A Robust Approach for Monocular Visual Odometry in Underwater Environments

Mario Alberto Jordan[1,2], Emanuel Trabes[2] and
Claudio Delrieux[2]
[1]Argentinean Institute of Oceanography (IADO-CONICET).
Florida 8000, Bahía Blanca, ARGENTINA.
[2]Dto. Ingeniería Eléctrica y de Computadoras- Universidad
Nacional del Sur (DIEC-UNS).

Abstract

This work presents a visual odometric system for camera tracking in underwater scenarios of the seafloor which are strongly perturbed with sunlight caustics and cloudy water. Particularly, we focuse on the performance and robustnes of the system, which structurally associates a deflickering filter with a visual tracker. Two state-of-the-art trackers are employed for our study, one pixel-oriented and the other feature-based. The contrivances of the trackers were crumbled and their suitability for underwater environments analyzed comparatively. To this end real subaquatic footages in perturbed environments were employed.

1 Introduction

Vision-based tracking of camera pathway in scenes which are, to some extent, rich in features and relief, in both indoor and outdoor environments is an issue of high interest in Computer Vision, see [1], [2] and [3]. Particularly this is a common framework in a broad spectrum of robotics applications like visual SLAM (Simultaneous Localization And Mapping), real-time decision processes, navigation and guidance systems.

This paper targets the monocular vision-based tracking which exists in autonomous underwater navigation, particularly at low altitudes over the seabottom in shallow waters. Also we focus on spatiotemporal lighting changes on the scene in the form of caustics like sunlight waves. This perturbation is caused by the refraction of sun rays when trespassing a wavy surface and commonly has a stochastic nature. Simultaneously, the presence of suspended particles, solar glares, bubbles and backscattering difficult the visibility. All these perturbations might heavily affect the information contained in a footage, rendering it almost unusable for many decision making processes in autonomous navigation.

While poor visibility underwater may be practically mitigated to some extent by reducing the altitude of the vehicle, the rapid variations of brightness produced by the refracted sun rays on the bottom is the most destabilizing element in camera tracking in short-terms. As counterproductive, the navigation almost at ground brings occlusions along. Thereby, in order to alleviate their adverse effects in motion estimation, some illumination invariant algorithms have been presented to model the global and local lighting changes, see an extensive overview of tools in [4].

For global illumination changes, some works employ the median value of pixel residuals [4], [5], [6], [7], [8]; some others an affine brightness transfer function [9], [10]. On the other side, for local lighting changes, some authors propose image gradients, rather than pixel intensities, to formulate the direct energy function, thus gaining local lighting invariance [11]; others combine low-pass filtering and dense computation of a deliberately designed local descriptor to obtain a clear global minimum in energy [12]; and finally [13] employs a binary descriptor to achieve local illumination invariance. In [4] a combination of the sound characteristics of affine and gradient methods is proved to provide the most reliable tracking estimates in tested datasets. However, at the expense of a higher computational load in comparison with all the other methods cited above.

In large part of those approaches, local and global solar glares are focused as the main perturbation for tracking failures in outdoor environments. While glimmering on the image may seriously disrupt the clean scene, they are generally more predictable than sunlight caustics on seafloor and in principle motion-dependent which could eventually be corrected in autonomous navigation. On the other side, sunlight flickers underwater preserves geometrically and temporally its stochastic (non-stationary) nature no matter the motion and has a wide frequency range of light changes. All these characteristics place the modelling approach in a more complex category than the one of sun glares in open air.

While mostly internal modifications of the existing motion estimation techniques are aimed to accomplish robustness in visual odometry, we will, in contrast, provide a solution in other framework, namely we propose to interpose a specific filter between incoming corrupted data and the specific tracker used in the application. Thereby, a deflickered footage of the scene could be obtained on-line for supporting camera tracking in autonomous navigation without modifying existing well-proved techniques.

