International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 12
April 2018

Underwater Image Enhancement using Image Fusion Technique

G. VenkataRao[1], N. Ashokkumar[2], M. Apparao[3], P. Bharath kumar[4] & K. Ravindra Reddy[5]
[1]Associate Professor, [2,3,4,5]Student,
Department of Electronics and Communication,
Lakireddy Balireddy College of Engineering, Mylavaram, India.

Abstract:

Underwater photography has been around long enough for all of the problems to be recognized and their respective solutions be discovered. Water being multiple times heavier and viscous than air, it has the ability to retain particles in suspension through which the light required for the purpose of imaging needs to travel. This causes the light levels to drop drastically the deeper we go, a phenomenon called as Backscatter. Other problems such as loss of color, loss of contrast, Refraction, Magnification, flooding etc., creep up in underwater imaging. The Proposed method is a single-image approach that doesn't require special hardware and knowledge of underwater conditions or scene structure. It is practically a blend of multiple images which are themselves derived from the color-compensated and white balanced version of the original input image. To avoid the sharp weight map transitions that results in creation of artifacts in the lower frequency components of the reconstructed image, we adapt a Multiple-scale fusion approach. The method enhances the contrast and sharpness of the image. The enhanced result is further improved by color-correction, lightened dark regions, naturalness preservation, and well enhanced edges and details.

Keywords

Underwater, Enhancement, Fusion Technique, Weight Maps, Backscatter.

1. Introduction

Under water imaging in the basic essence means the act of Photography or Capturing images under the water bodies using special Water-proof Imaging equipment. The very fact that the physical composure and properties of air and water differ from one another implies that Imaging in both mediums will not have the same quality or outputs.

Usually the speed with which light travels in vacuum is $3x10^8$ m/sec which is 3,00,000 km/sec. But this isn't the case with air, glass and other materials. The speed of light is highest in vacuum. As the density and permeability of the medium in which light propagates changes, the speed with which light travels also changes due to the viscous effect of the medium. As water is usually much denser than air and tends to hold suspended particles in it such as mud particles, wastage, plants and fishes, the speed with which light propagates in water faces a significant impact. As the water molecules are larger than wavelength of light, the light gets reflected and refracted multiple times the effective intensity of light reduces significantly. Many problems are identified in the underwater photography, some of them are as follows:

- **Density:** Water is many times denser than air and is capable of holding matter in suspension through which the scene needs to be imaged.
- **Attenuation**: The light which does penetrate is then absorbed far quicker than on land because water is much denser. This results in light levels dropping drastically the deeper we go.
- **Loss of color:** The most clear and pure water have a strong cyan or bluey/green cast and absorbs different colours at different rates. So, as a result colours gets attenuated completely after propagation of light to some extent, starting from red and all the way to green.
- Refraction, Magnification
- **Loss of Contrast**: Water limits our horizons, it also reduces the contrast and this in turn affects the clarity of your shots.

The light normally gets absorbed by the water particles due to multiple reflections. In many cases the attenuation undergone by light will be exponential with the depth and distance of propagation of light. In addition to this, the scattering of light particles takes place as the particles of light are smaller than molecules of water, this scattering causes the light to change its direction multiple times there by reducing the contrast of the scene a lot and gives a misty appearance of the objects under water. In Ocean water the visibility is around 10metres. Wavelength of the light plays a key role in understanding the extent of propagation of light in a given medium. In Water as the molecules are larger than wavelength of light, the wavelengths which are longer gets attenuated because the waves which are longer in length cannot propagate for a complete cycle with actually colliding or getting reflected by a water molecule. This is the very reason why all the scenes from under water appear in bluish shade colours, as the colours in the upper end of the visible part of the spectrum gets attenuated more. With propagation distance increasing only the bluer-components manages to reach the destination and the camera lens. So in under water imaging, the blue colour is the dominant and key component. Actually, the entire lower side of the visible part of the spectrum important to imaging. This phenomenon gets intensified as the depth of water increases because as the depth of water increases parameters such as pressure under water comes into play in addition to the normal scattering done by the water molecules. The hazed and blurred components of the images are due the poor contrast, which is induced because of the very fact that colours on the upper band of the visible spectrum gets attenuated heavily during propagation in water.

