REAL TIME VISIBILITY IMPROVEMENT FOR UNDERWATER VIDEO

SANAT UPADHYAY
LOLAARK LLC

NIKOLAOS MITSAKOS
DEPARTMENT OF MATHEMATICS, UNIVERSITY OF HOUSTON

MANOS PAPADAKIS
DEPARTMENT OF MATHEMATICS, UNIVERSITY OF HOUSTON

Proceedings of the 23rd Offshore Symposium, February 14th 2018, Houston, Texas
Texas Section of the Society of Naval Architects and Marine Engineers

ABSTRACT

Water turbidity is a frequent impediment for achieving satisfactory imaging clarity in underwater video and inhibits the extraction of information concerning the condition of submerged structures. Ports, rivers, lakes and inland waterways are notoriously difficult spots for camera inspections due to poor visibility. This problem motivated us to study methods to extract a cleaner image /video from the one acquired in an almost real-time setting (delay of the order of 6-7 secs). This type of problem arises in image post-processing as an illumination neutralization problem and, it can also be viewed as a blind deconvolution problem. We present a method which enables the derivation of a cleaner image from a poor visibility original by means of a combination of linear and non-linear deterministic mathematical transformations for illumination neutralization implementable in almost real-time on GPUs. Real time visibility improvement for marine and water environments is a suite of algorithms aiming to restore the visibility of images and videos acquired under the surface of every water body utilizing our illumination neutralization method. We are currently transitioning from an academic algorithmic suite to a product we will call ΑΛΣvision (ALSvsion) from the Homerian Greek word ΑΛΣ which means sea. Upon completion of the project, the GUI will enable the view of the original camera feed and of the visibility improved video side-by-side. This software is not intended to replace the original video feed, but to offer guidance for interpreting it. ΑΛΣvision can be used for a variety of visual inspections in marine and offshore industries. The method we present also works for visibility restoration in still images or videos acquired in poor atmospheric conditions such as fog, haze, smoke or with insufficient illumination. One of our main contributions is the development of a mathematical theory which enables the derivation of results showing that lines and textures significant for the identification of structures and of structural problems are made visible as they appear in an image or video acquired in good imaging conditions.

Keywords: Underwater video, water turbidity, illumination, singularity enhancement, visibility improvement, compression

1 INTRODUCTION, PROBLEM STATEMENT AND, STATE-OF-THE-ART

Water with or without suspended solid particles are mediums affecting light reflected back to the camera or directly sent to it from the light source. This is also true for diluted particles, such as salt. Therefore, various degrees of water clarity affects image and footage visibility. In this paper we present a new approach for removing or mitigating the effects of poor footage visibility, even in high water turbidity.

Our starting point, is to treat this problem, as a poor illumination scene problem modeled by the Lambertian principle of illumination:

$$Image(f) = luminance(\mathcal{S}) \times Reflectance(R). \tag{1}$$

The model postulates that Luminance, $\mathcal{S}$, is the function modeling the energy and spectrum of light rays carrying the structural information of objects in a scene. These light rays are either reflected (back-scattered) due to interaction with the surface of an object in a scene, or emitted directly by a light source present in the scene captured in still photo or video f which is what we see. Luminance is not scene illumination but it depends on it. All structural properties of objects in a scene, namely edges, curvature of surfaces, the structural texture of a surface together with its color and consistency determine its natural abilities to backscatter light. The percentage of light energy a surface can scatter back is called albedo in physics. When the light from an object is captured by a camera, there are two more factors affecting what the camera sees, the angle of incidence and the relative angle of the light trajectory backscattered from the object with respect to the camera axis. These two angles and the object's albedo determine what we call Reflectance R. This discussion indicates that R contains all of the geometric and structural information of scene objects and of their relative placement in the scene (foregroung vs background). Consequently, what we want to see in an image is R, but, instead, of R we see f. When illumination uniformly and adequately floods the scene (e.g. flood light) and the medium through which rays travel is vacuum or air with no suspended particles, we have*ideal illumination* or $\mathcal{S}$ constant at every point of the scene, which we conventionally set $\mathcal{S} = 1$ everywhere. In this case, the amount of light energy reaching the camera depends only on the surface properties of the object, incident and observation angles and is not be affected by the distance of the object from the camera plane (as long as this distance is not big). Here, we assume that the distance is such that does not exceed the distance within which the camera can generate a focused image. These remarks imply that in underwater imaging, $\mathcal{S}$ is not constant. Also, we may say that, *poor visibility is a condition where the viewer of f cannot infer the structural information R carries.*

Extracting R, when it is not constant, becomes as ill-posed of a problem as solving the equation $10 = xy$, with $x, y > 0$. However, if someone gives us some idea about the applicable range of x we can find all suitable y in the previous equation. A similar line of thought is adopted here for extracting an approximation of Reflectance containing all pertinent information from Eq. (1). However, this task is not easy. *A commonly used assumption is that there exists a linear basis of functions with respect to which the expansion of $\mathcal{S}$ is sparse, while R has different sparsity properties, if represented by the same basis.* Leveraging the different sparsity properties of $\mathcal{S}$ and R we develop a new method and a new mathematical line of thought to address the problem of extracting this informative approximation of R given f.

Only the key ideas of this method are presented in this paper in a format suitable for a technically educated yet, not mathematically savy audience. Our method is numerically efficient, faster than other state of the art best performers. Also, our approach is general enough to work even when significant turbidity degrades the visibility of underwater images and footage. We demonstrate, that even in these conditions, our method can be used to improve the visibility of underwater images and footage to the point that a lot of useful structural information can be inferred with the side by side analysis of original and improved footage. We also give indications why our method is suitable for real-time visibility improvement of such footage.

Water affects light propagation due to scattering, which is increased by suspended solid particles in the water. Light absorption is one way this effect manifests. Turbidity is an optical determination of the level of water clarity. It is caused by a variety of non-soluble particles, e.g clay, sand silt or algae. Suspended particles are those that have diameter larger than $2\mu m$. Anything smaller is considered soluble. Water turbidity is the amount of light scattered by suspended particles in water. Its opposite, water clarity is measured by means of Secchi disks. In this case we measure the depth at which such a disk is no longer visible when

submerged without the use of an external light source but in broad daylight and with no additional light sources. For a more detailed discussion of water turbidity the reader may refer to (Kemker 2014, Anderson and Davie 2004). Notably, marine environments have lower turbidity levels than freshwater sources. In fact, the higher the salinity is, the better clarity becomes (Kemker 2014). The theory we develop establishes that, under certain assumptions for luminance, we can generate an illumination neutral image which contains all useful information. Specializing this to underwater video our methods should lead to a neutralization of the effects of turbidity on $\mathcal{S}$, and, thus to restoration of visibility. Whether the general assumptions we make for luminance are satisfied by the turbidity modified luminance is not yet known to us. We need experimental verification of this fact which is part of our future research plans. In this paper we provide preliminary evidence showing that our method for illumination neutralization is able to restore visibility in underwater images and video affected by turbidity. This is an indirect proof that turbidity modified luminance obeys the general assumptions we make for luminance.

Our key contributions are:

1. The proposed approach for illumination neutralization is deterministic, so it is a robustly fast preprocessing unlighting tool.

2. We also build a mathematical framework for illumination neutralization enabling us to show rigorously that illumination is neutralized while desired information content is preserved.

Several attempts and a lot of analysis has been put into this problem by numerous groups, it still remains unsolved. The greatest need is for algorithms capable of neutralizing illumination effects in outdoor and other uncontrollable conditions with real-time speed. Many groups talk about illumination normalization or albedo extraction. First we avoid the term albedo extraction, because, no one has proposed a concrete and rigorous mathematical definition of albedo. This term is used to refer to the reflectivity properties of surfaces. The use of the term normalization leaves open a gap; With respect to what do we normalize illumination? These ambiguities in the terminology motivate us to coin a new term, illumination neutralization implying that we attempt to extract from an acquired image another one which is illumination or visibility conditions neutral. This will be made mathematically concrete in the statement of Definition 2.1.1. Illumination neutralization algorithms fit in two categories (Ochoa-Villegas et al. 2015): Those that treat the illumination problem as a *relighting* one and others which treat it as an *unlighting* one. Relighting methods improve similarity of illumination conditions between gallery and probe images (e.g. (Li et al. 2009, Shim et al. 2008, Guan et al. 2012, Shashua and Riklin-Raviv 2001, Shao et al. 2010)). Unlighting methods (e.g. (Kimmel et al. 2003, Wang et al. 2014, Guo et al. 2013) utilize Lambertian reflectivity, Eq. (1). The main difference between the two approaches is that unlighting uses only a single input image, while relighting uses multiple images of the same scene acquired under non-identical illuminations. Our Illumination Neutralization (IN) fits in the latter category. Most notably, Retinex theory (Land 1964, Land et al. 1977) is an attempt to explain color perception by the human visual system lead to a generation of unlighting algorithms. Here we list the most recent work which unifies most of the previous Retinex approaches (Zosso et al. 2015). Finally, we do not specifically mention contrast enhancement methods here, because these do not attempt to extract R from f. There are several popular methods of this sort, such as histogram equalization. However, contrast enhancement provides a better way to view f by adjusting the range of pixel intensity values when visibility conditions compromise an image. We will revisit this method in Section 3. The next section is somewhat theoretical and we expect to be less interesting for the majority of the conference's audience. The reader can skip what follows after Eq. (5) and look at the Fig. 1 to get an feeling of how the proposed method works. The main results of this section are summarized in the Conclusions Section.