Many modern direct and indirect methods both in the form of sparse and dense modality, have indistinctly proved outstanding performance in varied applications indoors as well as outdoors [3]. Nevertheless, their robustness in light-disrupted subaquatic sceneries are insufficiently researched, at least to be able to draw out outright similar conclusions as in open air. For our study we select two promising techniques for visual odometry, namely the direct method DSO (Direct Sparse Odometry) by [15] and on the other side a feature-based method ORB-SLAM by [14].

In this paper we qualitatively evaluate the performance of the proposed robust approach for visual odometry on the synthetic and real-world datasets regarding accuracy and robustness.

2 Preliminaries

In our intended application areas of the autonomous navigation, the basic structure required for a robust visual system underwater are portrayed in the Fig. 1. First, the remotion of sunlight caustics and flares on the footage is carried out in real time. According to our expectations, the navigation can be supported for any well-proved monocular visual odometer, namely a direct or indirect technique indistinctly, without employing any modification to achieve robustness against fast spatiotemporal changes of the photometric properties on the seafloor. Finally, the vehicle pose estimations and surrounding mapping enable the visual system to take decisions in navigation, for instance, for guiding the vehicle autonomously with the end of exploration or revisiting, or simply collision avoidance.

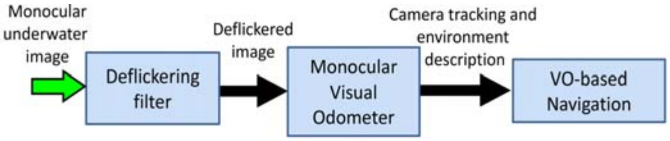

Monocular underwater image
Deflickering filter
Deflickered image
Monocular Visual Odometer
Camera tracking and environment description
VO-based Navigation

Figure 1 - Scheme for a robust visual odometry for navigation in subaquatic scenaries with sunlight caustics on the seafloor

Our primary objective in this structure is to confine the study to the proper conditioning of the footage in order to improve the performance and robustness of monocular visual odometry provided by direct and indirect methods.

3 Adaptive deflickering algorithm

The footage underwater of the perturbed luminance scenery have two kinematic components superposed, one is consistent with the egomotion of the camera and the other accounts for a rather chaotic dynamics of sunlight trails on the bottom.

The mayor challenge for a image conditioning consists in differentiate between these components since the visual information is shared to some extent in frequency and space as well. Other difficulty may happen when the stochastic caustics are non-stationary and hence no statistics could be invoked in the medium term. Consequently, the approach has to be able to adapt its performance to the changing lighting conditions, even to become completely insensitive in calm waters.

3.1 Filter structure

The basic algorithm for image deflickering is illustrated in Fig. 2.

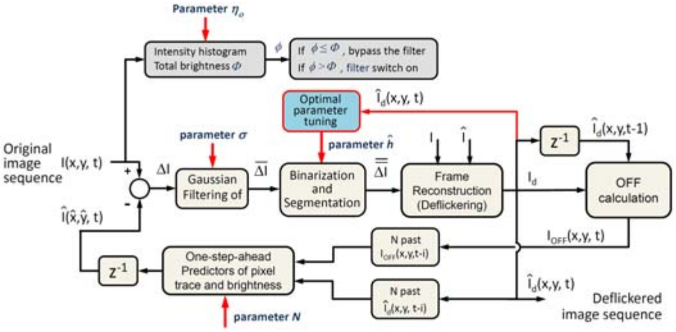


Figure 2 - On-line deflickering filter approach with self-tuning of the sensitivity

The foundation of the filter lays on a loop structure, in where the input is the raw camera video termed $I(x,y,t)$ and the output the synchronized deflickered video $I_d(x,y,t)$. Thanks this feedback, the filter may accomplish a better performance and a greater stability than common open-loop algorithms.

In a first level of the filter structure, the global brightness of the affected zones of the image is checked up in order to avoid unnecessary video processing when the caustic waves are insignificant. To this end, the intensity histogram of the raw image is computed and the accumulated brightness ϕ in the interval $[255-\eta_0, 255]$ is thresholded. Here η_0 is and integer-valued threshold and Φ is the total image brightness. If $\phi > \Phi$, the deflickering filter is activated and vice versa.

