

A Statistical Learning-Based Method for Color Correction of Underwater Images

Luz A. Torres-Méndez and Gregory Dudek

Centre for Intelligent Machines, McGill University,
Montreal, Quebec, H3A 2A7, CA,
latorres,dudek@cim.mcgill.ca,

WWW home page: <http://www.cim.mcgill.ca/~latorres/research.html>

Abstract. This paper addresses the problem of color correction of underwater images using statistical priors. Underwater images present a challenge when trying to correct the blue-green monochrome shift to bring out the color visible under full spectrum illumination in a transparent medium. We propose a learning-based Markov Random Field (MRF) model based on training from examples. Training images are small patches of color depleted and color images. The most probable color assignment to each pixel in the given color depleted image is inferred by using a non-parametric sampling procedure. Experimental results on a variety of underwater scenes demonstrate the feasibility of our method.

1 Introduction

Image restoration in general, involves the correction of several types of degradation in an image. The restored image must be more suitable than the original image for a specific application. Traditionally, the most common sources of degradation are due to imperfections of the sensors, or in transmission. We are interested in restoring the color of underwater images. High quality image data is desirable for many underwater inspection and observation tasks. Particularly, vision systems for aquatic robots [2, 4, 7] must cope with a host of geometrical distortions: color distortions, dynamic lighting conditions and suspended particles (known as 'marine snow') that are due to inherent physical properties of the marine environment. All these distortions cause poor visibility and hinder computer vision tasks, e.g., those based on stereo triangulation or on structure from motion.

Underwater vision is plagued by poor visibility [10, 9] (even in the cleanest water). Additional factors are the ambient light, and frequency-dependent scattering and absorption, both between the camera and the environment, and also between the light source (the sun) and the local environment (i.e. this varies with both depth and local water conditions). The light undergoes scattering along the line of sight. The result is an image that is color depleted (typically appearing bluish), blurry and out of focus. In this paper, we focus on the specific problem of restoring/enhancing the color of underwater images. The term *color* refers to

the red, green and blue values (often called the color channels) for each pixel in an image.

Prominent blue color of clear ocean water, apart from sky reflection, is due to selective absorption by water molecules. The quality of the water determines its filtering properties. The greater the dissolved and suspended matter, the greener (or browner) the water becomes. The time of day and cloudiness of the sky also have a great effect on the nature of the light available. Another factor is depth, once at sufficient depth, no amount of filtration can effectively restore color loss. Due to the nature of underwater optics, red light diminishes when the depth increases, thus producing blue to grey like images. 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 [3]. Since many (if not all) of the above factors are constantly changing, we cannot really know all the effects of water.

Color recovery is not a simple linear transform since it depends on distance and it is also affected by quantization and even light source variations. We propose a learning based Markov Random Field model for color correction based on training from examples. This allows the system to adapt the algorithm to the current environmental conditions and also to the task requirements. As proposed in[5], our approach is based on learning the statistics from training image pairs. Specifically, our MRF model learns the relationships between each of the color training images with its corresponding color depleted image. Training images are small patches of regions of interest that capture the maximum of the intensity variations from the image to be restored. In the process it is important not to lose resolution or details which will create even a worse problem. To our knowledge this is the first formulation of using MRFs for the context of color correction of underwater images.

This paper is structured as follows. Section 2 briefly consider some of the related prior work. Section 3 describes our Markov Random Field model for color correction. Section 4 tests the proposed algorithm on two different scenarios with several types of experimental data each. Finally, in Section 5, we give some conclusions and future directions.