2 METHODS: MULTISCALE ILLUMINATION NEUTRALIZATION

2.1 Illumination variation and scene structure

As previously stated, our illumination model is Eq. (1). Obviously R is a scene invariant, because, by default carries all necessary information, edges, curvature of surfaces and, textures. Deriving R, though is an ill-posed problem, which we circumvent by proposing to obtain a surrogate of R which carries the information

content of R which is necessary to uniquely identify a scene. This type of information can be problem or application specific. Typically edges, curvature of surfaces and, textures are the most common elements the surrogate of R should contain. Moreover, this surrogate image should not vary with illumination. This leads us to the next definition:

Definition 2.1.1. *Illumination neutralization is a process that allows to extract from an image f a function r which is representative of the scene depicted in the image f, in the following sense: For a given scale, or spatial resolution, which is identical to those of f, two different scenes cannot be represented by the same r. Moreover, for a given scene the corresponding r should remain invariant under all possible luminance configurations.*

The definition and the preceding discussion suggest that we can have surrogates of R which are not identical to R. Nevertheless, a high fidelity approximation of reflectance can serve as a surrogate of R. When illumination is ideal $\mathcal{S} = 1$ and $f = R$. This suggests that surrogates should be, illumination invariant and close approximations of an f of a given scene acquired in ideal illumination or visibility conditions. This relaxation of the sought solution allows to solve the problem of illumination neutralization without having to deal with the fact that it is ill-posed. Below, we summarize the discussion that shows that this type of solution is the proposed illumination neutralization operator. The rest of this section is a summary of S. Upadhyay's thesis (Upadhyay 2014) or in (Upadhyay and Papadakis 2017).

We assume that all acquired images $f = \mathcal{S}R$ belong to a compact subset $\Omega \subset \mathbb{R}^2$ with non-empty interior, $0 < \gamma \leq R(x) \leq \Gamma < \infty$, and $0 < \tau \leq \mathcal{S}(x) \leq 1$ for every $x \in \Omega$. The assumptions are physically meaningful because they ensure that there are no regions in the scene with very small reflectance of light nor do we have locations entirely dark (no photons back-scattered from them).

2.1.1 Modeling Illumination variation

To formalize the statement that luminance lives mainly in coarse scales and varies slowly we use *Campanato Spaces*.

Definition 2.1.2. *Let Ω be a compact subset of $\mathbb{R}^n$, $1 \leq p < \infty$ and $\eta \geq 0$. A function g is said to be in* Campanato space $\mathscr{L}_{p,\eta}(\Omega)$*, if there exists a constant $\mu > 0$ such that for every n-dimensional cube $Q \subseteq \Omega$,*

$$\int_Q |g(t) - g_Q|^p \, dt \leq \mu^p \, |Q|^\eta \tag{2}$$

where $g_Q = \frac{1}{|Q|} \int_Q g(t)dt$

If μ is small the previous definition indicates that, in average, the values of g in a small image patch point are close to the mean value of g (the p-mean oscillation of g) in this patch and this remains true as the patch becomes smaller and smaller. The definition also suggests that this property remains true regardless of the location of the patch. For instance, uniformly continuous functions g would satisfy this property. Empirically, we have seen though that this formalism works for modeling visibility induced $\mathcal{S}$ in underwater images and footage. However, further experimental investigation is warranted to prove that turbidity-induced luminance lives in Campanato spaces and which are the parameters η and p determining those spaces.

We observe that when $(\eta - 1)/p > 0$, then the p-mean oscillation of g (the left-hand side in (2)) decays to zero as $|Q| \to 0$. Therefore, $(\eta - 1)/p$ reflects the rate of decay of p-mean oscillations as the scale increases. The minimum value of μ satisfying (2) is a semi-norm for the Campanato space $\mathscr{L}_{p,\eta}(\Omega)$ denoted by $\|\cdot\|_{\mathscr{L}_{p,\eta}(\Omega)}$. Semi norms generalize the concept of length of a vector. The semi-norm $\|g\|_{\mathscr{L}_{p,\eta}(\Omega)}$ for a function is small if the variations in g are locally small. Utilizing these properties we model illumination variation by a function $\mathcal{S} : \Omega \longrightarrow [\tau, 1]$ with $\tau > 0$ in a suitable Campanato space $\mathscr{L}_{p,\eta}(\Omega)$ with sufficiently small semi-norm and positive $(\eta - 1)/p$. Slowly varying luminance regardless of how poor visibility is, still belongs to a Campanato space identified by a high value of p.

2.1.2 Modeling scene structures

The structure of a scene of interest contains singular and smooth patches. The former have significant high frequency content and can either be jump discontinuities or continuous functions with high frequency oscillations or discontinuous higher order partial derivatives. To generate a faithful surrogate image of a scene from f, we must preserve this local oscillatory behavior. In practical terms, locally, all around an image we want to recover the degree of "sharpness" of various edges describing structures in the imaged scene. To be able to assess that we do so, we incorporate in our discussion an ensemble of singularity descriptors, obtained from different microlocal spaces measuring the local singular content at points of continuity of f. Microlocal spaces model mathematically local oscillations associated with edges, texture and other useful image content information. This is why microlocal analysis becomes relevant.

Definition 2.1.3. *Let $x_0 \in \Omega$ and $\alpha > 0$. A* locally bounded function $g : \Omega \to \mathbb{R}$ *belongs to $C^\alpha(x_0)$ if there exist a constant $\mu > 0$ and a polynomial P satisfying $deg(P) < \alpha$ and such that in a neighborhood of x_0 ,*

$$|g(x) - P(x - x_0)| \leq \mu \, |x - x_0|^\alpha . \tag{3}$$

The ***Hölder exponent (Jaffard et al. 2007)*** *of g at x_0 is,*

$$h_g(x_0) = \sup\{\alpha : g \in C^\alpha(x_0)\} \tag{4}$$

Inequality (3) mimics in non-smooth functions what Taylor polynomial approximations can do in analytic functions. This is why we have the presence of the polynomial term P in the LHS of this inequality. For example, for $g(x) = |x|$ at $x_0 = 0$ we can only choose $P(x) = 0$ and $\alpha = 1$. any value of $\alpha < 1$ will work as well, but we cannot choose an $\alpha > 1$. The wedge of the absolute value at the origin limits its smoothness at this point. If you pick another point you can choose $\alpha \in [1, 2]$ but no greater than 2 because you need to choose P to be a linear function. Keep in mind that (3) is only satisfied locally around x_0. Edges are lines of points of non-smoothness, that is approximating P will have low degree and α will be small, as in the wedge example of $g(x) = |x|$. Picture $g(x, y) = |x|$, with $-1 \leq x \leq 1$ and y in any interval. In local Hölder spaces we can characterize the sharpness or regularity at points x_0, that is the Hölder exponent of g at a point x_0, based on the asymptotic behavior of wavelet coefficients of g in a neighborhood of x_0. If we denote by $L_j^g(x_0)$ the "wavelet leaders" at x_0 (Jaffard et al. 2007), which is the largest magnitude of wavelet coefficients of g in the first order neighborhood of x_0 at scale j, then the *Hölder exponent* of g at x_0 is given by (Abry et al. 2014),

$$g_f(x_0) = \liminf_{j \to +\infty} \left(\frac{\log \left(L_j^g(x_0) \right)}{\log \left(2^{-j} \right)} \right) \tag{5}$$

Consequently, if the asymptotic rate of decay of wavelet coefficients between two functions at a point x_0 is identical, then the two functions have the same Hölder "singularity" exponent at x_0. This understanding is one of the pivotal motivations in formulating the $\mathcal{IN}$ operator as described in Section 2.2. The definition of $\mathcal{IN}$ ensures that this asymptotic behavior is inherited by the surrogate image $\mathcal{IN}(f)$.

In addition to characterizing structures locally, it is also important to characterize their topological organization along curves in the domain of the image. We can study this property by certain multifractal properties of the set containing identical singularity exponents.

