A Deep Learning Approach for Underwater Image Enhancement

Javier Perez[(✉)], Aleks C. Attanasio, Nataliya Nechyporenko,
and Pedro J. Sanz

Department of Computer Science and Engineering,
Jaume I University, Castellón de la Plana, Spain
{japerez,al366204,al365957,sanzp}@uji.es

Abstract. Image processing in underwater robotics is one of the most challenging problems in autonomous underwater robotics due to light transmission in water. Although image restoration techniques are able to correctly remove the haze in a degraded image they need many images from the same location making impossible to use it in a real time system. Taking into account the great results of deep learning techniques in other image processing problems such as colorizing images or detecting objects a deep learning solution is proposed. A convolutional neural network is trained with image restoration techniques to dehaze single images outperforming other image enhancement techniques. The proposed approach is able to produce image restoration quality images with a single image as input. The neural network is validated using images from different locations and characteristics to prove the generalization capabilities.

Keywords: Underwater robotics · Deep learning · Image dehazing

1 Introduction

One of the most challenging problems in underwater robotics is the processing of underwater images. Besides the well known problems to automatically interpret an image in order to interact with the environment, underwater robotics needs to deal with additional problems caused by the degradation of the image due to the light transmission in water.

A correct interpretation of the camera input is crucial to build autonomous robots capable to move and interact in an unknown environment. In the case of underwater robotics, there are many applications related to the underwater industry and, unfortunately, maritime disasters such as shipwrecks, leaks on offshore or aircraft accidents. These interventions are usually performed by Remote Operated Vehicles (ROVs) controlled by expert pilots through an umbilical communication cable. Nevertheless, in the last few years, a more autonomous architecture has been developed: Intervention Autonomous Underwater Vehicles (IAUV) [4]. This architecture has many advantages such as the absence of delay between commands and vehicle reaction.

Usually, the first step in this kind of systems consists in processing the input of the cameras to be able to localize the system, safely navigate and identify the targets of interest. Due to the nature of light transmission in the underwater environment, described in [15], images suffer from different degradation effects such as absorption, scattering, marine snow or vignetting. These effects make interpreting the scene a really challenging problem.

Absorption reduces the amount of light as the robot goes deeper or further from the camera, colors drop off one by one depending on their wavelengths. This effect is the cause of the bluish color of underwater images as this wavelength is the least attenuated in the medium. The scattering effect changes the direction of the light to the camera generating a characteristic veil that superimposes itself on the image and hides the scene blurring the objects. Besides this effects, a common problem is the presence of small floating particles known as marine snow, that also increase the amount of scattered light. Finally, vignetting is a light fade-out in terms of intensity in the corners of the image caused by the geometry of the lens and sometimes by the lens housing.

For this reasons a preprocessing step is needed in order to restore the original colors and enhance the image for further processing. This can be addressed from two points of view. Image restoration that aims to recover a degraded image using a model of the degradation and of the acquired image: it is essentially an inverse problem. The second option, image enhancement, consists in using qualitative subjective criteria to produce a more visually pleasing image. Both methods have their own advantages and drawbacks, but the main difference is image restoration produces more realistic results but requires to estimate or measure several parameters thus it is difficult to use in a real time system.

In this work a hybrid solution is proposed: using a deep learning architecture to learn an image enhancement function from image restoration techniques. A dataset of pairs of raw and restored images is used to train a convolutional network, thus it is able to produce restored images from degraded inputs. The results are compared with other image enhancement methods using the image restoration as groundtruth of the system.

The paper is organized as follows. In the next section a review of state of the art techniques for image dehazing is presented. Section 3 describes the deep learning method. The experiments and results of the proposed approach are showed in Sect. 4. Finally, in Sect. 5 conclusions and further work are given.

2   State of the Art

Restoring degraded underwater images requires modelling and estimating many parameters such as water absorption, scattering and distance to objects (depthmap). This kind of inputs are difficult to estimate from a single image. For this reason, a large set of images from the same location or a combination of different sensors are typically used for this purpose. There is a large amount of work on restoring underwater images, [20] offers a detailed review.

