

# Virtual High Dynamic Range Imaging for Underwater Drone Navigation

Bogusław Cyganek<sup>a,\*</sup>, Michał Woźniak<sup>b</sup>

<sup>a</sup>AGH University of Science and Technology, Al. Mickiewicza 30, Kraków, Poland

<sup>b</sup>Wrocław University of Science and Technology, Department of Systems and Computer Networks, Poland

\*Corresponding Author: cyganek@agh.edu.pl

## Abstract

Digital images contain few orders of magnitude smaller dynamic range than capabilities of a human eye. High dynamic range allows for perception of more details under wide lighting variations. Therefore acquisition of high dynamic range images is of high importance. However, the process of collecting of images at various exposures, then joining them together is not an easy one. Not always it is even possible to acquire a series of images of different exposures. However, even a single image due to not perfect conditions of acquisition usually contains much more information than available on the first sight. In this paper we show a method of so called virtual high dynamic range imaging formation which allows for obtaining at least virtually higher dynamical range just from an image acquired at a single exposure. Thanks to this the images convey more discriminative information as it is shown in the case of navigation of the underwater drone.

**Keywords:** high dynamic range imaging, drone navigation.

## 1. Introduction

Digital images from commodity cameras have very limited dynamic range of pixels, usually few order of magnitude lower than human eyes. This greatly limits amount of conveyed information, as well as discriminative properties of pixels. Thus, obtaining high dynamic range (HDR) images finds great interest. Usually this task is done by capturing a number of images of the same scene but each with different exposure settings. This way a series of dynamic range images is obtained which, after a fusion, form a high-dynamic range version. However, recovering HDR radiance map from images is a nonlinear tasks. For this purpose, a method of reconstructing image response curve and the radiance values from photographs was proposed by Debevec

and Malik<sup>(10)</sup>. Then only, knowing the response function, it is possible to reliably fuse multiple images in one HDR.

The other question is how to display such a HDR image since its dynamic range is usually much higher than capabilities of standard displays. For this purpose the tone mapping is applied which turns an HDR image into an low-dynamic range (LDR) one, suitable for display. An overview of methods and techniques for HDR creation can be found in the recent paper by Sen and Aguerrebere<sup>(18)</sup>. However, fusing a series of images captured with different exposures frequently leads to smearing and other artifacts. This is due to scene dynamics and increasing exposure times.

The other possibility of acquiring HDR images is to build special type image sensors and HDR cameras. For example Lapray<sup>(14)</sup> et al. propose an FPGA based camera that can output a real-time HDR video. The key element of this system is a novel CMOS sensor which allows real-time multi-frame acquisition with various exposures.

However, all of the above solutions suffer from a number of limitations, especially if ordinary cameras are the only solution in presence of dynamic scenes. In our developed underwater drone navigation system we faced such a situation. Thus, to enhance image dynamics we resorted to a third option which is a virtual HDR. The main idea is to work with a single image but post-processed with a number of tone adjusting curves, followed by a fusion and local contrast enhancement modules. Such methods were investigated e.g. by Chang<sup>(5)</sup> et al., who showed that local contrast enhancement is inversely proportional to the local standard deviation of an image. Further on, Cvetkovic<sup>(6)</sup> et al. proposed a multi-scale local contrast enhancement method. Then Ke<sup>(13)</sup> et al. proposed hardware-efficient virtual HDR. Their main idea is to construct a virtual photography out of a single low dynamic range image, followed by a local contrast enhancement. In this paper we follow the latter idea in the system of image enhancement for underwater drone

navigation. The problem with such images is their significant degradation due to poor visibility, noise, as well as loss of colors associated with light propagation properties in water.

In this paper we propose our version of the VHDR method for enhancement of underwater images. This writes into a broader framework of development of the visual front-end for underwater drone navigation aimed at underwater search. As will be shown, the VHDR imaging allows for better detail perception.

## 2. Computation of the Virtual High Dynamic Range Images

Architecture of the image processing chain is shown in Fig. 1. At first, an input LDR image is processed by a number of tone adjustment curves, resulting in a number of tone converted images. The role of these is to expose different ranges of the input image, potentially revealing hidden details. These are then fused back to form one HDR image. After that, image range conversion and contrast enhancement follow.