Definition 2.1.4. ***Iso-Hölder set and Singularity (or multi-fractal) spectrum (Jaffard et al. 2007):*** *Let g be a locally bounded function, and let $E_g(H)$ denote the iso-Hölder set, which is the set of points where the Hölder exponent of g takes the value H. The spectrum of singularities of g (denoted by $d_g(H)$) is the Hausdorff dimension (Triebel 1997) of $E_g(H)$.*

Preserving Iso-Hölder set and singularity spectrum ensures that the spatial organization of singularities, described by the Hölder exponent, are also maintained along with their pointwise behavior.

It is important to note that a single singularity descriptor does not suffice to characterize all forms of local oscillatory behavior, hence of local structures. For example, local Hölder spaces contain both chirps (Guilheneuf et al. 1998) and cusps. In this case, having a common Hölder exponent between two points does

not guarantee the presence of the same type of structure at both locations. But, we can conclude that both singularity sharpness and structure type are known to be identical when the chirp exponent (Abry et al. 2011) is also preserved in addition to the Hölder exponent. This however is true only if the set of possible structures consists only of chirps and cusps. At any rate, more microlocal descriptors are needed but we are not going to carry on this tedious discussion here. We have not fully identified all necessary descriptors to fully characterize a scene and we do not anticipate that a program of this sort is feasible.

What is mathematically a satisfactory surrogate image of a scene? This question can have various answers, but all boils down to our ability to classify without ambiguity the different classes of objects we anticipate to have in a scene. Of course, there may be structures that we have never encountered before. Objects or structures in this class should not be confused with those in other classes within the realm of our knowledge. For example, with a few strokes of a pen an artist can sketch the face of a person which we can unambiguously identify. If this type of image, is sufficient for facial identification of individuals, then sketches can be considered as sufficient surrogate images for faces. This example illustrates that the desirable properties of the surrogate image depend on the type of information we want to retrieve from an image. Nevertheless, what is important and can be considered as a minimal requirement for a surrogate image are boundary lines of objects, textures and, more generally, lines associated with local changes of surface smoothness. For instance lines associated with sharp gradients are boundaries of an object or signify other local changes of the continuity properties (as quantified by Hölder exponents or other microlocal properties) of the surface resulting from structural changes (e.g. a crack on a metal surface or a welded seam). Within this context, as a minimum, a surrogate image must preserve iso-Hölder sets and the singularity (or multi-fractal) spectrum resulting from objects present in a scene. As we will argue in the next subsection the proposed illumination normalization operator preserves iso-Hölder sets and the singularity spectrum of reflectance R regardless of luminance $\mathcal{S}$. Therefore, $\mathcal{IN}$ generates, or at least, has the hope to generate, surrogate images of scenes.

2.2 Illumination Normalization operator and associated properties

Let $\Omega \subset \mathbb{R}^2$ be a compact domain with non-empty interior for an Image $f = \mathcal{S}R$. The variation in illumination $\sigma \in \mathscr{L}_{\zeta,\eta}(\Omega)$ for some $1 \leq \zeta < \infty$. Also, as before, $0 < \gamma \leq f(x) \leq \Gamma < \infty$ and $0 < \tau \leq \sigma(x) \leq 1$ for some positive values γ, Γ and τ.

Given φ, ψ_i to be a compactly supported scaling function and its associated wavelets used for analysis and $\tilde{\varphi}$ and $\tilde{\psi}_i$ respectively be the dual family of scaling function and wavelets used for synthesis, The *Illumination Normalization operator* (Upadhyay and Papadakis 2017) $\mathcal{IN}(\cdot;\zeta) : L^2(\Omega) \longrightarrow L^2(\Omega)$ is defined by

$$\mathcal{IN}(f;\zeta) = \sum_{j=0}^{\infty} \sum_{k \in K_j} \sum_{i=1}^{3} \mathcal{A}_{j,k}(f,\zeta)(\langle f, \psi_{i,j,k}\rangle)\tilde{\psi}_{i,j,k} + \sum_{k \in K} \mathcal{A}_{0,k}(f,\zeta)(\langle f, \varphi_{0,k}\rangle)\tilde{\varphi}_{0,k} \tag{6}$$

where, the wavelet indices j, k determine the scale j of the decomposition/reconstruction and k determines the pixel at which the wavelet or the scaling function are centered at the scale j decomposition/reconstruction step. Inner products are denoted by $\langle ; \rangle$ and the wavelet coefficients of f at scale j and pixel k are the inner products $\langle f, \psi_{i,j,k}\rangle$. The tildes over the wavelets and scaling function underscore that we may use a different set of wavelets and scaling function at the reconstruction stage. Moreover, $\{\mathcal{A}_{j,k}\}_{j,k\geq 0}$ is a sequence of normalizers, with the property that $|\mathcal{A}_{j,k}(f,\zeta)\langle f, \psi_{i,j,k}\rangle|$ has the same asymptotic decay as $|\langle f, \psi_{i,j,k}\rangle|$ when $j \to \infty$, locally around every point $x_0 \in \Omega$ as long as the point is not close to the boundary of the image.

The operator $\mathcal{IN}$, as defined above, is proved to have several useful properties:

Theorem 2.2.5. *(Upadhyay 2014, Upadhyay and Papadakis 2017) Let $R \in \mathbf{L}^\infty(\Omega)$; $\mathcal{S} : \Omega \to [\tau, 1]$ $(\tau > 0)$ be a function in $\mathscr{L}_{\zeta,\eta}(\Omega)$ with $(\eta - 1)/\zeta > 0$, and $0 < \gamma \leq R(x) \leq \Gamma < +\infty$. Then for a suitable choice of normalizer sequence $\{\mathcal{A}_{j,k}\}_{j,k\geq 0}$, there exists a constant $C(\psi)$ which depends only on the compactly supported wavelet basis ψ used to define $\mathcal{IN}(\cdot;\zeta)$, such that for any $k \in \mathbb{Z}^2$ for which $|\Lambda_{0,k} \cap \Omega^o| \neq 0$ we have,*

$$\|\mathcal{IN}(\mathcal{S}R;\zeta) - \mathcal{IN}(R;\zeta)\|_{\mathbf{L}^\infty(\Lambda_{0,k})} \leq C(\psi) \cdot \|\mathcal{S}\|_{\mathscr{L}_{\zeta,\eta}(\Lambda_{0,k})} \|R\|^2_{\mathbf{L}^\infty(\Lambda_{0,k})}$$