The work in [1] uses a whole dataset of images and depthmaps from the same intervention to accurately estimate the water, light and camera parameters in

order to restore the colors of the image. The main drawback of this approach is, it requires a medium sized dataset of images and depthmaps of the same area, which may not be available making impossible to use it in real time applications.

Similarly, the authors in [21] propose a method using a depthmap and use it with a single image to estimate the rest of the parameters to restore the image. However, this method depends on a dense depthmap that may not be available when the environment is not textured enough.

Other works like [25], use specific hardware that dynamically mixes the illumination of an object in a distance dependent way by using a controllable multicolor light source in order to compensate color loss. This approach achieves a great color correction, but the main problem is the need of a specific hardware to solve the problem. Similarly, in order to deal with scattering some methods use specific hardware such as structured illumination [16] or polarizers [24].

In the context of single image dehazing there are a big family of algorithms that use the dark channel prior as [10]. Dark Prior techniques are based on the observation that in most of the non-background patches of outdoor haze-free images, at least one color channel has some pixels whose intensity is very low and close to zero. This has been proved to work in most outdoor air images and has also been adapted to underwater environments in [3] or [6]. The main disadvantage of this method is that it is based on a statistical observation that may not be valid for some cases.

In terms of image enhancement, a histogram equalization is typically used as described in [7]. This techniques analyse the histogram and transform it to accomplish a determined distribution that produces visually pleasing images. The main drawback of this approach is it amplifies the noise in homogeneous regions and creates false colours. Some research lines work to palliate this problems like [9,12] combining different techniques.

2.1   Deep Learning

In the last few years there have been a great variety of studies demonstrating the effectiveness of deep learning methods in different application domains. In addition to the classic Mixed National Institute of Standards and Technology (MNIST) handwriting challenge [5] many applications have been studied such as image classification [17] or speech recognition [11] and many others.

The growth of available data in computers for processing [19] combined with the increasing processing capabilities of computers initiated this revolution. Deep learning is the process that allows patterns to be found, discovered or learned in large, complex data. Although applications are not restricted to image processing tasks, this is the domain that has seen the biggest change in response to the introduction of these deep learning methods.

In the case of neural network for image dehazing there are only a few works and none of them, to the best of the authors knowledge, are tested in underwater environments. In [2,13] authors propose a deep learning solution to estimate transmission. In the case of [22], it uses a random forest and several haze-relevant features to estimate the transmission. The approaches proposed in [18,27]

perform a similar step generating synthetic images from non-hazy ones, but they also create a synthetic depth-map to produce the training images.

The main drawback of this learning approaches is they use synthetic images created from non-hazy images to train a neural network that estimates transmission due to the difficulty of finding hazy and non-hazy pairs. These images ignore many problems of real images and difficult its use in a real situation.

Other learning techniques have also been explored in this context, [23] has examined the use of Markov Random Fields (MRF) and a training stage to learn how to assign the most probable color to each pixel. The MRF is trained using pairs of input and output images learning transforms from a patch of degraded colors to restored colors. In order to acquire the desired output images, a light source is used, obtaining a better image to train. However, the method relies on a illumination system that obtains "groundtruth" images to train the system.

3   Proposed Method

The approach proposed in this work uses a convolutional neural network to learn the transformation from raw acquired images to enhanced images thus it can be used as input for other vision algorithms. In order to train and evaluate the system the images have been processed using the method in [1].
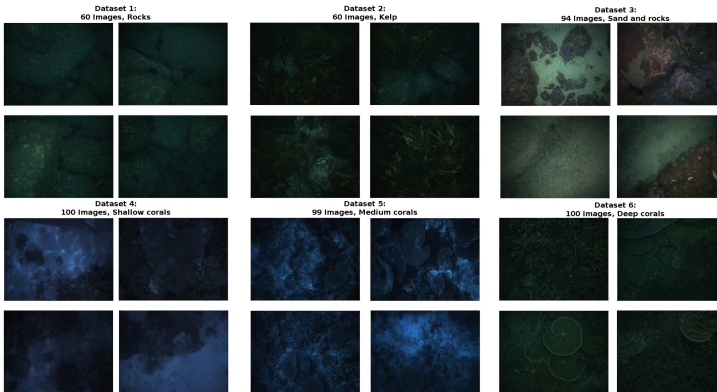

Fig. 1. Images of the different datasets used in the work.