2 Related Work

There are numerous image retouching programs available that have easy-to-use, semi-automated image enhancement features. But since they are directed at land-based photography, these features do not always work with underwater images. Learning to manipulate the colors in underwater images with computer editing programs requires patience. Automated methods are essential, specially for real-time applications (such as aquatic inspection). Most prior work on image enhancement tend to approximate the lighting and color processes by idealized mathematical models. Such approaches are often elegant, but may not be well suited to the particular phenomena in any specific real environment. Color restoration is an ill-posed problem since there is not enough information in the

poor colored image alone to determine the original image without ambiguity. In their work, Ahlen *et al.* [1] estimate a diffuse attenuation coefficient for three wavelengths using known reflectance values of a reference gray target that is present on all tested images. To calculate new intensity values they use Beer’s Law, where the depth parameter is derived from images that are taken at different depths. Additional parameters needed are the image enhancements functions built into the camera. In general, their results are good, but the method’s efficiency depends highly on the previously noted parameters. In [11] a method that eliminates the backscatter effect and improves the acquisition of underwater images with very good results is presented. Their method combines a mathematical formula with a physical filter normally used for land photography. Although the method does not perform color correction, the clarity achieved on the underwater images may allow for color correction.

3 Our MRF Approach for Color Correction

The solution of the color correction problem can be defined as the minimum of an energy function. The first idea on which our approach is based, is that an image can be modeled as a sample function of a stochastic process based on the Gibbs distribution, that is, as a Markov Random Field (MRF) [6]. We consider the color correction a task of assigning a color value to each pixel of the input image that best describes its surrounding structure using the training image pairs. The MRF model has the ability to capture the characteristics between the training sets and then used them to learn a marginal probability distribution that is to be used on the input images. This model uses multi-scale representations of the color corrected and color depleted (bluish) images to construct a probabilistic algorithm that improves the color of underwater images. The power of our technique is evident in that only a small set of training patches is required to color correct representative examples of color depleted underwater images, even when the image contains literally no color information. Each pair of the training set is composed by a color-corrected image patch with its corresponding color-depleted image patch. Statistical relationships are learned directly from the training data, without having to consider any lighting conditions of specific nature, location or environment type that would be inappropriate to a particular underwater scene.

3.1 The MRF Model

Denote the input color depleted image by $B = \{b_i\}, i = 1, \dots, N$, where $N \in \mathbf{Z}$ is the total number of pixels in the image and b_i is a triplet containing the *RGB* channels of pixel location i . We wish to estimate the color-corrected image $C = \{c_i\}, i = 1, \dots, N$, where c_i replaces the value of pixel b_i with a color value. The images in the training set are pairs of color depleted and color corrected regions of interest that capture the maximum of the intensity variations of the image to be restored. They are defined in a similar manner, let $B_p^t = \{b_i^t\}, i = 1, \dots, N_p^t$ be the color depleted training image p and $C_p^t = \{c_i^t\}, i = 1, \dots, N_p^t$,

the corresponding color corrected image. $N_p^t \in \mathbf{Z}$ represents the total number of pixels in the training image p .

The color correction problem can be posed as a labeling problem. A labeling is specified in terms of a set of *sites* and a set of *labels*. Sites often represent image pixels or regions in the Euclidean space. Let \mathcal{S} index a discrete set of N sites $\mathcal{S} = \{s_1, s_2, \dots, s_N\}$, and \mathcal{L} be the set of corresponding labels $\mathcal{L} = \{l_1, l_2, \dots, l_N\}$, where each l_i takes a color value. The inter-relationship between sites define the *neighborhood system* $\mathcal{N} = \{N_s \mid \forall s \in \mathcal{S}\}$, meaning any collection of subsets of \mathcal{S} for which 1) $s \notin N_s$, and 2) $s \in N_r \iff r \in N_s$. N_s is the set of *neighbors* of s and the pair $\{\mathcal{S}, \mathcal{N}\}$ is a graph in the usual way. Each site s_i is associated with a random variable F_i . Formally, let $F = \{F_1, \dots, F_N\}$ be a random field defined on \mathcal{S} , in which a random variable F_i takes a value f_i in \mathcal{L} . A realization $f = f_1, \dots, f_N$, is called a *configuration* of F , corresponding to a realization of the field. The random variables F defined on \mathcal{S} are related to one another via the neighborhood system \mathcal{N} .