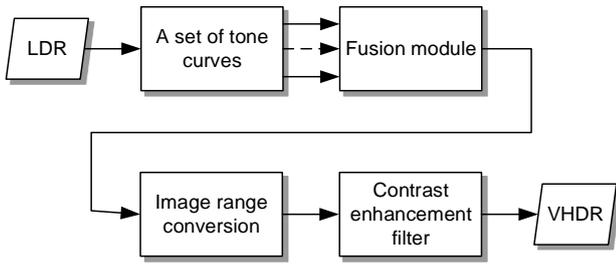


Fig. 1. Architecture of the proposed VHDR method.

Luminance adjustment is done with help of the logistic function, in its basic form given as follows

$$s_0(x) = \frac{A}{1 + e^{-k(x-x_0)}}, \quad (1)$$

where  $A$  stands for amplitude,  $k$  is a parameter, and  $x_0$  denotes the middle point. It is a “S” shape function with the inflection point at  $x_0$ . However, to generate series of virtual exposures we need three variants of the logistic function: convex, concave, as well as S-shaped, as shown in Fig. 2b. Also important is to assure that the domain and codomain cover the whole allowable range, which for 8-bits spans 256 levels. Thus we propose to modify (1), as follows

$$s_1(x, k, x_0) = \frac{A}{1 + e^{-Dk \frac{(x-x_0)}{R}}}, \quad (2)$$

where  $D$  is a domain range constant,  $A=P_{max}$  is a maximal pixel range of the input low dynamic range image. The parameter  $D$  controls the range of values of  $s_1$ , for which it is effectively different from 0 and 1, that is, these are the saturation values. In our system this parameter is set to 2 times 6 around  $x_0$ , which results in  $D=12$ .

Function (2) was further augmented with additional multiplicative and additive components to provide a concave, convex and S like shape, as shown in Fig. 2., as well as to cover full range of the codomain, i.e. in the range  $[P_{min}, P_{max}]$ , which in our case is  $[0,255]$ . The new proposed function reads as follows

$$s_2(x, k, x_0) = \left( \frac{|x_0 - P_{mid}|}{P_{mid}} + 1 \right) s_1(x, k, x_0) + s_1(0, k, x_0) \quad (3)$$

where for 8-bit input  $P_{mid}=(P_{max}-P_{min})/2=(255.0-0.0)/2=127.5$ , which denotes a middle value of the LDR image dynamical range. On the other hand, parameter  $x_0$  is chosen from the following three values

$$x_0 \in \{P_{min}, P_{mid}, P_{max}\}. \quad (4)$$

Each of the above values of  $x_0$  leads to one type of a curve as shown in Fig. 2. These are:

1. Convex case – The midtone data of the input image is stretched toward the highlight of the output side, resulting in a bright-toned image.
2. S-curve – The highlight and shadow of the input signal are enhanced, resulting in a higher contrast image.
3. Concave case – The midtone data of the input is stretched toward the shadow of the output, resulting in a dark-toned image.

The only one remaining parameter to consider is  $k$  which value was chosen experimentally, as shown in Fig. 2. In our experiments the following values were selected for each type of a tone mapping curve

$$k \in \begin{cases} \{0.3, 0.2\}, & \text{for } x_0 = P_{min} \\ \{0.4, 0.5\}, & \text{for } x_0 = P_{mid} \\ \{0.4, 0.3\}, & \text{for } x_0 = P_{max}. \end{cases} \quad (5)$$

With help of three  $x_0$  values, and two  $k$  parameters for each  $x_0$ , a series of  $V=6$  tone converted images  $I_v$  is generated. These parameters were chosen for experiments, although it is possible to set different values.

The consecutive fusion module operates with a number of  $V$  tone mapped images obtained in the previous step, as

shown in Fig. 1. Then, each pixel  $i_h$  in the output fused image  $I_h$  is computed as a weighted sum of all  $V$  pixels  $i_v$ , as follows<sup>(17)</sup>

$$i_h = \frac{\sum_{v=1}^V w_v i_v}{\sum_{v=1}^V w_v} \quad (6)$$

where  $i_h$  is an output pixel,  $i_v$  is an input pixel from one of the  $V$  tone mapped images  $I_v$ , and the weight  $w_v$  is computed as follows

$$w_v = e^{-\lambda \left( \frac{i_v - P_{mid}}{P_{mid}} \right)^2}, \quad (7)$$

where  $\lambda$  is a parameter controlling steepness of the weighting function. In our experiments  $\lambda$  is set to a value from the range 2.5-3.5. Let us observe that a weight  $w_v$  depends on a value of a relative distance of an actual pixel intensity  $i_v$  from the constant middle value  $P_{mid}$ . In effect, “promoted” are intensity values closer to  $P_{mid}$ , whereas the more distant ones get much less influence on the output pixel  $i_h$  in (6).