The images used to train the neural network have been taking by an underwater camera mounted in an autonomous underwater vehicle [26] during different real underwater interventions. The images have been divided in 6 sets depending on the characteristics of the images, as Fig. 1 shows. Furthermore, the images have been chosen to cover a wide variety of textures at different depths in order to train different kind of images.

These dataset division allows to train with some sets of images and validate the neural network with images from a different intervention. Thus it is possible

to test the system in the case of a different intervention. Besides this, each dataset has been organised in a training and testing set with images randomly selected to measure the training performance.

Several architectures have been tested to train the system, but the one proposed can be seen in Fig. 2. As can be seen, the neural net takes as input the whole image and goes through 6 convolutional steps. In the first one, the image size is reduced in an additional pooling step that extracts the most relevant features. Besides this, every convolutional step but the last one also includes a Rectifier Linear Unit (ReLU) layer as activation function.
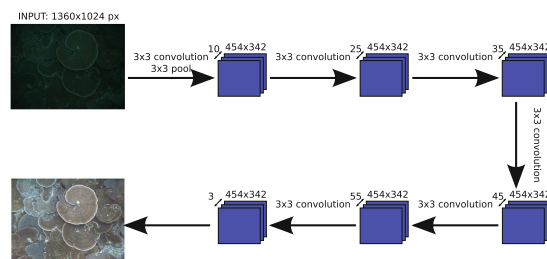
Fig. 2. Architecture of the convolutional network used to dehaze.

With each convolutional layer the number of features extracted increase from the 3 initial of the raw image (RGB) to 55 after five steps. At this moment the features are combined in the last neural network step to produce a matrix of 3 features that corresponds to the restored image.

In order to train the parameters in the neural network, the Adam optimizer, a gradient descent method, has been used with a $l2$ loss function as minimization function. The $l2$ loss function is a commonly used function that computes the squared sum of the differences between the estimated x and groundtruth y values: $l2 = \sum_{i=0}^{n}(y_i - x_i)^2$. In this case minimizing the l2 loss means minimizing the differences of intensities between the restored image and the ones estimated by the neural network.

As a consequence, the neural network learns to perform the same transform applied with the restoration methodology. However, the restoration method used to train requires a depthmap and a whole dataset of images while the neural network will need to do it with just a single image.

4 Results

Two experiments have been conducted to evaluate the precision of the neural network estimations. In the first case all the datasets have been used to train and evaluate it with the test images. However, this is not a realistic situation as training images for the intervention location are not usually available. For this

reason, the second experiment simulates this situation training with all but one dataset that is used to validate the system.

In order to evaluate the precision of the neural network predictions the images have been enhanced with two commonly used techniques and compared with the proposed approach. The first, histogram equalization, analyses the histogram of the raw image and displaces it to follow the desired distribution. The histogram equalization used in this paper is the most commonly used, modifies the pixel intensities to follow a normal distribution for every channel.

The second compared algorithm is an Automatic Color Enhancement (ACE), as explained in [8], that is also used in underwater environments in [14]. This technique enhances the image based on a simple model of the human visual system, inspired by different techniques such as gray world transformation, white patch assumption, lateral inhibition and local global adaptation. The main drawback of this technique is it is computationally complex, each image requires around 1.5 s in a i5 at 3.2 Ghz with a Geforce 960GTX while the time to process a single image in a neural network is 0.013 s.