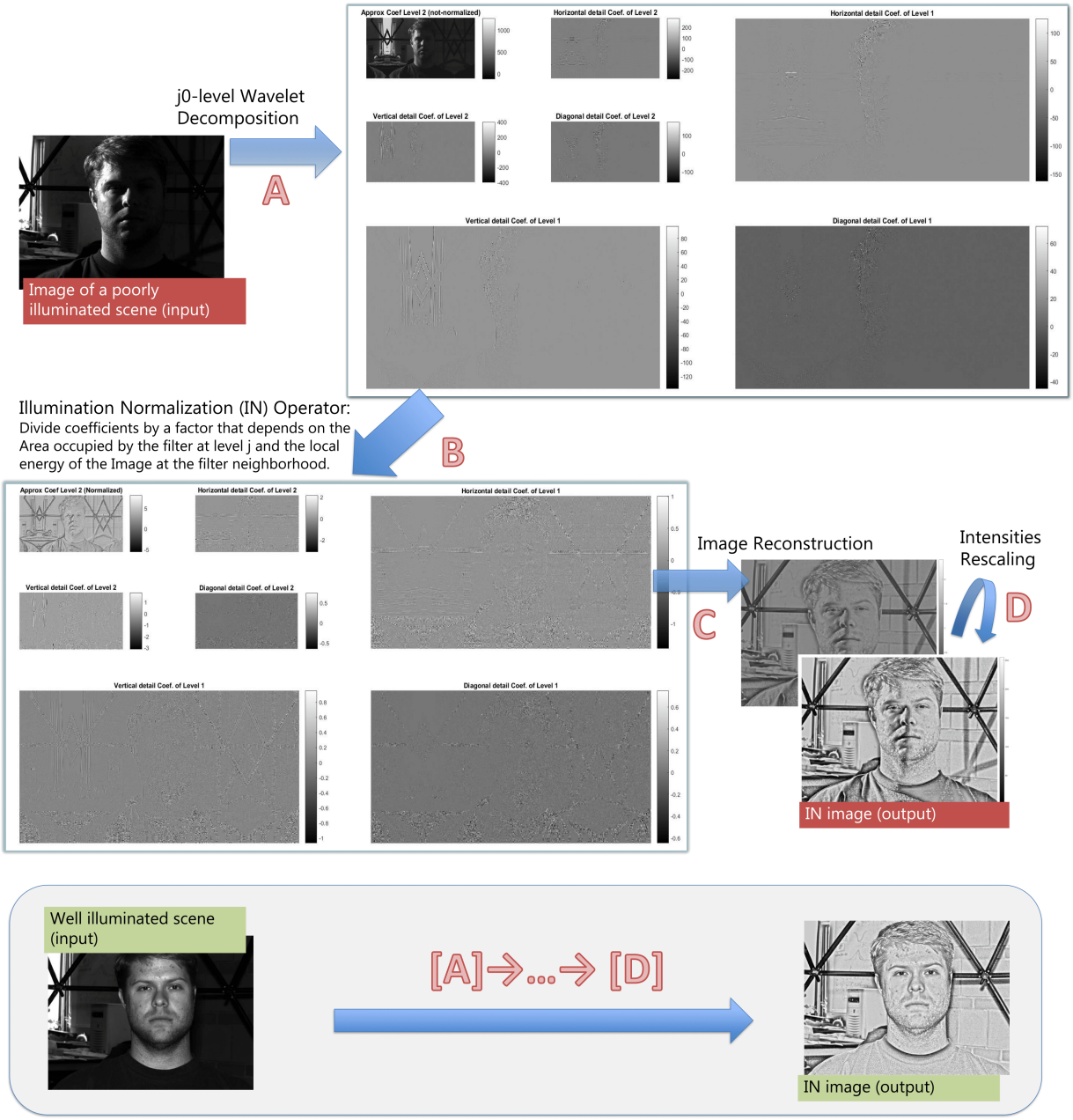


Figure 1: Schematic representation of algorithm implementation of proposed Illumination Neutralization by $A\Lambda\Sigma$vision. The number of wavelet decomposition levels J_0 and the Campanato space parameter ζ need to be tuned by the user. Part A generates the wavelet coefficients $\langle f, \psi_{i,j,k} \rangle$ and $\langle f, \varphi_{0,k} \rangle$. The gray scale images in A are formed by the values of the wavelet coefficients at scales j and k. In this example we use only two scales. Per scale we have three wavelets and one coarse scale filter induced by φ. Wavelet filters capture edges, hence the wavelet panels are dominated by a gray background of values close to zero. B implements the adjustments $\mathcal{A}_{j,k}(f,\zeta)(\langle f, \psi_{i,j,k} \rangle)$ and $\mathcal{A}_{0,k}(f,\zeta)(\langle f, \varphi_{0,k} \rangle)$. Now you see edges being more pronounced even those in dark regions of the original. At reconstruction C, we form the linear combination shown in the right-hand-side of Eq. (6). At this stage we have the surrogate image of the scene. Finally, in D we adjust the dynamic range of the surrogate image pixel intensities to please the eye of the beholder.

This theorem suggests that for illumination variation resulting in a small $\|\mathcal{S}\|_{\mathscr{L}_{\zeta,\eta}(\Lambda_{0,k})}$, we obtain

$$\|\mathcal{IN}(\mathcal{S}R;\zeta) - \mathcal{IN}(R;\zeta)\|_{\mathbf{L}^\infty(\Lambda_{0,k})} \approx 0.$$

It also predicts that regardless of how strong a shadow or turbidity are, as long as they only varies mildly on any image patch (an assumption consistent with the Lambertian principal), $\mathcal{IN}(\cdot;\zeta)$ essentially generates the same surrogate image depending only on the imaged scene and camera (or image acquisition) properties. *Thus, $\mathcal{IN}$ approximately satisfies the uniqueness property required by the definition of illumination neutralization.* But, does it generate informative surrogate images of any scene?

Nonwithstanding our remarks before closing the previous subsection, we are ready to address why $\mathcal{IN}$ generates a surrogate image. First, note that at each location k the non-linear adjustment operators $\mathcal{A}_{j,k}(f,\zeta)(\langle f,\psi_{i,j,k}\rangle)$ are defined in so that they maintain the asymptotic rate of decay of $\langle R,\psi_{i,j,k}\rangle$, using Eq. (5) it is not hard to infer that $\mathcal{IN}$ preserves the local Hölder exponents of R regardless of $\mathcal{S}$. On the other hand, if an edge is a line of jump discontinuities then, they are preserved in the form of a large but very localized oscillation at the point of discontinuity (Upadhyay and Papadakis 2017, Lemma 2). So to the extend that all edges correspond to edge discontinuities or to certain Iso-Hölder lines and everything else is smooth, then the surrogate image contains all pertinent information of R.

3 RESTORING VISIBILITY IN UNDERWATER IMAGES AND OTHER EXPERIMENTAL RESULTS

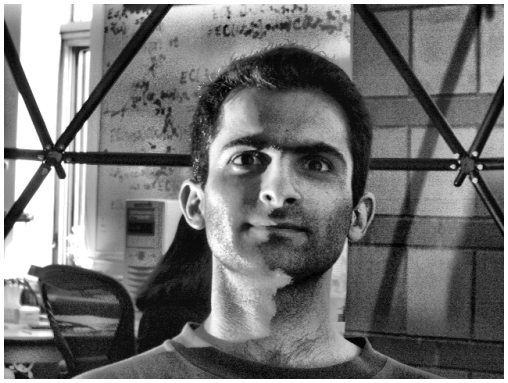
We already mentioned in the introduction that there is a deluge of techniques for illumination neutralization, normalization, as others call it, or global visibility enhancement. On the other hand, several existing techniques are either computationally inefficient or just have not established fidelity of visibility enhancement. By this term, we mean the faithfulness with which local patterns in the input image are preserved irrespectively of illumination. Moreover, marine environments pose more challenges as turbidity varies locally and globally. The problem with turbidity is not just a simple loss of contrast. Plain contrast enhancement will improve edge appearance to some extent but not overall, because turbidity changes the spectral properties of an image (in the sense of Fourier transform) and in a non-trivial way. Illumination neutralization methods are designed to remove to various degrees the effects of water turbidity on image visibility. Contrast enhancement methods will not remove the effects of turbidity on the image but they will make details to look more pronounced with the improvement mainly limited to objects in the foreground.

Moreover, for video enhancements for a real-time implementation computational efficiency and very small variability in processing time are of essence.

There are a few commercially available products for real-time underwater video visibility enhancement. All of them are equally computationally efficient and work in real-time but lack proper validation. This is something we plan to perform for our method in a systematic way and perhaps develop a set of validation videos for this small industry segment. Moreover, the creators of all other similar commercially available technologies do not describe their methodologies in any scientific publication, except for the Prohawk system (US Patent 9129406B2, of M. Kobayashi). Prohawk's main idea is local intensity histogram equalization. An improved version of this method is the Contrast-Limited Adaptive Histogram Equalization (CLAHE) (Zuiderveld 1994). This method is less computationally complex than ours, thus faster than ours, but qualitatively our method performs better than CLAHE. Fig. 2 provides a visual comparison of the visibility improvement results with our method and three other methods including CLAHE.

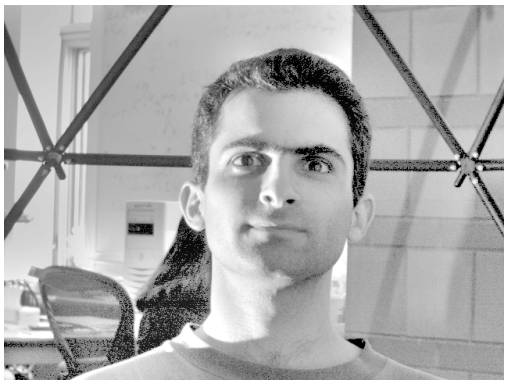
We observe that CLAHE increases the visibility of poorly illuminated regions of the image but trades uneven illumination with uneven illumination. This shows that CLAHE is not an illumination neutralization method. Primarily, it appears to function as a contrast enhancement tool and not necessarily as an illumination neutralization tool. Contrast enhancement will generate an image where edge-associated information will be more visible relative to neighboring structures. However, those enhancements are good locally and enhancement is not uniformly consistent in various regions of the same image. In video, these enhancements will result in non-consistent changes from frame to frame. With illumination neutralization we also have better contrast but in a uniform manner inter and intra frame because luminance is neutralized, something than contrast enhancement methods are not designed to perform.