Pixel values obtained in (6) are usually concentrated around middle values of the allowable pixel range since these are boosted by the weights in (7). Therefore, these are scaled by the image range conversion module, as shown in Fig. 1, to spread their histogram. The scaling is done in accordance with the following formula

$$\hat{i}_h = \frac{i_h - \min(I_h)}{\max(I_h) - \min(I_h)} P_{max}, \quad \text{for each } i_h \in I_h, \quad (8)$$

where  $\min(I_h)$  and  $\max(I_h)$  are minimal and maximal values of the whole  $I_h$  image, respectively.

The last step involves image contrast enhancement which is done with the one-scale local spatial filter, as proposed by Cvetkovic<sup>(6)</sup> et al. Each pixel is processed as follows

$$i_o = \hat{i}_h + \left( \hat{i}_h - m(\hat{i}_h) \right), \quad (9)$$

where the mean value  $m$  around a pixel  $\hat{i}_h$ , in the window controlled by two parameters  $s_w$  and  $t_w$ , is computed as follows

$$m(\hat{i}_h(p, q)) = \frac{1}{(2s_w + 1)(2t_w + 1)} \sum_{s=-s_w}^{s_w} \sum_{t=-t_w}^{t_w} \hat{i}_h(p + s, q + t). \quad (10)$$

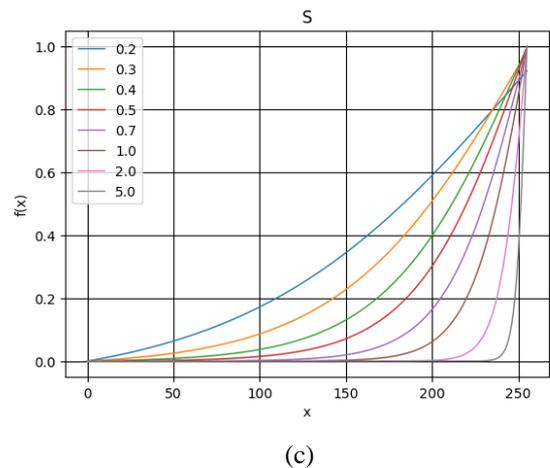
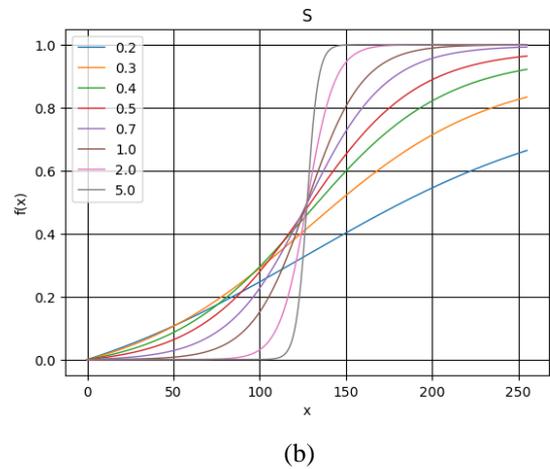
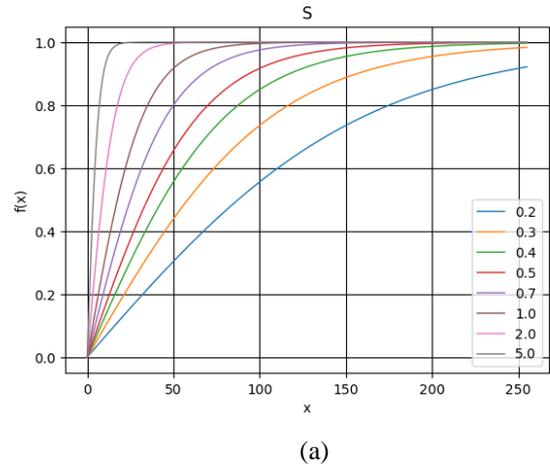


Fig. 2. Three versions of the sigmoid function used to generate virtual exposures. On the abscissa – the input brightness. On the ordinate – the output brightness. Convex part (a), full S-curve (b), concave part (c).

In our experiments the parameters were chosen as  $s_w=2$  and  $t_w=2$ , which boils down into  $3 \times 3$  local windows.

When processing color images, each color channel is processed separately in accordance with the aforementioned method.

### 3. Experimental Results

The presented method was implemented in C++, using the Microsoft Visual 2017 IDE. Experiments were run on a computer equipped with the Intel® Xeon E-1545 CPU @2.9GHz processor, 64GB RAM, and 64-bit Windows 10 OS.