4.1   Experiment 1

In this experiment all the datasets have been used to train the neural network keeping a few images of each in order to evaluate it. The system has been trained for 1700 epochs, reaching a 5.6% training error. This error is the mean difference between each intensity pixel and its groundtruth counterpart. In order to show percent errors intensities are transformed from 0–255 to 0–1 range.

The results for the test images of each dataset can be seen in Table 1 together with ACE and histogram equalization techniques. As can be seen the proposed method obtains the best results in all cases. This is not surprising as it is training with images from the same survey, thus it has similar examples that help to dehaze the raw image. But it is important that the neural network is able to learn the transform and correctly apply it to new images.

Table 1. Results for the experiment 1: training with every dataset.

Technique	Rocks	Kelp	Rocks-sand	Deep	Medium	Shallow
Proposed	3.5%	6.5%	5.1%	4.1%	3.4%	3.2%
ACE	6.8%	9.1%	15.3%	15.7%	7.5%	9.1%
Histeq	25.6%	37.9%	29.4%	20.5%	27.6%	27.5%

Another interesting result is ACE is obtaining results closer to the target image than the histogram equalization. Although the histogram equalization is enhancing the raw images is still far from the restored image. This means the colors generated by the histogram equalization are exacerbated producing false colors that were not present in the real objects.

It is also important to notice that the techniques perform very differently depending on the datasets. ACE has around 130% higher error in the case of deep corals or rocks-sand than in rocks dataset. However, in the case of histogram equalization deep corals is the best case scenario according to the used metric. This means the technique performance depends on the characteristics of the input images such as object colors or noise.
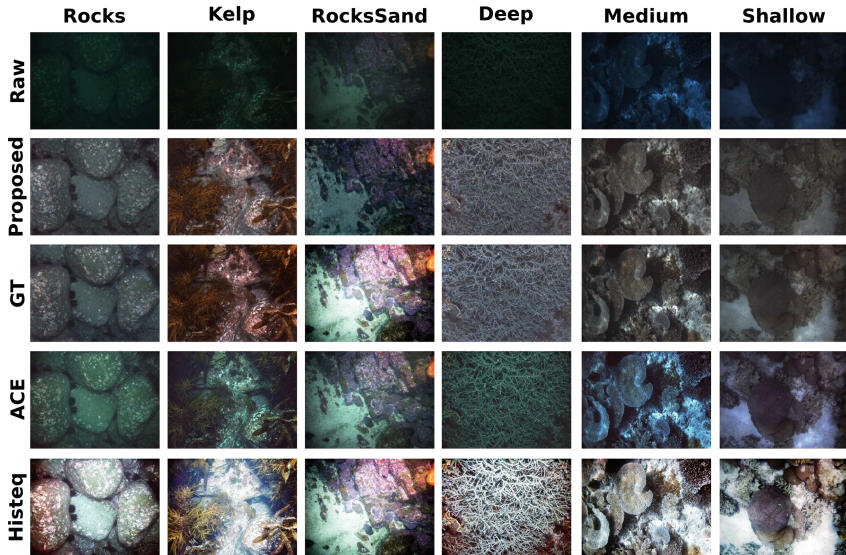
Fig. 3. Comparison of the image dehazing using different techniques in the first experiment. (Color figure online)

The visual results that can be seen in Fig. 3 reflect the numerical results. A test image of every dataset for each compared technique is showed together with the raw and groundtruth (GT). As can be seen the proposed method and the groundtruth images are indistinguishable in most cases, and in the cases that are different such as RocksSand dataset it is difficult to decide which one is better.