F is said to be an MRF on \mathcal{S} with respect to \mathcal{N} if and only if the following two conditions are satisfied [8]:

$P(f) > 0$ (positivity), and

$$P(f_i \mid f_{\mathcal{S}-\{i\}}) = P(f_i \mid f_{N_i}) \quad (\text{Markovianity}).$$

where $\mathcal{S} - \{i\}$ is the set difference, $f_{\mathcal{S}-\{i\}}$ denotes the set of labels at the sites in $\mathcal{S} - \{i\}$ and $f_{N_i} = \{f_{i'} \mid i' \in N_i\}$ stands for the set of labels at the sites neighboring i . The Markovianity condition depicts the local characteristics of F .

The choice of N together with the conditional probability distribution of $P(f_i \mid f_{\mathcal{S}-\{i\}})$, provides a powerful mechanism for modeling spatial continuity and other scene features. On one hand, we choose to model a neighborhood N_i as a square mask of size $n \times n$ centered at pixel location i . On the other hand, calculating the conditional probabilities in an explicit form to infer the exact MAP in MRF models is intractable. We cannot efficiently represent or determine all the possible combinations between pixels with its associated neighborhoods. Therefore, we avoid the usual computational expense of sampling from a probability distribution (Gibbs sampling, for example) and color correct a pixel value b_i with neighborhood N_i by using a non-parametric sampling strategy that is easy to implement, generates good results and is fast to execute.

To compute the MAP estimate for a color value, one first need to construct an approximation to the conditional probability distribution $P(f_i \mid f_{N_i})$ and then sample from it. For each new color depleted value $c_i \in \mathbf{C}$ to estimate, the samples (\mathcal{A}), which correspond to the set of small image patches in the training set, are queried and the distribution of C_i is constructed as a histogram of all possible values that occurred in the samples. \mathcal{A} is a subset of the real infinite set of all color images, denoted by \mathcal{N}_{real} .

Based on our MRF model, we assume that the color-depleted value c_i depends only of its immediate neighbors, i.e. of N_i . If we define a set

$$\Gamma(c_i) = \{N^* \subset \mathcal{N}_{real} : \|N_i - N^*\| = 0\} \quad (1)$$

containing all occurrences of N_i in \mathcal{N}_{real} , then the conditional probability distribution of c_i can be estimated with a histogram of all center color values in $\Gamma(c_i)$. Unfortunately, we are only given \mathcal{A} , i.e., a finite sample from \mathcal{N}_{real} , which means there might not be any neighborhood containing exactly the same characteristics in intensity and range as N_i in \mathcal{A} . Thus, we must use a heuristic which let us find a plausible $\Gamma'(c_i) \approx \Gamma(c_i)$ to sample from.

In the non-parametric approach, a color value c_p with neighborhood N_p , is synthesized by first selecting the most similar neighborhood (N_{best}) to N_p , i.e., the closest match to the region being filled in,

$$N_{best} = \underset{A_q \in \mathcal{A}}{\operatorname{argmin}} \|N_p - A_q\|, \quad (2)$$

Second, the k neighborhoods A_q in \mathcal{A} that are similar (up to a given threshold ϵ) to this N_{best} are included in $\Gamma'(c_p)$, as follows

$$\|N_p - A_q\| < (1 + \epsilon) \|N_p - N_{best}\| \quad (3)$$

The similarity measure $\|\cdot\|$ between two generic neighborhoods N_a and N_b is defined as the weighted sum of squared differences (WSSD) over the two neighborhoods. The "weighted" part refers to applying a 2-D Gaussian kernel to each neighborhood, such that those pixels near the center are given more weight than those at the edge of the window. We can now construct a histogram from the color values c_p in the center of each neighborhood in $\Gamma'(c_p)$, and randomly sample from it. c_q is then used to specify c_p . For each successive augmented voxel this approximates the maximum *a posteriori* estimate.