Fig. 3(a) shows a monochrome frame from an underwater video recorded by our drone. Its VHDR version obtained with the presented here algorithm is visible in Fig. 3(b). In the latter a number of details is discernable which were hidden in the original frame. Such level of details can improve further processing, as will be discussed.



(a)



(b)

Fig. 3. Underwater monochrome frame (a) and its VHDR version (b).

Fig. 4 depicts color frames from the underwater sequence. As alluded to previously, in the case of multi-channel images, each color channel is processed separately by the VHDR

method. VHDR processed frames are shown in the right column of Fig. 4. Again, more details are easily perceived but also the whole image is brightened and with overall better contrast. Since each channels is processed separately, the color gamut also has changed. However, in this case such a change further improves image perception since red color, which was missed due to water filtering properties, is recovered.

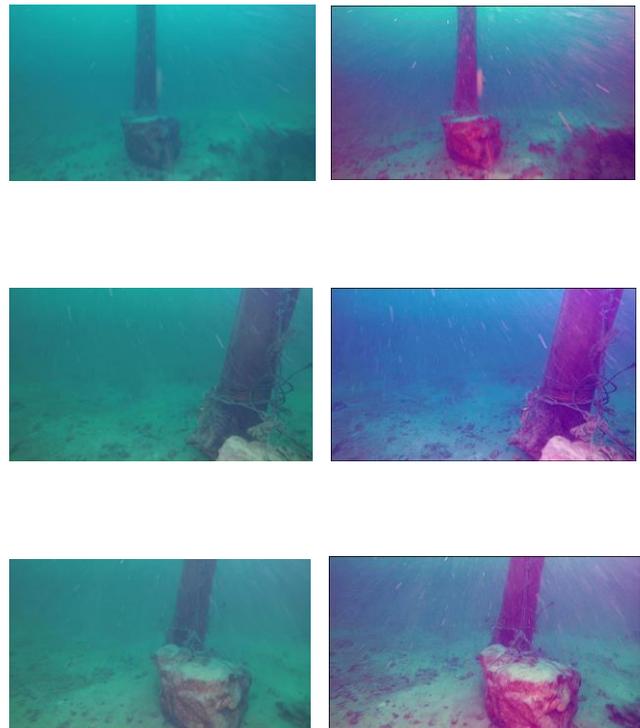


Fig. 4. Frames from an underwater video – original (left column), VHDR processed (right). Right images show enhanced details, e.g. marine snow.

A qualitative evaluation of the VHDR method is difficult due to lack of ground truth data, especially for the underwater conditions. Thus, to assess its performance in the background subtraction task<sup>(11)(15)(16)(19)</sup>, an annotated *Wallflower*<sup>(20)(12)</sup> dataset is used. In this experiment we measured influence of VHDR preprocessing of the input images on accuracy of the output background subtraction method obtained with the PCA based algorithm, originally proposed by Oliver<sup>(15)</sup> with implementation from Cyganek<sup>(7)</sup>. Results are presented in Table 1. For almost all test sequences, preprocessing with the proposed VHDR method leads to accuracy higher by 2-3%. Further research will be devoted to measure some classification methods, as well as to build an annotated database for underwater sequences.

Table 1. *F* value measured on the Wallflower dataset with ground-truth data provided.

Test sequence	F - Original	F - VHDR
<i>WaivingTrees</i>	0.80	<b>0.83</b>
<i>Bootstrap</i>	0.59	<b>0.61</b>
<i>Camouflage</i>	0.92	0.92
<i>ForegroundAperture</i>	0.63	<b>0.65</b>

Considering processing speed, the proposed method in serial implementation reaches real-time conditions, i.e. 30 frames of 640x480 pixels can be processed in 1s.

#### 4. Conclusions

In this paper a method for image enhancement is proposed. The method computes a high dynamic range image out from a single low dynamic range one. In the proposed method, first a series of tone mapped versions of the original image is generated. These are then fused by the weighted summation which aims at boosting middle level values of pixels. Further on, the fused image is range extended and finally contrast enhanced with a local high-pass filter. Although the method was developed for detail enhancement of underwater images, it can be used in other image processing applications requiring detail enhancement. As shown, image preprocessing with the VHDR filter leads to enhancement of image details which can be useful in further image processing and inferring levels. The method in our serial C++ implementation allows full real-time operation. However, it can be easily improved and parallelized.