2. Enhancement Techniques

Many attempts were made till date to enhance the visibility, quality and accuracy of underwater images. Algorithms along with hardware and software solutions were proposed for enhancement of the degraded images. The popular methods of dealing with this kind of degraded images is using multiple imaging techniques, Specialized hardware components/Add-ons such as polarizing filters etc., to the existing camera set-up exclusively for the underwater-imaging purposes. Though the techniques mentioned are quite effective they are not feasible to apply in the Realtime operations as the capital and maintenance costs sky rockets as the components are very expensive. Also, if in order to use polarized filters, we'll need a custom-made filter with specific degrees of polarization for every different scene and this reduces practicality of this method drastically because we cannot just make filters with required degree of polarization for every different color casted scenery for enjoying the perfect results. Other methods such as Laser-gated technology and Synchronous scanning increases the investment, complex design & implementation and also the imaging time giving rise to shutter lag.

Coming to the multiple imaging techniques (either software or hardware), they require the scene under question to be shot in multiple angles and environments, which obviously cannot be done when the imaging is being done under water. Firstly, the setting up of imaging equipment underwater is difficult. Secondly, we cannot wait for hours along to just let the environmental conditions to change just to have another shot at it which is very impracticable as the details we need to capture might have gone already. So, the above approach is highly impractical and impossible to apply for videos.

3. Proposed Method

In this paper, we are introducing a modified approach to deal with the degraded underwater images which utilizes a single-image to extract all the details and improve the image quality to the possible extent. The fact that the method uses a single-image to enhance implies that it could be implemented & extended to process videos with dynamic content filled scenes too with ease.

The underlying principle of this very method is Image-Fusion Technique which is already being used extensively for Realtime-applications like Magnetic-Resonance Image, Image compositing, Computed Tomography, Multi spectral video enhancement, Positron emission Tomography, defogging, single-photon emission computed tomography, and HDR imaging. All of these methods require multiple images, which our method doesn't. The proposed method does require multiple images for fusion to work but these multiple inputs are obtained from the input image itself and are used as inputs for fusion. The derivatives of the input image are used to extract details and develop required weight maps to finally fuse them together.

We are striving to make this approach best in parameters computational power requirement and to be straight to the point so that it will be faster in execution even on legacy devices with very low specifications.

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 12
April 2018

The degradation that the image undergoes could be modelled as sequence of multiplicative & additive occurrences of noise. As the noise is additive and multiplicative, the basic image enhancement/correction techniques such as white balance, histogram equalization, contrast stretching, color correction etc., fails to perfectly enhance the image as only certain components gets highlighted. So, if we manage to combine the goodness of these techniques they will result in better enhanced images depending on the sequence those techniques were applied and the weightage given to each method. Technically, we are just filtering the given input image directly with a fusion based method whose inputs were the extracted innate characteristics of the given input image. Once again, the quality of the images enhanced is strictly dependent on the choice of inputs to the fusion technique and the weights used. So, we need to obtain certain operators which works the best for underwater conditions. Thus, the basic idea of the principle of image fusion is to merge multiple input images by taking and maintaining the most important features intact. Each input of the fusion represents a unique characteristic of the input image. The images to be fused, as well as their associated weight maps, are well defined to promote the transfer of edges and color contrast to the output image. The weight maps specify the spatial and pixel relationships. These weights map the value to pixels to get the optimum version of the image. Higher weight ensures that the pixel will be present in the final output. The method uses multiresolution processing by using Laplacian pyramids to make itself robust. Techniques like temporal bilateral filter can be used to retain the coherence between frames of a video. Even though the suggested method is quite different from well-established models, our method incorporates certain inputs and weight maps carefully chosen to deal with images taken in underwater environment

Most of the images have a constituent backscatter element which arises due to artificial illumination of light from the same direction of imaging (the light particles hit the water molecules rather than the object and gets reflected back to the camera). The effect of backscatter is reduced using this approach. Backscatter can be avoided in the first place at the time of imaging by using decent illumination angles when it is required to use artificial illumination source.

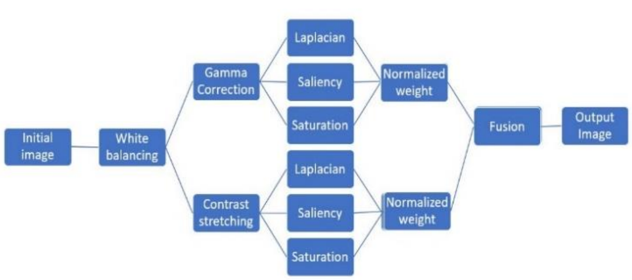


Figure 1. Flowchart of the proposed method

The mains steps of this enhancement strategy are Deriving inputs from the given input image, defining weight maps and applying multiscale fusion with the obtained weight measures.

White Balancing

The first and foremost step we need to do is to white balance the given input image so that any kind of color casts present in the given input image will be removed. The color cast is the component which is introduced by the light source in a given medium which causes the images to appear in that particular part of the color spectrum. For example, the images taken in a room lit by a fluorescent lamp appears whitish, whereas an image taken with an incandescent light gives a yellowish tint due to the warm nature of the light source. Correction of these color casts is white balancing. The white balanced image gives a more natural appearance and forms the base to our method. This white balanced image is further enhanced in order to remove the noise. The other inputs of the fusion technique are derived from this white balanced image so as to extract details from the entire range of intensities.

As the depth of water increases and exceeds 30ft, white balancing starts to lose its potency since the colors that got attenuated becomes very hard to restore. Also, the underwater images suffer from poor contrast due to the properties of medium. A wide range of white balancing techniques are available. Methods such as Finlayson's technique Shades-of-Grey [i] computes the illumination of the scene for each channel by using the Minkowski p -norm. For $p = 1$, this expression is a particular case of the Gray-World [i] while for $p = \infty$ it approximates the White-Patch hypothesis [ii]. Despite of its simplicity, the low-level approach of Finlayson and Trezzi [i] has shown to yield comparative results to those of more complex white balance algorithms such as the re- cent method of [iii] that relies on natural

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 12
April 2018

image statistics. The Grey-Edge hypothesis of Weijer and Gevers [iv], similarly with Shades-of-Grey [i] can also be formulated by extending the *p*-th Minkowski form.

In actual study, it had been found that solutions from the White-Patch algorithm [ii] usually fail as the underwater images contains only attenuated regions with reflection. Also, the method of Gray- Edge algorithm [iv] performs miserably in such cases, mainly due to the fact that underwater images are usually low in contrast and have less visible edges than normal images. The most appropriate strategy for underwater environment is the Gray-World approach of Buchsbaum et al. [ii]. But one issue of color-deviation is noticed whenever the illumination is estimated poorly. For example, in the underwater images, where the scene is overall blue, the parts that are wrongly balanced will show reddish shaded for the entire scene by increasing the average value estimated with a percentage λ instead to variate the norm value of *p*. So, the illumination is obtained from the measure μI which is computed from the mean of the image μref *and* adjusted by the parameter λ:

$$\mu I = 0.5+\lambda\mu ref$$

The value of λ is set to high value if the set of colors in the given image is small in number. λ varies in range of 0 to 0.5 based on the number of colors. Study showed that λ value of 0.2 produces good results. As white balancing alone cannot solve the problem of visibility we also derive other images from the input image.

The first derived input is represented by the color corrected version of the white balanced image while the second is computed as a contrast enhanced version of the underwater image after a noise reduction operation is performed.