Measuring the dissimilarity between image neighborhoods is crucial for obtaining quality results, especially when there is a prominent color (blue or green) as in underwater images. Color information can be specified, created and visualized by different color spaces (see [12] for more information about color spaces). For example, the *RGB* color space, can be visualized as a cube with red, green and blue axes. Color distance is a metric of proximity between colors (e.g. Euclidean distance) measured in a color space. However, color distance does not necessarily correlate with *perceived* color similarity. Different applications have different needs which can be handled better using different color spaces. For our needs it is important to be able to measure differences between colors in a way that matches perceptual similarity as good as possible. This task is simplified by the use of *perceptually uniform* color spaces. A color space is perceptually uniform if a small change of a color will produce the same change in perception anywhere in the color space. Neither RGB, HLS or CIE XYZ is perceptually uniform. We use the CIE *Lab* space, which was designed such that the equal distances in the color space represent equal perceived differences in appearance.

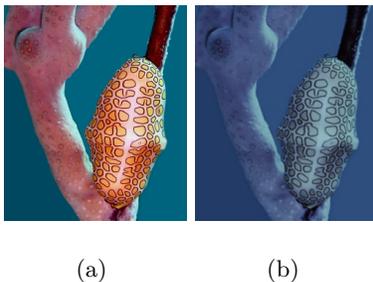


Fig. 1. (a) The ground truth (color) image. (b) The simulated color depleted image (this is the test image to be color corrected by our algorithm).

4 Experimental results

We test the proposed approach in two different scenarios. In the first scenario, we use color underwater images available on the web ¹ as our ground truth data. These images were taken with a professional camera. The second scenario, involves the acquisition of underwater video by our aquatic robot. Sections 4.1 and 4.2 describe these scenarios with the experimental results.

4.1 Scenario 1

In order to simulate the effects of water, an attenuation filter were applied to each of the color underwater image. Figure 1a shows the ground truth (color) image and Figure 1b, the simulated (color depleted) image after applying the attenuation filter. The images in the training set correspond to small image regions extracted from the ground truth image and the color depleted image (see Figure 2). These images correspond to regions of interest in terms of the variations in pixel color values, thus the intention is that they capture the intrinsic statistical dependencies between the color depleted and ground truth pixel values. The size of the neighborhoods in all experiments were 5×5 pixels, and the number of possible candidates k , was fixed to be 10. Figure 3a shows the training image

¹ <http://www.pbase.com/imagine> (used with the kindly permission of Ellen Muller.)

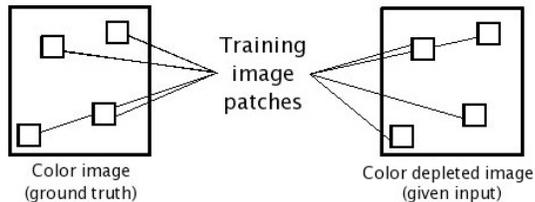


Fig. 2. Diagram showing how the training image pairs are acquired for the Scenario 1.

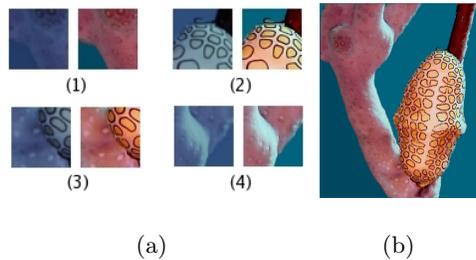


Fig. 3. (a) The training image patches. (b) The color corrected image.

patches from where our algorithm learns the compatibility functions and Figure 3b shows the resulted image after running our learning-based method. The color-corrected image *looks* good, the discontinuities and edges are preserved since our method assign colors pixel by pixel, thus avoiding over-smoothing. Also, there are no sudden changes in color which are typically both unrealistic and perceptually unappealing. To evaluate the performance of our algorithm, we compute the mean absolute residual (MAR) error between the ground truth and the color corrected images. For this case, the MAR error is 4.4. Note that while our objective is perceptual similarity, this is difficult to evaluate and we use this objective measure to obtain quantitative performance data. For comparison purposes, we calculate the MAR error between the input (color depleted) image and the ground truth image, this is 28.85.

Using the same input image (Figure 3b), we now show how the final result varies depending on the training data. In Figure 4, 4 examples when using different training pairs are shown. For example, Figure 4a shows a color-corrected image when using training pairs (1) and (3) (see Figure 3a). The MAR errors are 5.43, 6.7, 8.54 and 25.8, respectively. It can be seen that the resulting images are limited to the statistical dependencies captured by the training pairs.