A second level comprises a loop structure with a direct path and a feedback path.

In the direct path, the raw image $I(x,y,t)$ at time point t is compared with the *a-priori* estimation of the deflickered image $\widehat{I}(x,y,t)$. The difference yields the function $\Delta I(x,y,t)$ that contains (with some degree of accuracy) the first approximation of the caustic fringes at t. Also, ΔI will usually contain many relatively small specks, someone of them are related to light scattering and some others are due to high-frequency noise originated in the image processing. Essentially, the information of the egomotion has been removed in this instance.

While the effect of scattering is commonly reduced with temporal averaging that takes place in the feedback block, the spatially sprinkled noise is smoothed by Gaussian filtering with standard deviation σ. This yields $\bar{\Delta}\bar{I}(x,y,t)$. Large values of σ may change the sharpness of the final deflickered image significantly. A default value $\sigma = 9$ is proposed for good results.

Hereafter, $\bar{\Delta}\bar{I}$ is segmented by means of binarization, choosing for this purpose a threshold h. In this way, areas containing the most brilliant points are supposed those containing sunlight fringes. This yields the set $\overline{\bar{\Delta}\bar{I}}(x,y,t)$ in white color, which is the footprint of the caustics composed of many nonconnected areas referred to as S_i. Also its complementary sets, denoted by $\overline{S}_i$, will be useful later. The black background in $\overline{\Delta \bar{I}}$ represents regions of the seafloor which are much less perturbed by fringes.

Actually, the contours of fringes are by no means sharp, but rather completely fuzzy. Thus the proper selection of h is not trivial, not even to the naked eye. The best choice of h may be obtained by Otzu´s method under the supposition of bimodal distribution of $\overline{\Delta \bar{I}}$, or alternatively, by a more advanced method suggested here which is described next.

Having an estimation of the flicker regions S_i, the a-posteriori image $\widehat{I}_d$ can be reconstructed as follows. First, we define $\widehat{I}_d$ in every S_i as the intensity of the a-priori estimation in $\widehat{I}$. This results in I_{S_i}. Second, $\widehat{I}_d$ is defined with the intensity of the raw image I in the complements $\overline{S}_i$, yielding $I_{\overline{S}_i}$.

3.2 Filter feedback loop

Now we focus on the feedback path, wherein the main operation is the estimation $\widehat{I}$ for the next step $t+1$. The main idea is to track every pixel of I_d at t within a short period of time backwards in order to be able to predict a new a-priori estimation of the deflickered image. The pixel tracking is supported by an optical flow field on I_d during N steps, resulting in the sequence $D_{OF}(x, y, t{+}1{-}l)$, with $l = 1, ..N$. To this end, the celebrated algorithm of Farnebäck was implemented.

Hence, every pixel $[x, y]$ of $\widehat{I}_d$ is threaded frame by frame according to its motion direction pointed out by I_{OF}.

Certainly, there exist some inconsistent cases of interrupted threads, that might occur often. The causes for that may be occlusions, simply image noise or natural motion that make threads to eventually bifurcate or to meet themselves halfway. Since the number of inconsistencies is negligible relatively to the whole set, the affected pixels can preserve their original brightness without impairing the final result .

Once all the threads are computed at t, one proceeds to predict one step ahead the next link $[\widehat{x}(t+1), \widehat{y}(t+1)]$. This is accomplished by a N-th-order interpolation of the thread positions.

The feedback description ended with the estimation of the *a-priori* brightness $\widehat{I}$ for every link $[\widehat{x}(t+1), \widehat{y}(t+1)]$. A suitable estimation of $\widehat{I}$ for the pixel positions is accomplished by a weighted average of the intensities on each thread. A default value for N may be 9, while the interval $[3, 12]$ for N was identified as suitable. Since N depends on the pixel motion, it may be particularly tuned for every thread according to its rate D_{OF}. The faster the pixel rate the smaller N and vice versa.

Finally, the *a-priori* estimated $\widehat{I}(x, y, t{+}1)$ will be employed just in the next step to be compared with the incoming frame at $t+1$ as shown in the feedback path on Fig. 2.

3.3 Self-tuning of the filter sensitivity

The success of the filter depends to a large extend on the correct tuning of the threshold h for determining real fringe contours.

Even in cases in where h is appropriately set, the changing conditions of illumination in complex sceneries underwater might demand a new tuning. According to the complex wavy surface dynamics and its magnifying-glass effect of the light passing through the air/water layer, flicker borders are typically blurred.

A solution for this specific problem is proposed here as a process of continuous self-tuning of h in real-time. The key observation is that the deflickered image $\widehat{I}_d$ often reflects patterns of thin and faint trails related to caustic residues even for suitable set of threshold values. These residues reflects the brightness disparity between the contours of the identified flickers $\overline{\overline{\Delta I}}$ and the contours of their morphologically dilated regions. However, by decreasing h continuously one can notice the existence of a breakpoint, and from this point onwards the residues do no longer increase. Particularly in this situation, the residues will be typically quite small and are only accounted for the subtle superiority in quality of the *a-posteriori* image over the *a-priori* image.

So in the following, we attempt to estimate the breakpoint of h which is the optimal threshold value named h_{opt}. In our proposal it is not assumed a bimodal density function of the brightness as for instance the Otsu's method does. Indeed, we have no evidence of a bimodal distribution of residues to both sides of the contours, particularly when the residuals are small.

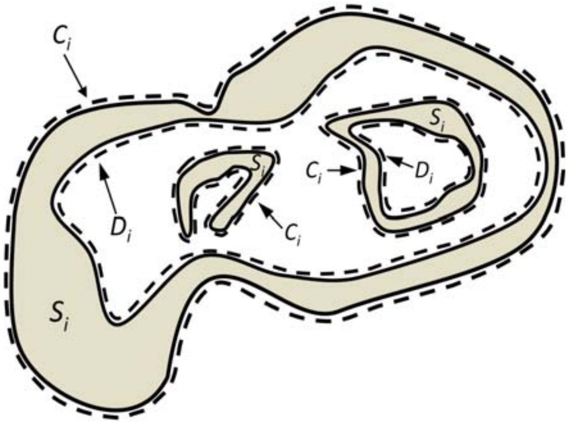


Figure 3 - Typical topology of a sunlight fringes composed of nested genus-0/1 areas S_i. Indication of 1-pixel dilated boundaries C_i and 1-pixel compressed boundaries D_i

To estimate h_{opt} we propose a suitable cost function $V(h)$ of the residues based on the *a-posteriori* image $\widehat{I}_d(x, y, t; h)$. For any value of h there exists a number of estimated non-connected fringe areas encompassed in the set S_i.

Typically they may be conformed by nested genus-n areas with $n \geq 0$, see Fig. 3. This means that there exist fringes with external and internal boundary as well referred to as S_i^C and S_i^D, respectively. Herein we distinguish between a boundary C_i which results from dilating 1 pixel an external boundary S_i^C of some S_i, and a boundary D_i which results from compressing 1 pixel of the internal boundary S_i^D of this genus-1. Notice that simply-connected areas do not have S_i^D. The following step is to evaluate the brightness along the natural contours of the S_i's as well as on the C_i's and D_i's with the end of constructing a cost function $V(h)$ of residues. The residues for every S_i yield from aggregated intensity differences between S_i^C and C_i from one side and between S_i^C and CD_i from the other side. They provides the errors for the cost $V(h)$ as in the following

$$e^C_{S_i}(h,t) = \sum_j \widehat{I}_d(S_i^C, t; h) - \sum_j \widehat{I}_d(C_i, t; h) \tag{1}$$

$$e^D_{S_i}(h,t) = \sum_j \widehat{I}_d(S_i^D, t; h) - \sum_j \widehat{I}_d(D_i, t; h) \tag{2}$$

$$\text{with pixels } (x_j, y_j) \in \left\{S_i^C, S_i^D, C_i, D_i\right\} \tag{3}$$

$$V(h) = \sqrt{\frac{1}{n_s}\left(\sum_{i=1}^{n_s} \left(e^C_{S_i}(h)\right)^2 + \left(e^D_{S_i}(h)\right)^2\right)}, \tag{4}$$

where n_s is the number of sets S_i.

The optimization process is carried out for every frame employing a steepest-descent algorithm. However as this procedure has to be done frame by frame it is computationally convenient to perform the estimation discretely in the domain of h. In general, the cost function looks approximately piece-wise linear in h with a breakpoint h_{opt} which is the optimal point. For the first time, one can search for h_{opt} iteratively starting conveniently from a rather high value of the initial condition $\widehat{h} = h_0$ and descend stepwise until one passes from large to small variations of $V(h)$ abruptly. In non-stationary caustics or by spurious as well, one can work in the modus self-tuning of h_{opt}, and start with an initial condition whose value is slightly superior to the previously identified $\widehat{h} = h_{opt}$.

4 Monocular visual odometry

The target now is to evaluate the performance of specific monocular visual odometry approaches with the deflickering filter embedded in the structure portrayed in Fig. 1.

Here we chose two well-proved and probably the most representative exponents of the categories feature-based and photometric-based methods in the state of the start. They are respectively ORB-SLAM [14] and DSO [15]. Other hybrid approaches that combines feature extraction in keyframes only but elsewhere they operate directly on pixel intensities such as DSO. So we will draw out our own conclusions in working with these pure opposite classes in the framework of underwater applications.

Here below, we summarized their features, operating mechanisms and design parameters. We attempt to stress particular characteristics of each method that we believed to play a clear role in its performance and make differences on equal terms in subaquatic sceneries.

4.1 Monocular ORB-SLAM

ORB-SLAM is a versatile and accurate real-time monocular SLAM approach that uses features for tracking, mapping, relocalization, and loop closing as well, in both small and large indoor and outdoor environments.

4.1.1 Features

The system is robust to severe motion clutter, it also enables place recognition from substantial viewpoint change and good invariance to light changes.

It is able to match features with a wide baseline, due to a relatively good invariance to viewpoint and illumination changes. It includes an automatic and robust initialization from planar and non-planar scenes.

It uses the same features for all tasks of the front- and backend. This makes the system more efficient, simple, and reliable, avoiding the need of feature interpolation. It employes bundle adjustment over features.

It was tested with datasets and benchmarks in small and large indoor and outdoor environments, hand-held, car and robot sequences, for instance the TUM RGB-D Benchmark and the odometry benchmark KITTI [15] with multiple loops.

Its hardware requirements are not so evolved, actually author´s tests were carried out in real-time employing an Intel core i7 and without the need to employ GPU acceleration.

Even so, there is no report about applications underwater. Moreover, there is in the literature no systematic study focusing on, in a way similar perturbations in aerial environments like glares, haze or moving cloud shadows that might serve to extrapolate results for subaquatic environments with light scattering and caustics.

4.1.2 Design parameters

Number of features per image (ORBextractor.nFeatures, for instance 1000 % 6000).

Scale factor between levels in the scale pyramid (ORBextractor.scaleFactor, for instance 1.2).

Number of levels in the scale pyramid (ORBextractor.nLevels, for instance =8).

Fast threshold. Image is divided in a grid. At each cell FAST are extracted imposing a minimum response. Firstly one imposes an initial value (ORBextractor.iniThFAST, for instance 20). If no corners are detected one imposes a lower value (ORBextractor.minThFAST, for instance 7). The values can be lowered if images have low contrast.

4.2 Monocular DSO

The direct sparse visual odometry sustained by DSO is based on continuous optimization directly over image pixel photometric errors over a window of recent frames, taking into account a both geometrically and photometrically calibrated model for image formation. It possesses the ability to use and reconstruct all points instead of only corners like feature-based techniques. Besides, it jointly optimizes for all involved parameters (camera intrinsics, camera extrinsics, and inverse depth values) employing Gauss-Newton optimization, effectively performing the photometric equivalent of windowed sparse bundle adjustment. As customary in direct methods, the geometry representation employed is composed of 3D points represented as inverse depth in a reference frame. A careful calibration of a high-performance camera is necessary.

4.2.1 Features

Camera model enhancement: A photometric model of the image formation along with the traditional geometric model are beneficially combined.

Point Dimensionality. In DSO a point is parametrized by the inverse depth in the reference frame in contrast to three unknowns as in the indirect model

Global solution in a multidimensional space: photometric error is defined in a Lie group $SE(3)^n \times R^m$ with n optimizing variables, including camera pose, lens attenuation, gamma correction, and known exposure times, inverse depth values and camera intrinsics, and m being the size of the sliding window

Optimization is performed in a sliding window using the Gauss-Newton algorithm on the total error, where old camera poses as well as points that leave the field of view of the camera are marginalized

Since DSO does not depend on keypoint detectors or descriptors, it can naturally sample pixels from across all image regions that have intensity gradient, including edges or smooth intensity variations on mostly white walls

Its hardware requirements are above all focused on high-performance cameras (global shutter, precise lenses and high frame-rates) with the end of squeezing the full potential of direct formulations

4.2.2 DSO parameters

Active points N_p (for instance maximal $N_p = 2000$, reduced $N_p = 800$).

Active frames N_f in the window (for instance $N_f = 6$).

Number of neighbor pixels for residual pattern $\mathcal{N}_p$ (typical$\mathcal{N}_p = 8$ pixels).

Image resolution (for instance 424×320 pixels).

Number of Gauss-Newton iterations after a keyframe is created (typical ≤ 6).

Size of blocks for splitting the image (for instance 32×32 blocks).

Global constant for the median of absolute gradient threshold over all pixels in a block (typical $g_{th} = 7$).

5 Case studies

The respective algorithms for visual odometry were employed with the original public source code.

5.1 Environment and settings

For providing a ground truth for tests and being able to achieve acceptable reproducibility of the results, a scenery is staged in a basin containing expected elements with good semblance with the natural underwater landscape, for instance rocks, gravel, sand banks and benthos among others, wherein a diversity of subaquatic-like visual effects can be obtained, see Fig. 4. Thereby one can replicate blurriness, rapid illuminance changes, self-similarities, occlusions and sunlight lens glares as well.

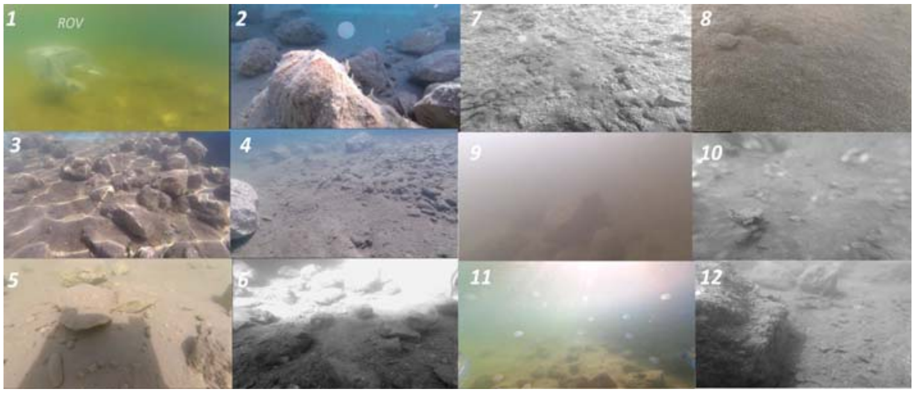


Figure 4 - Diversity of the staged sceneries for tests underwater. 1) OPEN-ROV with frontal monocular camera, 2) 3D scene with lens flares, 3) Well textured scene with strong sunlight caustics, 4) Well textured scene with clear visibility,, 5) Sharp vision with vehicle shadow, 6) Transition from a dim to a shining scenario, 7) Self-similar floor with moderate relief, 8) Self-similar floor with granular texture, 9) Scene with extreme blurriness, 10) Motion blur, 11) Glare and bubbles, 12) Scene with occlusions

The bioactivity of microorganisms and particles change permanently the texture characteristics and water transparency. This causes bubbles which may adhere uniformly to the floor and raise in front of the camera. In our test phase, the visual state on the basin floor was periodically reestablished to nearly the initial state of transparency and texture. At low altitudes the scene appearance may be volumetric and hence occlusions arise very often. Also the own shadow of the vehicle on the floor might appear on the floor and change the lighting conditions significantly.

For comprehensive assessment of the approach adequacy we have classified the tests according to the traits of the perturbation, see Fig. 4.

As platform for the camera motion we have employed a ROV (OpenROV v2.8) whose paths were steered via teleoperation in the coordinates x-y while its altitude z was automatic regulated around a previously defined reference depth. The ROV model was modified by coupling two hydro stabilizers in the bow for pitch steadying, regarding high speed changes. The path lengths for the tests, range from about 12 meters up to 30 meters, which are enough to manifest performance differences.

The vehicle has two independent cameras, namely a low-resolution model (Genious f100) and a high-performance model (GoPro Session H4). They operate with different frame rates and are both rolling shutter. In order to subdue the effect of rolling shutter, above all in the DSO performance that suggests built-in global shutter, we have used a high fame rate of typically 120 fps with an image size 848x480 pixels. Moreover, we have ensured moderate movements of the camera on the path.

The scenery is illuminated by direct and indirect sunlight. Sunlight flickers were produced by uniformly shacking the water surface in two roughly orthogonal directions of the basin. The visibility was reduced by discharging on purpose particles of silt on the surface that rest suspending in the column for a while.

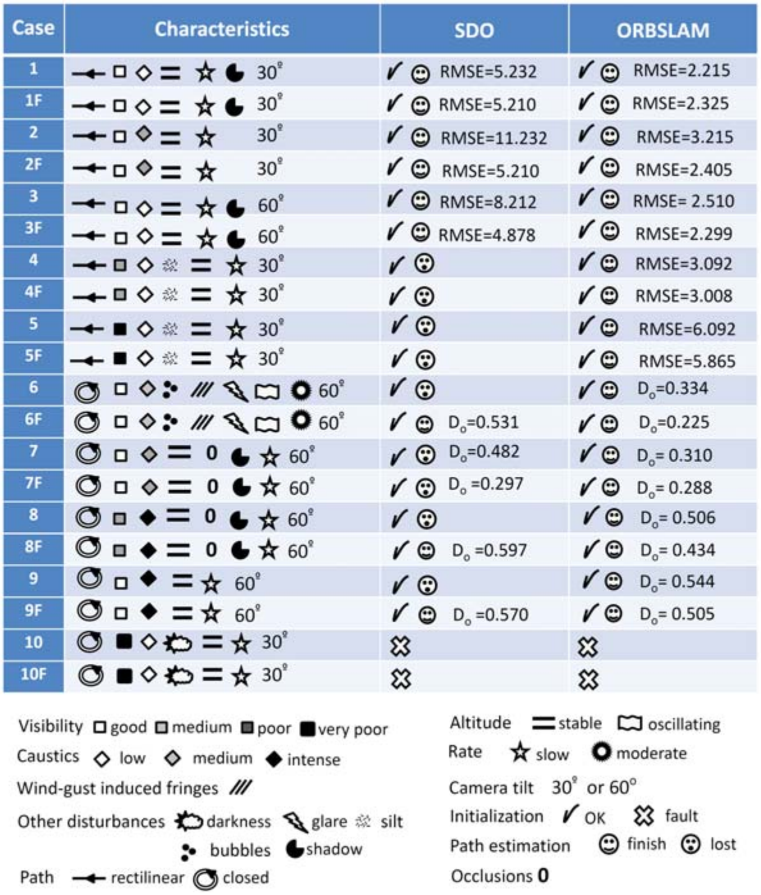
Even when SDO is computationally more intensive than ORB-SLAM due to a complex cost optimization in a multidimensional domain, hardware aspects will be not considered in this work, Thereby, we will remove the dependency on the host machine's CPU speed by not enforcing real-time execution of the experiments. The source code of every approach is taken from public domains. The datasets and source code of the deflickering filter are propriety of the authors.

5.2 Case studies

We analyze 10 datasets comprehending a subset of straight-ahead paths, and a second subset an irregular close path with identical starting and ending camera poses. Both subsets share different lighting conditions, visibility levels and different altitudes. Moreover, the datasets involve two different fields of vision, namely camera tilts of 30 and 60 degrees, see Table 1. The small tilt accomplishes a rather distant horizon with commonly marked blurriness at the top of the image, while the larger tilt represents a nearby horizon with a better visibility all around. Occasionally, the vehicle altitude and pitch became slightly oscillatory in the water column due to surface waves or sudden rate changes. This is contemplated in the experiments.

The initialization phase for each method is performed before the assessment of performance is started.

In rectilinear paths, the ground-truth geometry was represented by a line across the length of the basin. The scale factor for scaling the estimations was in this case roughly determinable. Thereby, we can straightforwardly calculate the estimated path deviations in root-mean-square error (RMSE) with respect to the physical line of navigation in the scale of the estimation.



Case	Characteristics	SDO	ORBSLAM
1	⟵ □ ◇ = ☆ ◕ 30°	✓ ☺ RMSE=5.232	✓ ☺ RMSE=2.215
1F	⟵ □ ◇ = ☆ ◕ 30°	✓ ☺ RMSE=5.210	✓ ☺ RMSE=2.325
2	⟵ □ ◈ = ☆ 30°	✓ ☺ RMSE=11.232	✓ ☺ RMSE=3.215
2F	⟵ □ ◈ = ☆ 30°	✓ ☺ RMSE=5.210	✓ ☺ RMSE=2.405
3	⟵ □ ◇ = ☆ ◕ 60°	✓ ☺ RMSE=8.212	✓ ☺ RMSE= 2.510
3F	⟵ □ ◇ = ☆ ◕ 60°	✓ ☺ RMSE=4.878	✓ ☺ RMSE=2.299
4	⟵ ▣ ◇ ※ = ☆ 30°	✓ ☹	✓ ☺ RMSE=3.092
4F	⟵ ▣ ◇ ※ = ☆ 30°	✓ ☹	✓ ☺ RMSE=3.008
5	⟵ ■ ◇ ※ = ☆ 30°	✓ ☹	✓ ☺ RMSE=6.092
5F	⟵ ■ ◇ ※ = ☆ 30°	✓ ☹	✓ ☺ RMSE=5.865
6	◎ □ ◈ ⁘ /// ⚡ ⌓ ✹ 60°	✓ ☹	✓ ☺ D_o=0.334
6F	◎ □ ◈ ⁘ /// ⚡ ⌓ ✹ 60°	✓ ☺ D_o=0.531	✓ ☺ D_o=0.225
7	◎ □ ◈ = 0 ◕ ☆ 60°	✓ ☹ D_o=0.482	✓ ☺ D_o= 0.310
7F	◎ □ ◈ = 0 ◕ ☆ 60°	✓ ☹ D_o =0.297	✓ ☺ D_o= 0.288
8	◎ ▣ ◆ = 0 ◕ ☆ 60°	✓ ☹	✓ ☺ D_o= 0.506
8F	◎ ▣ ◆ = 0 ◕ ☆ 60°	✓ ☺ D_o =0.597	✓ ☺ D_o= 0.434
9	◎ □ ◆ = ☆ 60°	✓ ☹	✓ ☺ D_o= 0.544
9F	◎ □ ◆ = ☆ 60°	✓ ☺ D_o =0.570	✓ ☺ D_o= 0.505
10	◎ ■ ◇ ☁ = ☆ 30°	✖	✖
10F	◎ ■ ◇ ☁ = ☆ 30°	✖	✖

Visibility □ good ▣ medium ▩ poor ■ very poor

Caustics ