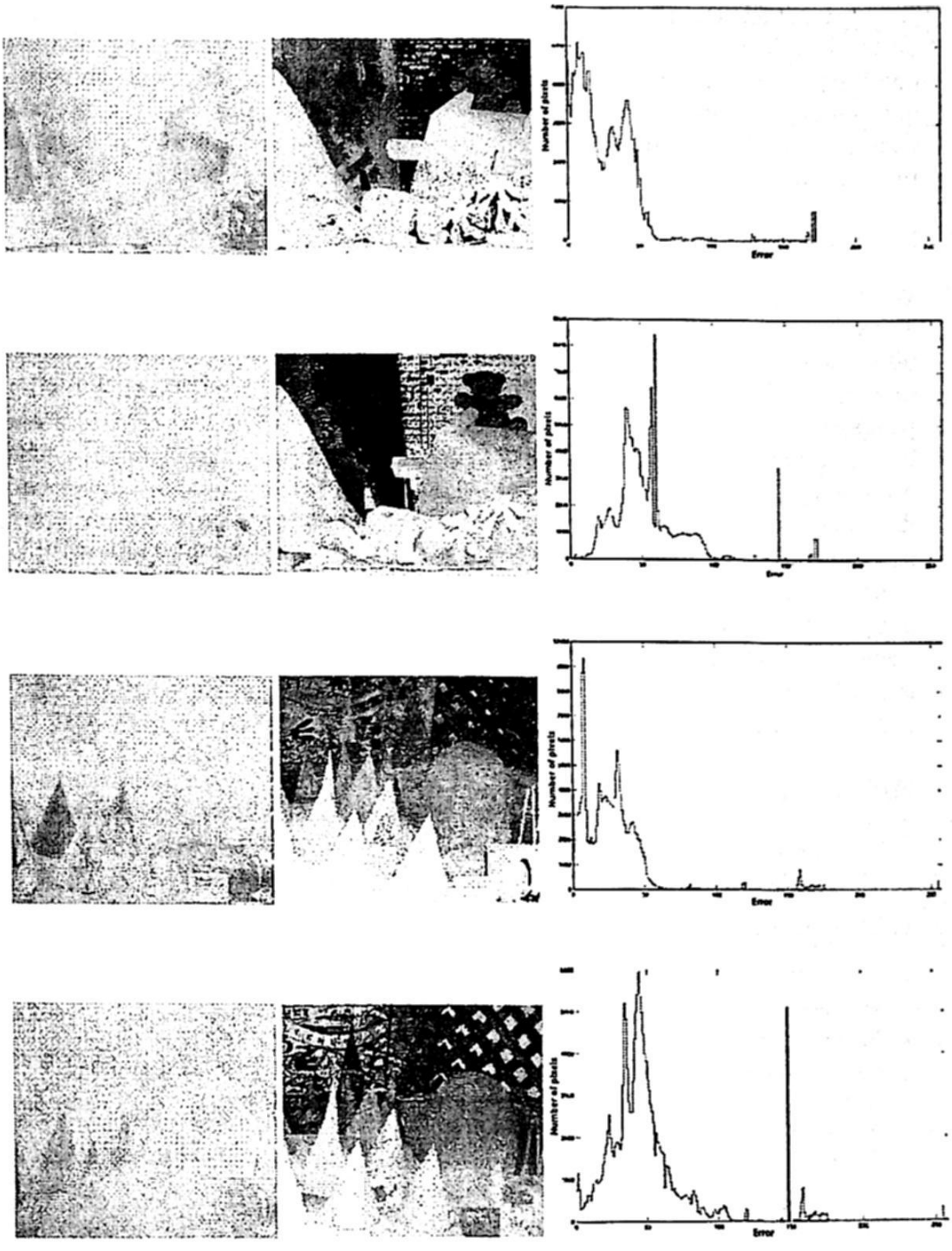

**Fig. 3.** Experimental results for the depth estimation model. The input aquatic images are in the first column. The resulting depth maps in the second column and in the third, the error histograms when compared to ground truth range maps (shown in Fig. 1.)

In our case, previously known objects (landmarks) will be located in the underwater environment to help a robot to autonomously navigate to a given goal. Another advantage of the model is that having different values for the degradation coefficients at each RGB channel, we can estimate three range maps for one image and eliminate an important amount of the errors present when using only one estimation. These calculations are simple and therefore the execution time is short. This is particular important for real-time applications, especially in underwater environments, where the camera cannot be fixed.

## 5 Conclusions

We present a method to estimate the relative depth of objects in aquatic images. To evaluate the performance of our method different tests were carried out under different scenarios. On one hand, an image formation model to simulate aquatic images was implemented that reflects the three main factors affecting the light propagation in this type of ecosystem. The photometric characteristics present in the simulated aquatic images, after applying this model, are similar to those observed in images taken underwater. On the other hand, we have developed a depth estimation model based on the image formation model with some simplifications, such as not taking into account the forward scattering present in real aquatic images and considering the color intensity values as a depth cue. In general, we obtain a good estimation of the relative depth of objects in underwater scene, even when testing with images that include this effect. The results are specially good in textureless objects, with few variation in the color intensities. Future work involves using real underwater images to evaluate our model.

## Acknowledgements

We would like to thank to Conacyt for funding this research work.

## References

1. Boult, T.: DOVE: Dolphin omni-directional video equipment. In Proc.Int. Conf. Robotics and Automation, pp. 214-220 (2000).
2. Foresti, G.L.: Visual inspection of sea bottom structures by an autonomous underwater vehicle. IEEE Trans. Syst. Man and Cyber., Part B, 31:691795 (2001).
3. C. Georgiades *et al.*: AQUA: an aquatic walking robot. Proc. of IROS, Vol. 3, pp. 3525-3531 (2004).
4. Jaffe, J. S.: Computer modeling and the design of optimal underwater imaging systems. IEEE J. Oceanic Engin. 15, pp. 101-111 (1990).
5. McGlamery, B. L.: A computer model for underwater camera system.Proc. SPIE 208, pp. 221-231 (1979).
6. Narasimhan, Srinivasa G., Nayar, Shree K.: Vision and Atmosphere. International Journal of Computer Vision, Vol. 48, No. 3, pp. 233 - 254 (2002).
7. Schechner, Youv Y. and Karpel, Nir: Clear Underwater Vision. Proc. of the Computer Vision and Pattern Recognition, Vol 1, pp. 536 - 543 (2004).
8. Torres-Méndez, L.A., Dudek, G.: Color Correction of Underwater Images for Aquatic Robot Inspection. EMMCVPR, pp. 60-73 (2005).