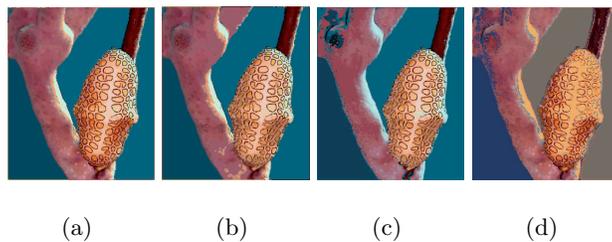


Fig. 4. Color correction results using different training sets. The input image is shown in Figure 1b. The training pairs (labeled) are shown in Figure 3a. Results using training pair (a) (1), (2) and (3); (b) (1) and (2); (c) , and (d) (1).

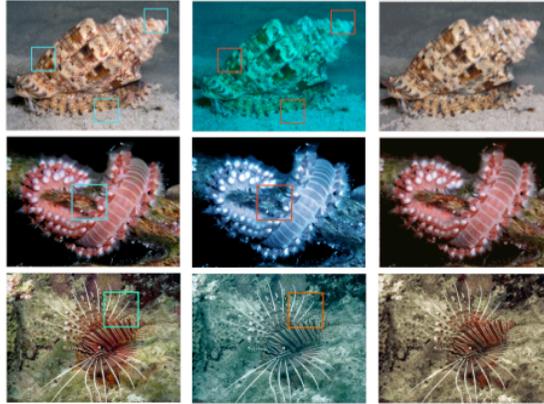


Fig. 5. The training pairs are indicated by the squares in the original and input images.

Three more examples of underwater scenes are shown in Figure 5. Each row shows from left to right, the ground truth color image, the input bluish image and the color corrected image after running our algorithm. The training image regions are shown by squares in the corresponding color and bluish images. In general the results look very good. For the last two examples, the size of the image patches in the training set is very small and enough to capture all the statistical dependencies between bluish and color information, as a result, the number of total comparisons in our algorithm is reduced and speed is achieved. The average computation time for an input image of 300×400 pixels with small number of training pairs (4 or less) of size 50×50 is 40 seconds on generic PC's.

4.2 Scenario 2

The following scenario is of special interest as it can be applied to color correct underwater images in real-time with no user intervention. Our application is specifically for aquatic robot inspection. As our aquatic robot [7] swims through the ocean, it takes video images. In order to be able to correct the color of the images, training data from the environment that the robot is currently seeing needs to be gathered. How can better images be acquired? As light is absorbed selectively by water, not only does it get darker as you go deeper, but there is a marked shift in the light source color. In addition, there are non-uniformities in the source amplitude. Therefore, the aquatic robot needs to bring its own source of white light on it. However, due to power consumption, the light cannot be left turned on. Therefore, only at certain time intervals, the robot stops, turns its light on and take an image. These images are certainly much better, in terms of color and clarity, than the previous ones, and they can be used to train our algorithm to color correct neighboring frames (under the assumption that neighboring frames are similar). Figure 6 shows this scenario, here frame t_3 represents the image pair to be used to train our model for color correction.

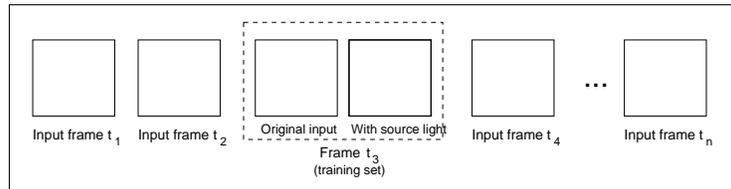


Fig. 6. The scenario 2.

Now we show an example. Figures 7a,b show the training image pair captured at time t . The robot moves around and then at time $t + \delta$ takes an image (Figure 7c), which is input to our algorithm. The resulting color-corrected image is shown in Figure 7d. Since we do not have ground truth data for this scenario, we cannot measure the performance of our algorithm, however it can be seen that the resulting image looks visually good.