Noise Reduction

The impurities present in the water along with conditions such as special illumination conditions give rise to noise in images. Removing this noise while preserving edges of the images is a complicated process but if done enhances the sharpness the image. This can be achieved by using different techniques such as median filtering anisotropic filtering and bilateral filtering. Applying these techniques for images is pretty easy but implementing the same for videos is quite difficult as we have to deal with both spatial and temporal coherence in case of videos. Of these, bilateral filtering turned out to be better as it is a non-iterative edge preserving smoothening filter and

$$J_s = \frac{1}{k(s)} \sum_{p\in\Omega} f(p-s,\sigma_f)g(D(p,s),\sigma_g)I_p$$

has proven useful in many solving many problems such as local tone mapping mesh smoothing and dual photography enhancement. If we look at the operation of bilateral filter in the spatial domain, the kernel blends center pixel s of the kernel with the neighboring pixels of the image p that penalizes pixels across edges that have large intensity differences.

On the other hand, for videos, the bilateral filter fails to enhance as much as it does to an image. In the absence of motion, the simple average of all the pixels at a coordinate through time will represent a decent solution but for dynamic scenes this approach yields ghosting artifacts.

For this case a temporal bilateral filter must be used on the frames of the white balanced version of the image that aims to reduce noise and smoothing frames while preserving temporal coherence. Choosing the value of σg is a compromise because choosing high value of σg leads to appearance of halos whereas the small values of σg won't be able to reduce noise levels to the required level. Computing the difference between the squares of intensities to compare them is more efficient than the conventional comparing of the individual intensities, also we do this by computing sum of the squared difference of the small spatial neighborhood around the center pixel of sub image and the center pixel of the kernel by using a gaussian weight. The usual size of neighborhood ψ is 3 x 3 or 5 x 5. This approach decreases the ambiguity between edges and noises as

$$D(p,s) = \sum_{x}^{\Psi}\sum_{y}^{\Psi}\Gamma(x,y)(I_p - I_s)^2$$

larger neighborhoods have lesser effect on temporal noise.

The second to the fusion technique is derived from the noise less and color corrected instance of the original image. This input is supposed to help us overcome the loss caused due to scattering by increasing the contrast of the image which is achieved in our case by a simple contrast local adaptive

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 12
April 2018

histogram equalization technique as it causes less amount of distortion while improving intensity range of the image and also it doesn't sacrifice some areas for others. The contrast operators are to be either specified or should be calculated from the image.

Weights for the Fusion process

The weights decide the quality and amount of enhancement of the image. So, the design of the weight maps is crucial for the algorithm to get the most out of the degrade image. Image restoration is highly dependent on the color correction and it is highly subjective to the perception of human eye. So, it is important to preserve components of an image such as salient features, global and local contrasts or exposedness/saturation. But they are difficult to blend or merge on per pixel basis. Here the weights come into play as higher weights imply that having that pixel will be advantageous to the final image.

Laplacian contrast weight (WL)

It deals with global contrast of the image by using a Laplacian filter to improve the luminance at each channel; of the input and by computing the absolute value of the filter result. This method is extensively used for tone mapping and extension of depth in images as it assigns higher level values to the edges and the textures. But the results of this method are not sufficient because it cannot differentiate flat and ramp regions. So, to overcome this we go for local contrast maps that can calculate the independent local distribution.

Local contrast weight (WLC)

It is nothing but the relation between each pixel and the average of its neighborhood. This improves the local contrast appearance as it advantages the transitions and mainly the parts that were shadowed in the second input. Its is the standard deviation of the luminance level and the local average of its surrounding region.

The (WLC) is computed as the standard deviation between pixel luminance level and the local average of its surrounding region:

$$WLC(x, y) = I^k - I^k_{whc}$$

Here I^k is the luminance channel of the input and the I^k_{whc} is the low pass version of it. The filtered version I^k_{whc} is obtained by employing a small 5 × 5 (1/16 [1, 4, 6, 4, 1]) binomial kernel with high frequency cut-off value $\omega hc = \pi/2.75$.