#### Acknowledgment

This work was supported by the National Science Centre, Poland, under the grant no. 2016/21/B/ST6/01461.

This work was supported by the statutory fund of the Department of Systems and Computer Networks, Faculty of Electronics, Wrocław University of Science and Technology.

#### References

(1) Becker, S., Candes, E., Grant, M.: TFOCS: Flexible first-order methods for rank minimization, Low-rank Matrix Optimization Symposium, SIAM Conf. on Optimization, 2011.

(2) Benezeth Y., Jodoin P.-M., Emile B., Laurent H.,

Rosenberger C.: Comparative study of background subtraction algorithms. *SPIE Journal of Electronic Imaging*, 19(3), 2010.

(3) Bingham E., Hyvärinen A.: A Fast Fixed-Point Algorithm For Independent Component Analysis of Complex Valued Signals. *International Journal of Neural Systems*, Vol. 10, No. 1, World Scientific Publishing Company, 2000.

(4) Bouwmans T.: Traditional and Recent Approaches in Background Modeling for Foreground Detection: An Overview. *Computer Science Review*, 2014.

(5) Chang D.C., Wu W.R.: Image contrast enhancement based on a histogram transformation of local standard deviation. *IEEE Trans. MI*, vol. 17, no. 4, pp. 518-531, 1998.

(6) Cvetkovic S.D., Schirris J., de With P.H.N.: Locally-Adaptive Image Contrast Enhancement without Noise and Ringing Artifacts. *IEEE International Conference on Image Processing*, vol. 3, pp. 551-560, 2007.

(7) Cyganek, B., *Object Detection and Recognition in Digital Images: Theory and Practice*, Wiley, 2013.

(8) Cyganek B., Recognition of road signs with mixture of neural networks and arbitration modules, 3rd International Symposium on Neural Networks (ISNN), Springer, Lecture Notes in Computer Science, Vol. 3973, pp. 52-57, 2006.

(9) Cyganek B.: Soft system for road sign detection. *IFSA World Congress Analysis and Design of Intelligent Systems Using Soft Computing Techniques*. Springer Advances In Soft Computing, Vol. 41, pp. 316-326, 2007.

(10) Debevec P.E., Malik J.: Recovering high dynamic range radiance maps from photographs. *Proceedings of the 24th annual conference on Computer graphics and interactive techniques (SIGGRAPH)*, pp. 369-378, 1997.

(11) Guyon C., Bouwmans T., Zahzah E.: Robust Principal Component Analysis for Background Subtraction: Systematic Evaluation and Comparative Analysis. *Principal Component Analysis*, Edited by Parinya Sanguansat, InTech, 2012.

(12) <http://research.microsoft.com/en-us/um/people/jckrumm/wallflower/testimages.htm>

(13) Ke W.-M., Wang T.-H., Chiu C.-T.: Hardware-efficient virtual high dynamic range image reproduction. *Proceedings of the 16th IEEE International Conference on Image Processing (ICIP'09)*. Piscataway, NJ, USA:

IEEE Press, pp. 2665–2668, 2009.

- (14) Lapray P.-J., Heyrman B., Rosse M., Ginhac D.: Smart camera design for real-time high dynamic range imaging. Distributed Smart Cameras (ICDSC), Fifth ACM/IEEE International Conference, pp. 1-7, 2011.
- (15) Oliver N. M., Rosario B., Pentland A. P.: A Bayesian Computer Vision System for Modeling Human Interactions. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 8, pp. 831–843, 2000.
- (16) Piccardi M.: Background subtraction techniques: a review. IEEE International Conference on Systems, Man and Cybernetics, Vol. 4, pp. 3099–3104, 2004.
- (17) Robertson M.A, Borman S., Stevenson R.L.: Dynamic Range Improvement through Multiple Exposures. International Conference on Image Processing ICIP 99. Vol. 3, pp.159–163, 1999.
- (18) Sen P., Aguerrebere C.: Practical High Dynamic Range Imaging of Everyday Scenes. IEEE Signal Processing Magazine, pp. 36-44, 2016.
- (19) Stauffer C., Grimson W. E. L.: Adaptive background mixture models for real-time tracking. Proc. Int. Conf. on Computer Vision and Pattern Recognition, Vol. 2, IEEE, Piscataway, NJ, 1999.
- (20) Toyama K., Krumm J., Brumitt B., Meyers B.: Wallflower: Principles and Practice of Background Maintenance. Seventh International Conference on Computer Vision, September 1999, Kerkyra, Greece, pp. 255–261, IEEE Computer Society Press, 1999.