The ACE methodology obtains slightly uncorrected images. The algorithm is not able to completely remove the haze obtaining bluish or greenish images depending on the input. However the images are greatly enhanced. On the other hand, histogram equalization overcorrect the images producing unnatural images with extreme colours in some cases such as the kelp dataset, but it shows at least part of the error comes from the fact that the images are brighter than the produced by the restoration method.
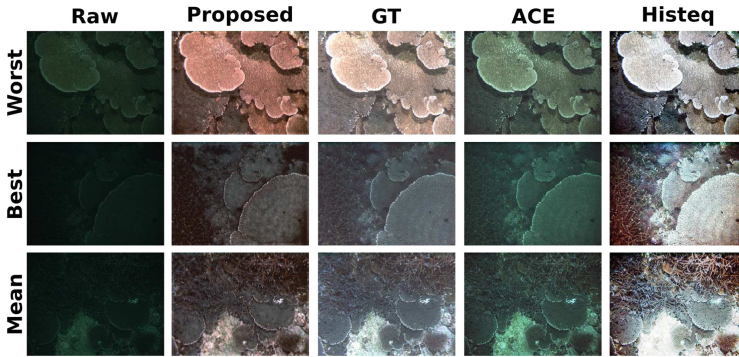
4.2   Experiment 2

The results of the previous experiment show the system is able to learn dehaze transformations given training images, but a generalization experiment is needed.

Taking this into account the second experiment focus on training with all but one dataset and use this last dataset as validation.

The deep corals dataset has been chosen as validation dataset as it seems to be the most complex for other techniques, thus it is a more challenging dataset. For this reason the resting datasets are included in the training scheme and a new neural network has been trained with them.

The results in this case are closer to the ACE performance, but the neural network still performs better. The neural network obtained a 14.1% error in the validation set while maintaining a 5% in the validation set. This is close to the 15.7% error of the ACE technique. However, processing an image with ACE requires 1.5 s making difficult to use it in a real time environment while the neural network needed 0.013 s per image. The histogram equalization is far from this results producing a 20.5% error for the validation set.

Fig. 4. Comparison of the image dehazing using different techniques in the second experiment. (Color figure online)

The visual results are displayed in Fig. 4. In this case three images from the validation set are showed, the worst and best performing images have been chosen together with one close to the mean error. As can be seen the neural network solution slightly overcorrects the images compared with the groundtruth result. In the case of ACE, images are not completely corrected showing a greenish color. Finally histogram equalization overcorrects inputs even more than the proposed image resulting in colors very different from the target with too dark and too bright zones.

This experiment proves the neural network performance is good, although numerical results are close to the ACE performance visual results look much more natural. Furthermore, the computation time is extremely shorter allowing to include it in a real time system as a preprocessing step. Finally, it is important to remark that the training never saw an image of the validation dataset permitting to use images to train in a different location from the intervention.

5   Conclusions and Future Work

In this work a real time deep learning solution for image dehazing is proposed and compared with other state of the art alternatives. The system is trained with other restoration methodologies that require several inputs that are hard to estimate at the intervention time. However, when the system is trained it is able to correctly dehaze images in real time with only a still raw image as input.

The results show that the system is able to generalize and learn to dehaze with images from a location and be used in a different location. However, in this situation the results are slightly worse, but still outperform other state of the art alternatives for real time dehazing.

Furthermore, when images from the same location as the final intervention are available to include in the training stage the results are visually indistinguishable from restoration techniques. This allows to obtain restoration results with only a single image as input in real time.

References

1. Bryson, M., Johnson-Roberson, M., Pizarro, O., Williams, S.B.: True color correction of autonomous underwater vehicle imagery. J. Field Robot. **33**(6), 853–874 (2016)
2. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: an end-to-end system for single image haze removal. IEEE Trans. Image Process. **25**(11), 5187–5198 (2016)
3. Chiang, J.Y., Chen, Y.C.: Underwater image enhancement by wavelength compensation and dehazing. IEEE Trans. Image Process. **21**(4), 1756–1769 (2012)
4. De Novi, G., Melchiorri, C., García, J., Sanz, P., Ridao, P., Oliver, G.: A new approach for a reconfigurable autonomous underwater vehicle for intervention. In: 2009 3rd Annual IEEE Systems Conference, pp. 23–26. IEEE (2009)
5. Deng, L.: The MNIST database of handwritten digit images for machine learning research. IEEE Sig. Process. Mag. **29**(6), 141–142 (2012)
6. Drews, P., Nascimento, E., Moraes, F., Botelho, S., Campos, M.: Transmission estimation in underwater single images. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 825–830 (2013)
7. Garg, R., Mittal, B., Garg, S.: Histogram equalization techniques for image enhancement. Int. J. Electron. Commun. Technol. **2**(1), 107–111 (2011)
8. Getreuer, P.: Automatic color enhancement (ACE) and its fast implementation. Image Process. On Line **2**, 266–277 (2012)
9. Ghani, A.S.A., Isa, N.A.M.: Underwater image quality enhancement through integrated color model with rayleigh distribution. Appl. Soft Comput. **27**, 219–230 (2015)
10. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Trans. Pattern Anal. Mach. Intell. **33**(12), 2341–2353 (2011)

11. Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N., et al.: Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Sig. Process. Mag. **29**(6), 82–97 (2012)

12. Hitam, M.S., Awalludin, E.A., Yussof, W.N.J.H.W., Bachok, Z.: Mixture contrast limited adaptive histogram equalization for underwater image enhancement. In: 2013 International Conference on Computer Applications Technology (ICCAT), pp. 1–5. IEEE (2013)

13. Hussain, F., Jeong, J.: Visibility enhancement of scene images degraded by foggy weather conditions with deep neural networks. J. Sens. **16**, 1–9 (2016)

14. Iqbal, K., Abdul Salam, R., Osman, M., Talib, A.Z., et al.: Underwater image enhancement using an integrated colour model. IAENG Int. J. Comput. Sci. **32**(2), 239–244 (2007)

15. Jaffe, J.S.: Computer modeling and the design of optimal underwater imaging systems. IEEE J. Oceanic Eng. **15**(2), 101–111 (1990)

16. Jaffe, J.S.: Enhanced extended range underwater imaging via structured illumination. Opt. Express **18**(12), 12328–12340 (2010)

17. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)

18. Mai, J., Zhu, Q., Wu, D., Xie, Y., Wang, L.: Back propagation neural network dehazing. In: 2014 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 1433–1438. IEEE (2014)

19. McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D., Barton, D.: Big data. The management revolution. Harvard Bus. Rev. **90**(10), 61–67 (2012)

20. Raimondo, S., Silvia, C.: Underwater image processing: state of the art of restoration and image enhancement methods. EURASIP J. Adv. Sig. Process. **2010**, 746052 (2010)

21. Roser, M., Dunbabin, M., Geiger, A.: Simultaneous underwater visibility assessment, enhancement and improved stereo. In: 2014 IEEE International Conference on Robotics and Automation (ICRA), pp. 3840–3847. IEEE (2014)

22. Tang, K., Yang, J., Wang, J.: Investigating haze-relevant features in a learning framework for image dehazing. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2995–3002. IEEE (2014)

23. Torres-Méndez, L.A., Dudek, G.: Color correction of underwater images for aquatic robot inspection. In: Rangarajan, A., Vemuri, B., Yuille, A.L. (eds.) EMMCVPR 2005. LNCS, vol. 3757, pp. 60–73. Springer, Heidelberg (2005). doi:10.1007/11585978_5

24. Treibitz, T., Schechner, Y.Y.: Active polarization descattering. IEEE Trans. Pattern Anal. Mach. Intell. **31**(3), 385–399 (2009)

25. Vasilescu, I., Detweiler, C., Rus, D.: Color-accurate underwater imaging using perceptual adaptive illumination. Auton. Robots **31**(2–3), 285–296 (2011)

26. Williams, S.B., Pizarro, O.R., Jakuba, M.V., Johnson, C.R., Barrett, N.S., Babcock, R.C., Kendrick, G.A., Steinberg, P.D., Heyward, A.J., Doherty, P.J., et al.: Monitoring of benthic reference sites: using an autonomous underwater vehicle. IEEE Robot. Autom. Mag. **19**(1), 73–84 (2012)

27. Zhu, Q., Mai, J., Shao, L.: A fast single image haze removal algorithm using color attenuation prior. IEEE Trans. Image Process. **24**(11), 3522–3533 (2015)