Additional results on images, using the same training pair of previous example, are shown next. Figures 8a,c show the color depleted frames and Figures 8b,d, the color corrected images. Our algorithm performs very well in both examples. Note that the size of the training image pair is considerably bigger than those considered in previous section. The computation time is increased for the searching of best matching candidates in the training data. We have implemented a kd -tree structure on the training data, thus significantly decreasing computational burden, taking on average one minute for an image of 400×300 pixels. Details on this implementation are skip due to space limitations.

5 Concluding remarks

Color restoration and image enhancement are ubiquitous problems. In particular, underwater images contain distortions that arise from multiple factors making them difficult to correct using simple methods. In this paper, we show how to formulate color recovery based on using statistical learning constraints. This approach's novelty lies in using a pair of images to constrain the reconstruction. There are some factors that influence the quality of the results, such as the ade-



Fig. 7. (a)-(b) The training image pair captured at frame t . (c) Image taken at frame $t + \delta$ and input to our algorithm. (d) The color corrected image.

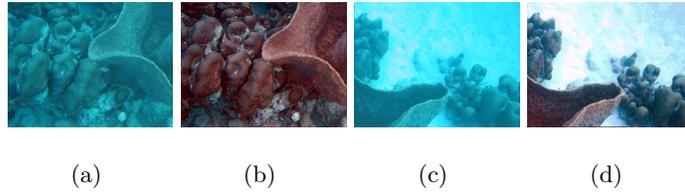


Fig. 8. (a)-(b) The training image pair captured at frame t . (c) Image taken at frame $t + \delta$ and input to our algorithm. (d) The color corrected image.

quate amount of reliable information as an input and the statistical consistency of the images in the training set. In some cases, ambiguities on local information is obtained, this is due to the fact that both the color corrected and the color depleted classifiers look only at a small image patches. Therefore, propagating information between regions can be used to resolve the ambiguity. More specifically, the marginal probability of the MRF model can be calculated by using belief propagation (BP) [13]. Results in this direction will be reported elsewhere.

References

1. J. Åhlén and D. Sundgren. Bottom reflectance influence on a color correction algorithm for underwater images. In *Proc. SCIA*, pages 922–926, 2003.
2. T. Boulton. DOVE: Dolphin omni-directional video equipment. In *Proc.Int. Conf. Robotics and Automation*, pages 214–220, 2000.
3. J. Dera. *Marine Physics*. Elsevier, 1992.
4. G.L. Foresti. Visual inspection of sea bottom structures by an autonomous underwater vehicle. *IEEE Trans. Syst. Man and Cyber., Part B*, 31:691–795, 2001.
5. W.T. Freeman, E.C. Pasztor, and O.T. Carmichael. Learning low-level vision. *International Journal of Computer Vision*, 20(1):25–47, 2000.
6. S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. on PAMI*, 6:721–741, 1984.
7. C. Georgiades, A. German, A. Hogue, H. Liu, C. Prahacs, A. Ripsman, R. Sim, L. A. Torres-Méndez, P. Zhang, M. Buehler, G. Dudek, M. Jenkin, and E. Milios. AQUA: an aquatic walking robot. In *Proc. of IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*, volume 3, pages 3525–3531, Sendai, Japan, 2004.
8. J.M. Hammersley and P. Clifford. Markov field on finite graphs and lattices. In *Unpublished*, 1971.
9. S. Harsdorf, R. Reuter, and S. Tonebon. Contrast-enhanced optical imaging of submersible targets. In *Proc. SPIE*, volume 3821, pages 378–383, 1999.
10. J.S. Jaffe. Computer modeling and the design of optimal underwater imaging systems. *IEEE J. Oceanic Engineering*, 15:101–111, 1990.
11. Y.Y. Schechner and N. Karpel. Clear underwater vision. In *Proc. of Intl. Conference of Vision and Pattern Recognition*, volume 1, pages 536–543, 2004.
12. G. Wyszecki and W.S. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulae*. Wiley, NewYork, 1982.
13. J. Yedidia, W. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. Technical report, Mitsubishi Elect. Research Labs., Inc., 2001.