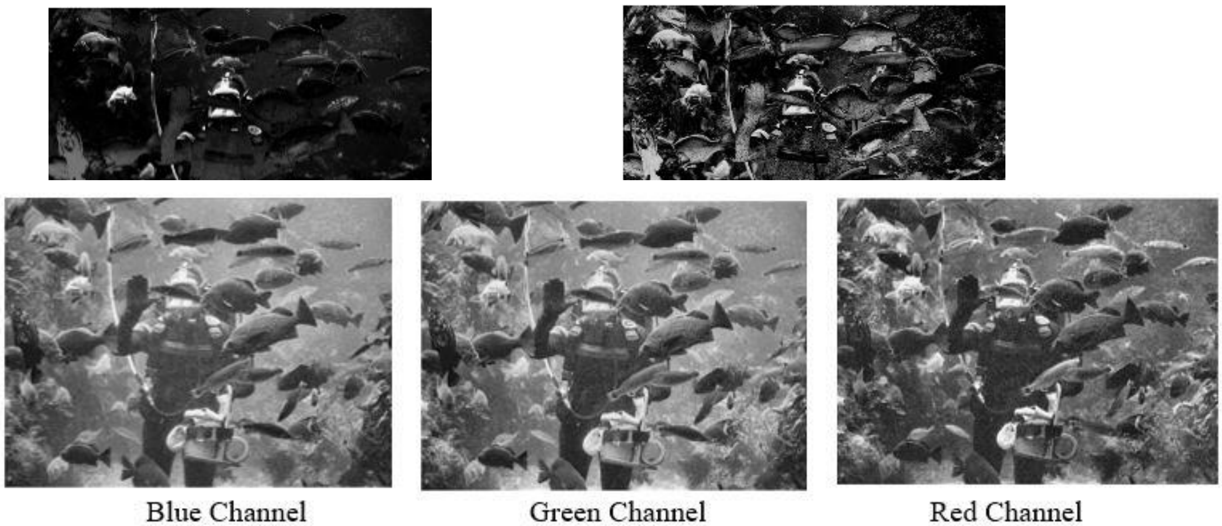
Saliency weight (WS)

Saliency is the components that gives the discrimination information in the image. It is the uniqueness of details in the image that are often lost in the underwater images. This weight is supposed to improve the saliency of the given image. To measure this quality, the saliency algorithm of Achanta et al. [v] is used in this approach. The algorithm is straightforward in implementation. Saliency

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 12
April 2018

Figure 2. Figure showing the inputs to the algorithm, white balanced, histogram equalized versions and all the derived weight maps for the Fusion Technique except saliency maps

Saliency-Weight-Maps



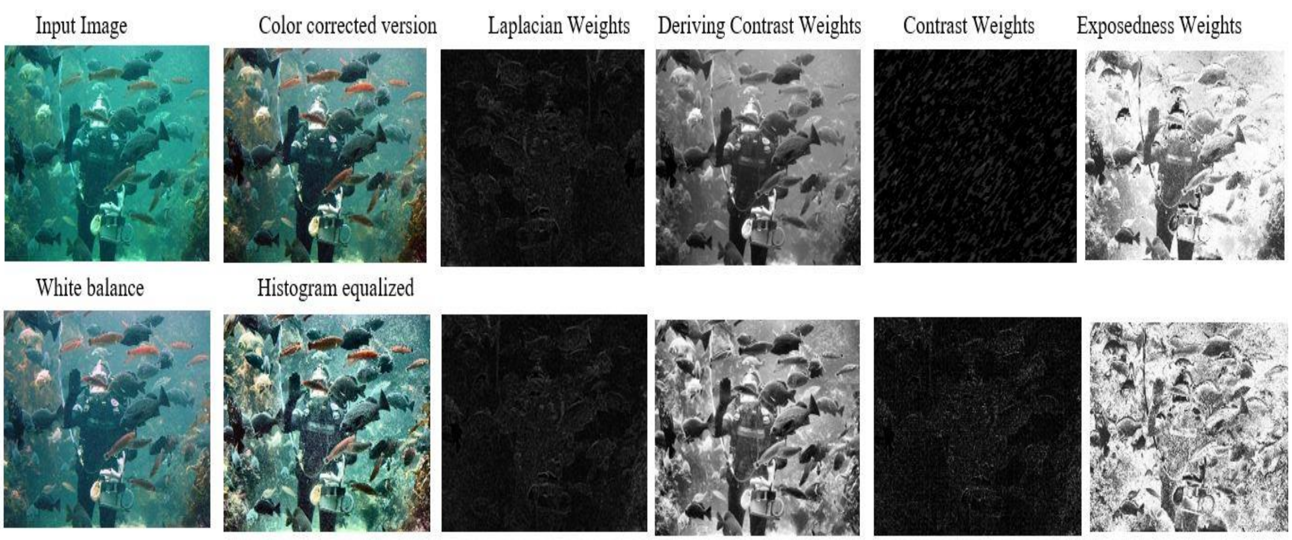


Figure 3,4. Saliency weight maps of color corrected input and contrast stretched inputs respectively