(a) Raw Image



(b) CLAHE

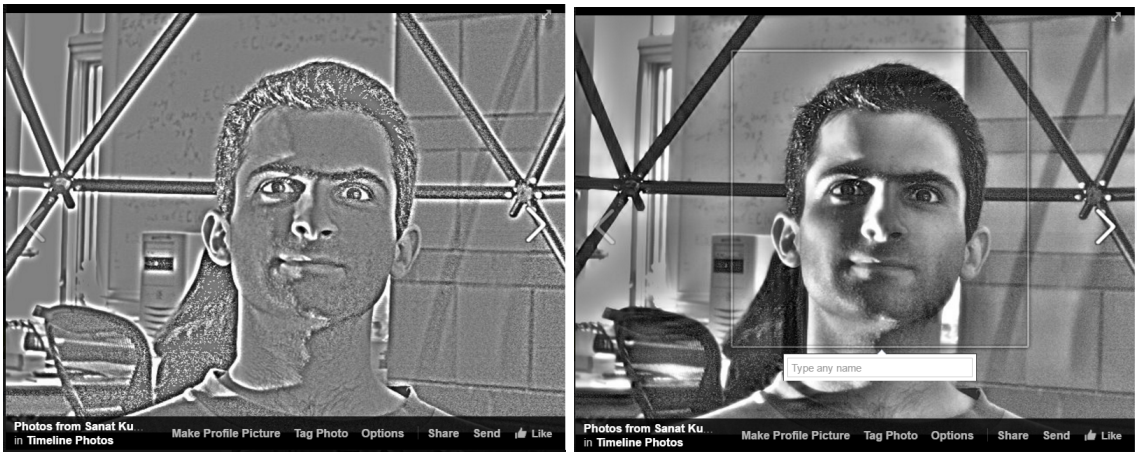
(c) Subtracting Gaussian blur



(d) Non-Local Retinex

(e) Proposed method

Figure 2: An example of some popular methods and our illumination neutralization on an image taken from Extended Yale B face database (Georghiades et al. 2001). CLAHE (Zuiderveld 1994) does not neutralize the illumination and only enhances contrast. Subtracting a Gaussian smoothed image from the raw image in the log-domain eliminates some of the medium spatial frequency structures. This method is more commonly known as Difference of Gaussians (DoG). There is also a colored version of it. Non local Retinex (Zosso et al. 2013, 2015) achieves illumination neutral output and retains all structure, but is computationally slow. Our method for illumination neutralization is both computationally fast and retains all structures. Compared to all other methods ours and Non local Retinex give the best results in the sense that both balance illumination and retain all structures



(a) Subtracting Gaussian Blur

(b) Proposed $\mathcal{IN}$ operator

Figure 3: We observe that simple methods like subtracting a low pass version of image, using a Gaussian filter Difference in the log domain, maintain sharp edges but do not enhance some medium frequency structures. On the other hand, our method preserves structures corresponding to all medium and high frequencies. As a result, the Facebook face detector fails to detect a face when illumination is neutralized using the log-Gaussian filter difference, but it is successful when the proposed method is used.

Also, CLAHE is volatile to noise. In underwater imaging noise may be significant and it will be amplified unless removed prior to the application of contrast enhancement. This may affect the clarity of edges after restoration. Another problem of this method is that one needs to tune several parameters. In the facial images of Figure 2 we used the default CLAHE parameters but we had to test several percentile values to trim outliers before we ended up with one descent image. Another classical off the self method is to subtract the Gaussian blur from the original image in the log domain. This idea goes back to the work of Marr and Hildreth on vision Marr and Hildreth (1980). It also neutralizes illumination across the image, but does not adequately maintain all local structures in the scene. On the lower left side of panel (a) in Figure 2 there is a coffee thermo-mug. This mug becomes more visible in panels (d) and (e) than in (c), perhaps, due to the trimming of mid-frequencies by the Gaussian density. The Non-Local Retinex achieves illumination neutralization and also maintain local structures, but at the cost of high computational cost as it involves an L^1 optimization. Here, L^1 underscores that what is minimized is the L^1-norm of a certain cost functional C, that the optimizer seeks to find the solution R which minimizes

$$\int_{\mathbb{R}^2} |C(R)(x,y)| dxdy \ .$$

We did not accidentally used R here. Approximations of reflectance are obtained by Non-Local Retinex using this L^1 minimization, but this method is slow, because it cannot be implemented with gradient descent. Nonetheless, compared with all other methods Non-Local Retinex gives the best visual results.

Our method achieves equally good results as Non-Local Retinex, but it is much faster and requires less parameter tuning than Non-Local Retinex. One of the strengths of our method is that any wavelet coefficient in the surrogate image is zero if and only if the corresponding wavelet coefficient in the raw image is zero. This property ensures that structures corresponding to medium and high spatial frequencies (in the sense of Fourier, one might also call them wavenumbers) are never lost in the process of neutralizing illumination or revert the effects of water turbidity (see Fig. 3). This is also another advantage of our method relative to CLAHE and Retinex: We apply different enhancements for details in an image corresponding to different spatial frequencies. The other two methods are essentially mono-scale, that is treat all spatial frequencies the same.

Among all other methods Non-Local Retinex gives the best visual results. However, this method uses

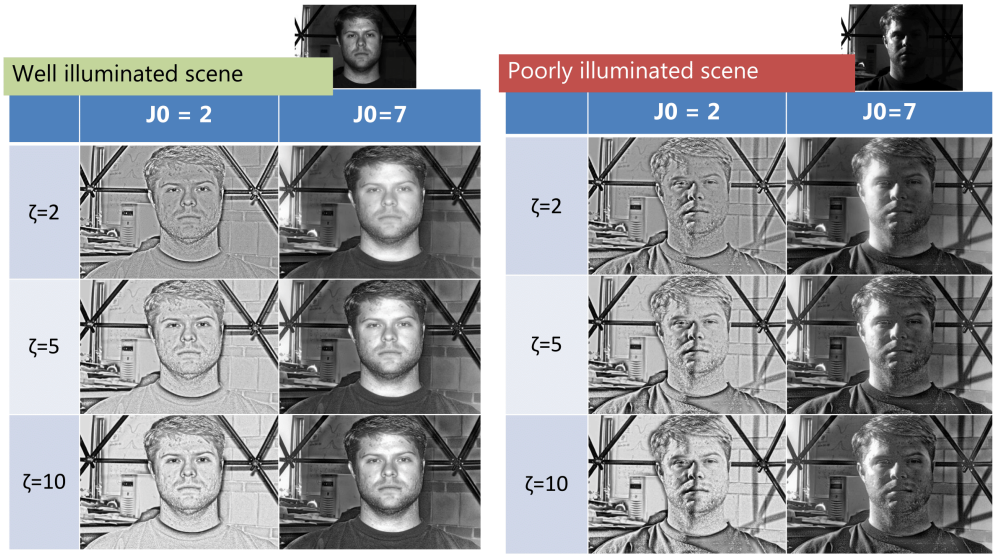


Figure 4: A higher value of ζ seems to work better for both illuminations, due to the high smoothness and low oscillation of $\mathcal{S}$ in both input images of the same scene. Note that both inputs are encoded with JPEG with lossless compression. These examples highlight how qualitatively close outputs can be despite the different illumination in the inputs. The best outputs need not be obtained with the same ζ setting, as expected by Theorem 2.2.5. Finally, notice that the shadow boundary on the nose of the subject remains unaffected in accordance to the theoretical prediction. Illumination is neutralized in the interior of shadows, but a shadow boundary line remains visible.

optimization,hence processing times can be long and vary significantly, rendering the usefulness of the method questionable for real-time underwater video visibility enhancement.

In Fig. 1 we give a schematic description of our algorithm. We use images from our facial detection experiments only for illustration purposes. The effect of the use of different parameter values for ζ and J_0 in the enhancement of significant image content can be seen in Fig. 4, where the boundary lines of facial characteristics are enhanced with varying emphasis. This observation begs the question of how to choose the best values for these two parameters. We believe that there is no universal answer to this problem. Both of these parameters are determined by the properties of $\mathcal{S}$, which we are not known and may vary based on the nature of $\mathcal{S}$: Optimal values for ζ and J_0 in underwater imaging are not expected to be the same for images of dry scenes with illumination affected by a shadow as in Fig. 4. We have not systematically studied yet how water turbidity affects the choices for ζ and J_0. However, we have observed that the best ranges of values for ζ and J_0 depend on the bandwidth of the camera and its focus. When the resolution is higher it is better to use bigger values for J_0.

The proposed method has not been implemented yet outside of a typical academic environment. Lolaark LLC[1], a partnership with the University of Houston, is currently working to develop a real-time visibility video enhancement software system available for commercial use by July 2018 under the brand name $A\Lambda\Sigma$vision (ALSvision) from the Homerian Greek word $A\Lambda\Sigma$ which means sea. With the use of a graphics processing unit (GPU) the required computations can be executed fast. We estimate that from the time the input video is acquired till to the time the video will be enhanced and shown to the user there will be a delay ranging from 5 to 10 seconds. The system will be deployed on a conventional laptop with an external graphics card unit and interface with input devices from the underwater camera feed. $A\Lambda\Sigma$vision is being designed to work with analog, AHD (those are HD cameras but output analog images) and SDI cameras (those output digital images). In Figs. 7 through 10 we showcase how $A\Lambda\Sigma$vision enhances underwater in-

[1]IP is exclusively licensed to Lolaark LLC by the University of Houston BOR.

spection footage acquired with a conventional analog camera with visibility at 1.5m, much less than the 5m suggested visibility requirement for a marine underwater inspection. This shows the potential of $A\Lambda\Sigma$vision for eliminating the need to move a vessel to clear waters for an underwater inspection. The same technology can be utilized for underwater inspections of deep sea off-shore facilities or in murky waters or wherever water visibility is an issue. The $A\Lambda\Sigma$vision GUI will be designed to allow the user to choose the values of ζ and J_0 optimizing the appearance of the improved video.

Currently, industry uses two types of cameras with the second type gradually becoming more mainstream, analog and digital. Most commercial footage acquisition systems either apply some compression right at the camera or apply compression in what the industry knows as DVRs. DVR's are the primarily interface between the camera and the monitor performing the visual inspection. Those also serve as storage devices. When, the video input is captured in an analog form DVRs also perform the conversion from analog to digital using MPEG. There are lossy and lossless compression forms of MPEG. MPEG is an algorithm for video encoding and compression. We stress, that once an image is acquired in a digital form all media require it to be encoded, as if it is written in words using a universal alphabet, in oder to be used, transmitted or displayed on monitor. Encoding automatically uses some rate of compression and the latter is of two kinds: Lossless where the original image can be fully reconstructed to avoid loss of information, or lossy where the choice of what is significant to the eye of the beholder is made by the algorithm based on transmission or storage constraints. MPEG maintains a high level of clarity and contrast in regions where there is significant amount of information (lots of pixel intensity gradients) and applies much higher compression rates where an image is smooth (small intensity gradients not varying rapidly) as shown in Fig. 6. In the latter regions the high compression rates create pixel value differences that illumination neutralization will enhance because it is designed to perceive them as lines of an object which is barely visible (see Fig. 5). For this reason our method and any other illumination neutralization method will perform best in raw (uncompressed) images or in images stored with lossless compression. Also relatively big suspended particles in water will contribute to the overall pixelization. On the other hand, we avoid a lot of these undesirable effects when we enhance HD-video stored with small compression loss rates (see Figs. 11 and 15).

4 CONCLUSIONS

We presented a novel, multiscale non-linear transform which enables the enhancement of singularities obscured by low illumination, water turbidity or other factors resulting from multiple scattering of light rays. The highlights of the method are the following:

- We introduce the concept of a surrogate image of a scene to formulate in a concrete mathematical manner the concept of illumination neutralization.
- We presented a numerically efficient, multiscale method for illumination neutralization that can be applied to still images and video capable of generating a unique output containing all pertinent structural information of a scene regardless of the conditions which inhibit visibility or degrade illumination.
- Contrast enhancement is not illumination neutralization, because this type of transformation only amplifies differences between pixel value intensities and does not achieve illumination neutral outputs. In other words, it provides a more informative version of f in Eq. (1) but does not separate $\mathcal{S}$ from R.
- The presented enhancement method can generate an output video with significantly improved visibility even when turbidity is high.
- As a result we anticipate that marine and off-shore industries can improve their operations by expanding the domains in which marine inspections can be credibly performed.
- The proposed method can be implemented with a system that original footage and improved should be watched side by side and conclusions should be derived by inspecting both the original and the improved footage.

Figure 5: Examples of three images used in an object detection experiment using the proposed method. All three original inputs are shown on the left column and have been retrieved from a public internet image database. In all of them visibility is reduced due to low and uneven illumination. Top and middle originals have been encoded with a lossy compression JPEG. Lossless compression has been applied to the bottom original. In the right column you can observe the outputs of $\mathcal{IN}$ for each one of the originals. The pixelization appearing mostly where the original images are dark is the result of the enhancement of pixelization existing in the input images. We show this in Fig. 6. Observe that the output of the bottom image does not have these artifacts.

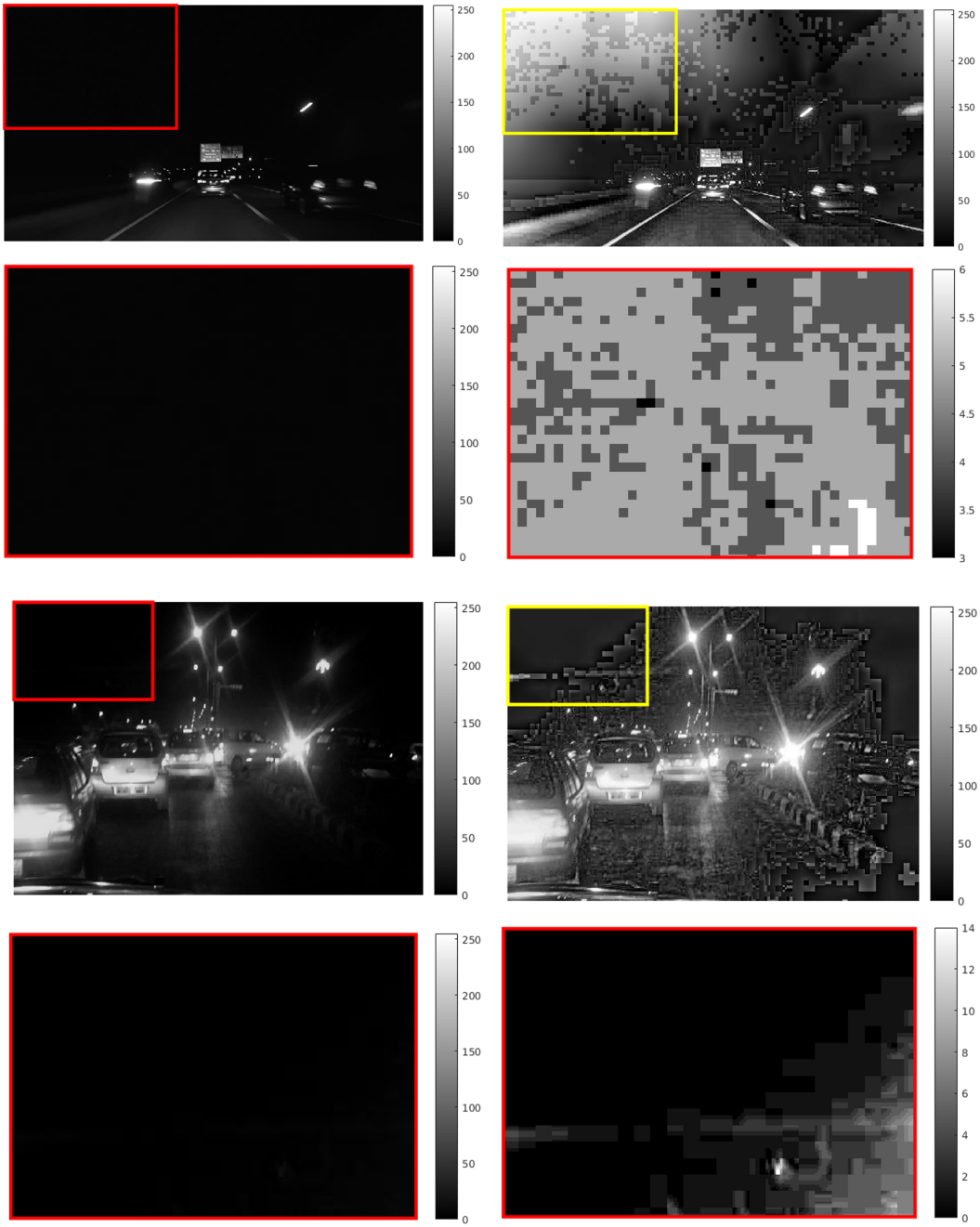
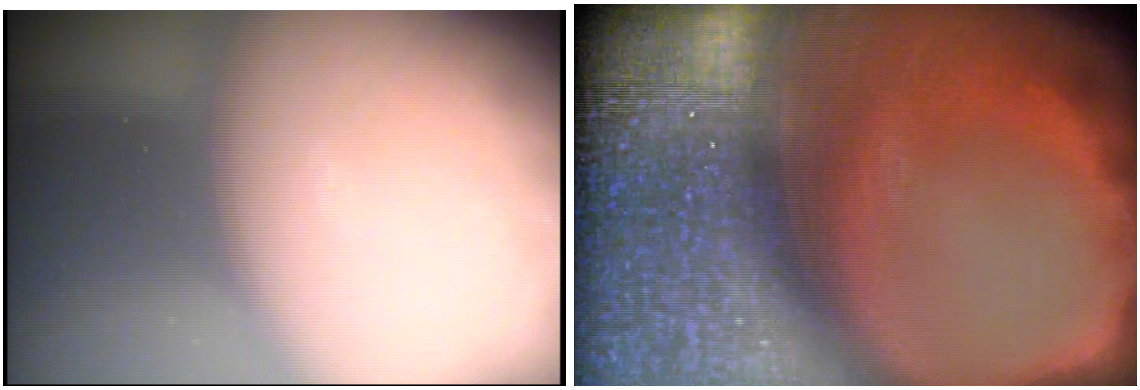


Figure 6: Left column: In rows 1 and 3 we have the two originals from Fig. 5 encoded with lossy compression. Observe the pixelization already present in the compressed original images due to very high compression rates applied on the dark regions marked by the red rectangles. Those rectangles are blown up below each original. In the right column, next to each original, we have the outputs of $\mathcal{IN}$. The red rectangles correspond to the yellow rectangles. In each one of them we observe pixelization. In rows 2 and 4 of the right column we re-display the red rectangles of the originals in the left column by merely adjusting the dynamic range of pixel values in each of these red rectangles. These adjustments reveal the same pixelization pattern we see in the yellow rectangular regions in the outputs of $\mathcal{IN}$. These two examples underscore that high compression rates will generate pixel artifacts due to the nature of JPEG encoding. For best results $\mathcal{IN}$ should be applied to images with lossless compression or with small compression rates.



(a) Original video frame (b) Illumination Enhanced

Figure 7: Frame from video inspection of a fixed tunnel bow thruster. The camera used was analog and compression was performed by an old DVR. Here we see how extreme glare results in mediocre enhancement. Glare appears to act almost as an overwhelming smoothing filter. Visibility is about 1.5m. All enhancements in this and in all subsequent figures are performed with $A\Lambda\Sigma$vision.

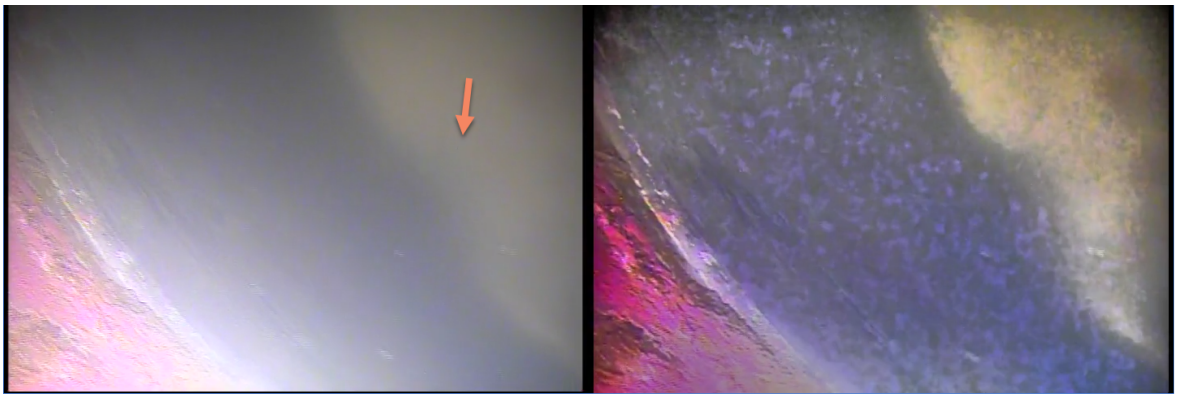


Figure 8: Frame from the same video inspection of a fixed tunnel bow thruster. Visibility is about 1.5m. (a) Blade damage of tunnel bow thruster shown in original footage shot with an analog camera. Visibility was about 1.5m. (b) Metal missing from the blade is clearly visible in this panel after enhancement.

(a) Original video frame (b) Enhanced

Figure 9: Frame from the same video inspection of a fixed tunnel bow thruster.

(a) Original video frame (b) Enhanced

Figure 10: Frame from the same video inspection of a fixed tunnel bow thruster. Some mild pixelization is observed in the region between the blades due to suspended particles and high compression rate. Notice how the strong glare is gone in (b).

(a) Sea Lion swimming (b) Enhanced

Figure 11: Sea Lion swimming in Galapagos. Frame from HD video with small compression rate. Water clarity high. Enhancement with $A\Lambda\Sigma$vision, no compression related pixelization artifacts. This image was compressed with a very low compression rate protocol.

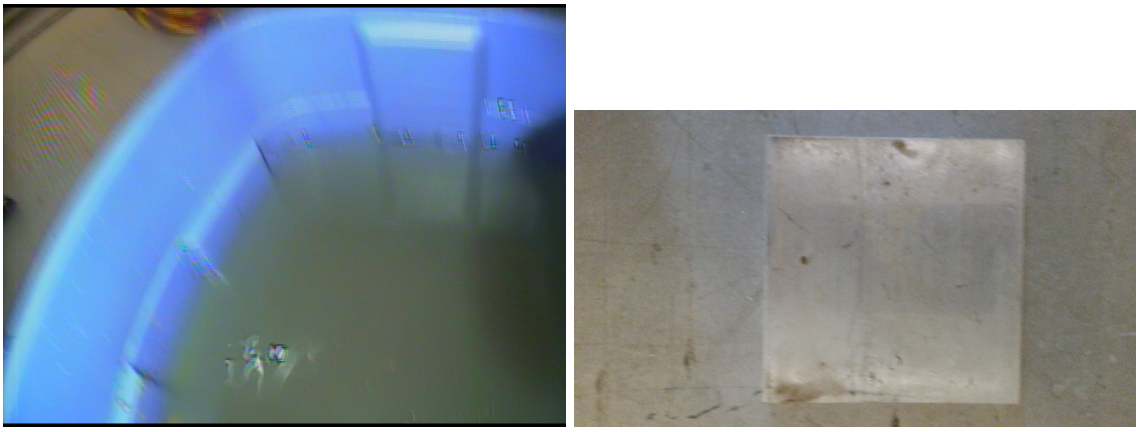
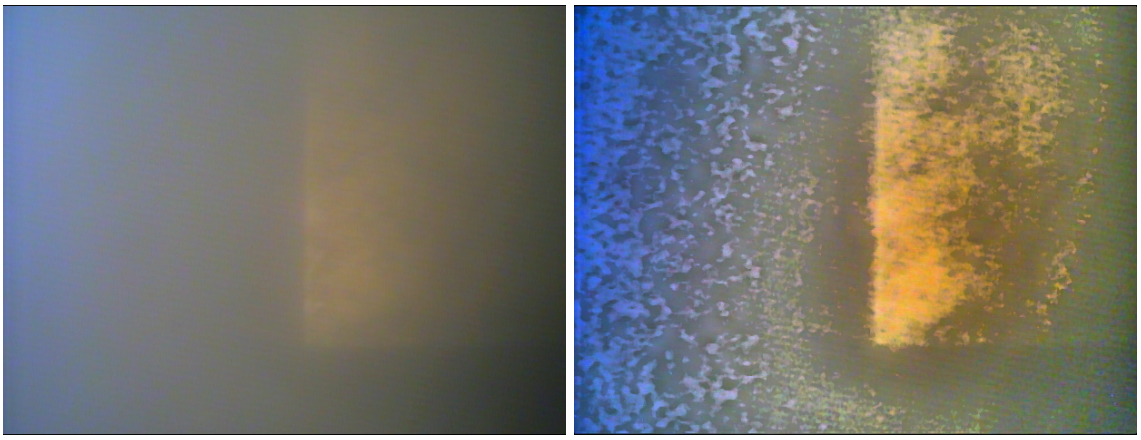


Figure 12: High turbidity simulation experiment in a small tank. The water depth is 9.5in. The metal plate shown on the right panel was resting on the bottom of the tank.



(a) Original (b) Enhanced

Figure 13: With an analog camera and the use of the Empia video capture software configured for the minimum loss of information we acquired a short video. The camera was moving in contact with the light source at varying depths inside the small tank. The plate image after enhancement is shown on the right panel. The plate is still visible in the original (a). The circular pattern in (b) is caused by the enhancement of the boundary of the light beam. The height dimension of the plate is also enhanced as a darker boundary in (b). Pixelization due to turbidity and high compression is also visible.

Figure 14: In this frame the enhanced is overlayed on the original in the turbidity simulation experiment (Fig. 12). At this point the camera is only 9.5in hovering over the metallic plate which is somewhat visible in the enhanced frame. Again we see the circular pattern of light source.

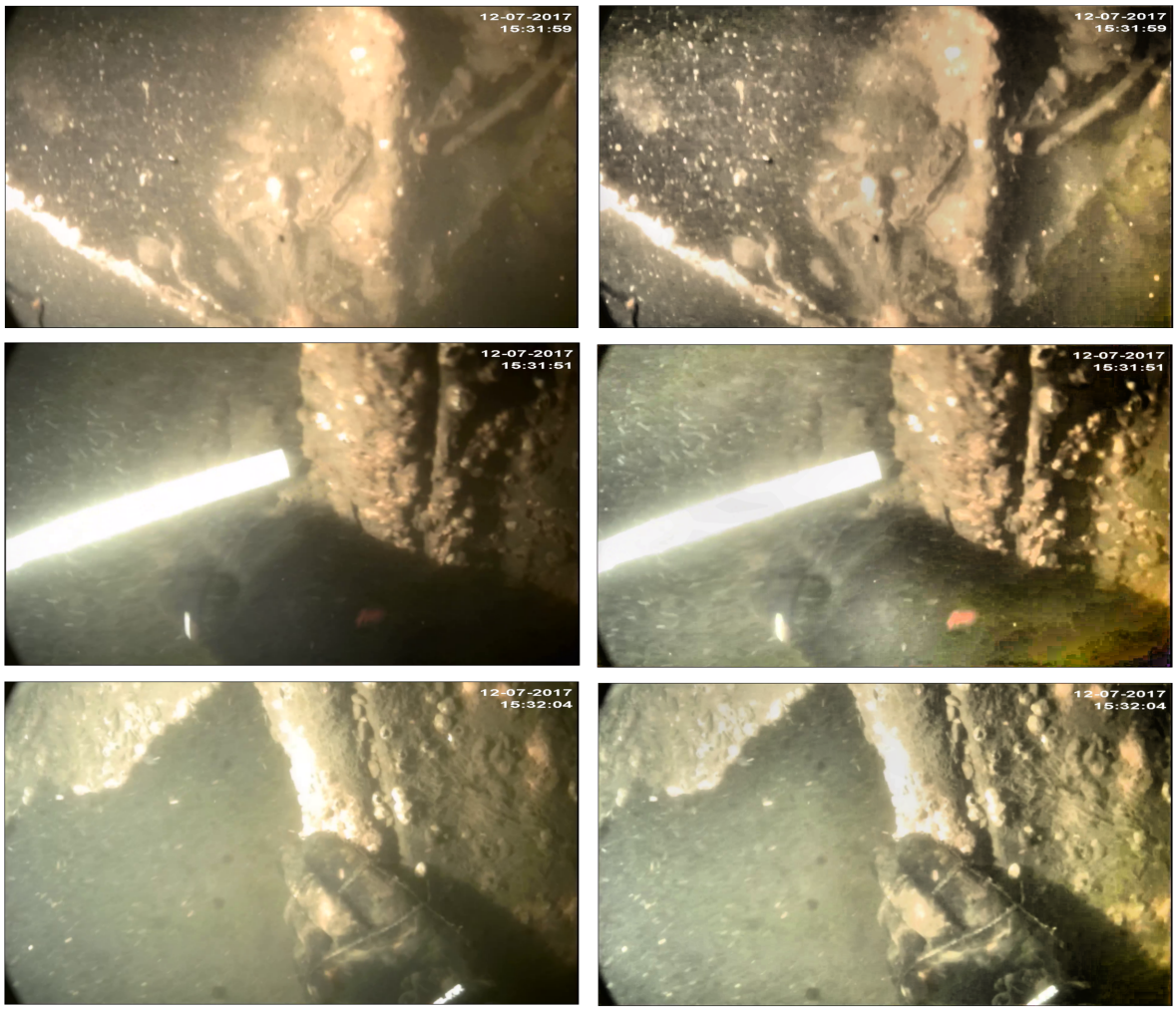


Figure 15: Sample frames from a dock inspection. The camera used is digital and the footage is compressed with a lossy compression protocol H.264 high profile 10. The loss rate is medium. Pixelization can be observed on the darker area of the enhanced image but a lot less under the dock despite of the turbidity.

5 ACKNOWLEDGMENTS

This work was partially supported by NSF awards DMS-1720487 and DMS-1320910 and by a University of Houston Seed Funding for Advanced Computing (SeFAC), 2017 "Overcoming Poor Illumination for Real-time, Large Scale Face Recognition Applications". We also want to thank all of those who helped with videos, expert opinions, advice and, encouragement. We will list them in alphabetical order: S. Adamopoulos, G. Argyros, G&G Marine, P. Noble, Outland Technology Inc., E. Psylos, Texas Commercial Diving Inc., Valsamis Inc. and Dr. Tom Campell of the UH Technology Transfer Office. Finally, many thanks to the anonymous reviewer of the 2018 SNAME Off-shore Symposium who truly went off the beaten path to provide us with valuable and honest feedback.

6 DISCLOSURES

Dr. Manos Papadakis is the managing member of Lolaark LLC in which he holds a significant financial interest. Lolaark LLC is the beneficiary of certain research results described in this study and is currently undertaking the translational effort to commercialize the method described in Section 2 as a real-time underwater video enhancement system under the brand name $A\Lambda\Sigma$vision.

References

Abry P, Jaffard S & Wendt H 2014 *in* 'Operator-Related Function Theory and Time-Frequency Analysis, The Abel Symposium 2012' Vol. 9 Springer.

Abry P, Roux S G & Jaffard S 2011 *in* 'ICASSP' IEEE pp. 4328–4331.
URL: *http://dblp.uni-trier.de/db/conf/icassp/icassp2011.htmlAbryRJ11*

Anderson P & Davie R D 2004 *Lake and Reservoir Management* **20**(2), 110–120.

Georghiades A, Belhumeur P & Kriegman D 2001 *IEEE Trans. Pattern Anal. Mach. Intelligence* **23**(6), 643–660.

Guan W, You S, Sucontphunt T & Neumann U 2012 *in* '2012 19th IEEE International Conference on Image Processing' pp. 2941–2944.

Guilheneuf B, Jaffard S & Lévy Véhel J 1998 *Applied and Computational Harmonic Analysis* **5**(4), 487–492.
URL: *https://hal.inria.fr/inria-00593247*

Guo R, Dai Q & Hoiem D 2013 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**(12), 2956–2967.

Jaffard S, Lashermes B & Abry P 2007 *in* T Qian, M Vai & Y Xu, eds, 'Wavelet Analysis and Applications' Applied and Numerical Harmonic Analysis Birkhuser Basel pp. 201–246.
URL: *http://dx.doi.org/10.1007/978-3-7643-7778-6*$_1$7

Kemker C 2014 *Fundamentals of Environmental Measurements. Fondriest Environmental, Inc* **13**.

Kimmel R, Elad M, Shaked D, Keshet R & Sobel I 2003 *International Journal of Computer Vision* **52**(1), 7–23.

Land E H 1964 *American Scientist* **52**(2), 247–264.

Land E H et al. 1977 *The retinex theory of color vision* Citeseer.

Li Q, Yin W & Deng Z 2009 *The Visual Computer* **26**(1), 41–49.

Marr D & Hildreth E 1980 *Proceedings of the Royal Society of London B: Biological Sciences* **207**(1167), 187–217.

Ochoa-Villegas M A, Nolazco-Flores J A, Barron-Cano O & Kakadiaris I A 2015 *IET Computer Vision* **9**(6), 978–992.

Shao M, Wang Y & Liu P 2010 *in* H Zha, R ichiro Taniguchi & S Maybank, eds, 'Computer Vision – ACCV 2009: 9th Asian Conference on Computer Vision, Xian, September 23-27, 2009, Revised Selected Papers, Part III' Springer Berlin Heidelberg Berlin, Heidelberg pp. 108–117.

Shashua A & Riklin-Raviv T 2001 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**(2), 129–139.

Shim H, Luo J & Chen T 2008 *IEEE Transactions on Image Processing* **17**(8), 1331–1341.

Triebel H 1997 *Fractals and spectra related to Fourier analysis and function spaces* Birkhauser Basel Boston.

Upadhyay S 2014 Extraction and Normalization of Directional Characteristics of Images and Textures using Multiscale Transforms PhD thesis University of Houston.

Upadhyay S & Papadakis M 2017 *Applied Computational Harmonic Analysis* . under revision.

Wang L, Xiao L, Liu H & Wei Z 2014 *IEEE Transactions on Image Processing* **23**(8), 3381–3396.

Zosso D, Tran G & Osher S 2013 *in* 'Proc. SPIE, Computational Imaging XI' Vol. 8657 pp. 865702–865702–16.
URL: *http://dx.doi.org/10.1117/12.2008839*

Zosso D, Tran G & Osher S J 2015 *SIAM Journal on Imaging Sciences* **8**(2), 787–826.

Zuiderveld K 1994 *in* 'Graphics gems IV' Academic Press Professional, Inc. pp. 474–485.