Laplacian Pyramid Outputs (Grey Scaled)

Figure 5. Laplacian Pyramid Outputs

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 12
April 2018

Normalized Weight Maps and Output

Figure 6. Final Normalized inputs for Fusion and final output of the Fusion technique

map inclines to favor highlighted areas in the image. In order to improve the quality of the results, we go for the exposedness map to preserve the mid-tones that could be manipulated in some special scenarios.

Exposedness weight (WE): It calculates how much a pixel is exposed i.e., to light. This deduced quality helps to estimate and preserve a uniform appearance of the local contrast, which ideally is neither exaggerated nor undermined. Usually, the pixels lean towards having a high exposed appearance whenever their normalized values are nearer to the average of 0.5. The weight map is modelled as a Gaussian-

$$W_E(x, y) = exp\left(-\frac{(I^k(x, y) - 0.5)^2}{2\sigma^2}\right)$$

modeled distance to the average normalized range value (0.5):

where $I^k(x, y)$ represents the value of the pixel (x, y) of the input image I^k, the standard deviation is set to $\sigma = 0.25$. This map assigns higher values to tones whose distance is closer to zero, whereas the pixels which are characterized by large distances, are associated with the over and under-exposed regions. In result, this weight manipulates the output of the saliency map and gives a well-preserved appearance of the fused image. In order to get consistent results, we use the normalized weight values $\bar{W}$ (for an input k the normalized weight is computed as,

$$\mathcal{R}(x, y) = \sum_{k=1}^{K} \bar{W}^k(x, y) I^k(x, y)$$

by limiting that the sum at each pixel location of the weight maps W equals to one.

Multiscale Fusion Process: The improved image version R(x, y) is obtained by fusing the above said inputs with the weight measures at each and every pixel location (x, y):

where I^k is the input (k is the index of the inputs - K = 2 in our case) which is weighted by the normalized weight maps $\bar{W}k$.

$$\bar{W}^k = W^k / \sum_{k=1}^{K} W^k)$$

The normalized weights $\bar{W}$ are obtained by normalizing all k weight maps W such that all the values are limited to give sum of one. From extensive study it's been found that direct naïve fusing of the image inputs with the weight maps introduces halos and undesirable artifacts undesirable halos. A good approach to solve this problem is to go for multi-scale filters (Both linear and nonlinear). The design of non-linear filters is quite complex and also its been found that they don't provide significant improvement in quality of the images when compared to the Laplacian and gaussian pyramids. So, it is computationally effective and efficient to use Laplacian pyramid style decomposition. In Laplacian pyramid decomposition model every image is modelled as a sum of patterns evaluated at multiple layers which are derived from a Laplacian operator/filter. The inputs to this Laplacian pyramid are convolved with a gaussian kernel to obtain a low-pass version of them. The standard deviation set by us controls the higher cutoff frequency. At every level of the pyramid the difference between the original and low passed version to obtain the difference levels. Also, the difference between outputs of adjacent gaussian pyramid are performed. The result is the Laplacian pyramid which is nothing but a

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 12
April 2018

quasi-bandpass version of the image. In this algorithm the input images are decomposed using Laplacian pyramids and the weight maps are decomposed using gaussian pyramids. The gaussian and the Laplacian pyramids are ensured to have same number of levels so that mixing/fusing them together at each level to obtain the fused pyramid.

$$\mathcal{R}^l(x,y) = \sum_{k=1}^{K} G^l\left\{\bar{W}^k(x,y)\right\} L^l\left\{I^k(x,y)\right\}$$

Where l is the number of the pyramid levels employed (typically the number of levels is 5), L {I} is the Laplacian version of the input I, and G & W¯ ' is the Gaussian version of the normalized weight map W¯. This step is executed successively for each pyramid layer, in a bottom-up manner. The enhanced output is obtained by summing the fused contribution of all inputs. The Laplacian multi-scale strategy performs relatively fast implying a good trade-off between speed and accuracy. The artifacts in the image are reduced as the sharp transitions between the weight maps are minimized as we are performing the fusion process independently at every scale level. The Multi-scale fusion technique is modelled from the way a human eye works as it is sensitive to local contrast and changes in edges and the corners in the image.

3. Results: The proposed approach was tested for real underwater images taken from various sources. Which directly implies that images were collected from different camera setups. it is to be noticed that all the processing being done in this approach is on 8-bit images, but there are professional cameras with option to store the images in RAW format of 12-bit unprocessed data. The proposed technique is computationally efficient it took only 0.5 seconds to process an input image. This approach with certain optimizations could be able to run on common hardware in real time, processing video frames just as they arrive. By just having a glance at the inputs and outputs of this approach on can easily understand the improvement in the output in terms of edges, contrast, color accuracy etc., by comparing the images. By using this technique on the white balanced version of one of the inputs we are able to obtain a more pleasing image. The results produced by this method are much better than the results produced by other popular methods in many aspects. But in some scenarios, such as deep scene under water images and darkened scenarios with poor strobes and poor artificial illumination, the proposed theory showed limitations. But even in such cases some amount of enhancement is obtained and can be easily noticed.

The color accuracy is checked using iqm metric which is itself modeled from the way the human eye percepts images. This metric utilizes a model of the human visual system being sensitive to three types of structural changes: loss of visible contrast (green), amplification of invisible contrast (blue) and reversal of visible contrast (red). The professional cameras introduce different color casts. This approach managed to enhance the image despite those colorcasts as the white balancing algorithm utilizes image data to remove color casts rather than using prefixed models and managed to maintain the uniformity in the color appearance of images taken with different cameras. Distant parts of the scene cannot be recovered reliably if the illumination is poor. This can be overcome using Dark channel prior method.

Conclusion

The approach is suitable for many applications. Matching images by local feature points is a fundamental task of many computer vision applications. The results of this approach when compared with the original images gave higher number of matches with details, which implies that there are very low number of artifacts introduced during enhancement or none at all. Promising metrics demonstrate that this algorithm doesn't introduce artifacts but restores both global contrast and local features of underwater images. Segmentation is the division of images into disjoint and homogeneous regions with respect to some characteristics. In the underwater images processed with this approach the segmentation result is more consistent and the filtered boundaries are more accurate. Image dehazing is the process of removing the haze and fog effects from the spoilt images. Because of similarities between hazy and underwater environments due to the light scattering process, we found our strategy appropriate for this challenging task. However, as explained previously, since the underwater light propagation is more complex We can assume safely that image dehazing could be seen as a subclass of the underwater image restoration problem. Comparative results with state-of-the-art single image dehazing techniques.

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 12
April 2018

References

[i] G. D. Finlayson and E. Trezzi. Shades of gray and colour constancy. IS&T/SID Twelfth Color Imaging Conference: Color Science,Systems and Applications, Society for Imaging Science and Technology, pages 37–41, 20

[ii] M. Ebner. Color Constancy, Wiley 1st edition. 2

[iii] A. Gijsenij and T. Gevers. Color constancy using natural image statistics and scene semantics. IEEE Trans. Pattern Anal. Mach. Intell., 33(4):687–6

[iv] J. van de Weijer, T. Gevers, and A. Gijsenij. Edge based color constancy. IEEE Transactions on Image Process

[v] R. Achantay, S. Hemamiz, F. Estraday, and S. Susstrunky. Frequency-tuned salient region detection. IEEE CVPR, 20

[vi] C. O. Ancuti, C. Ancuti, C. Hermans, and P. Bekaert. A fast-semi-inverse approach to detect and remove the haze from a single image. ACCV, 20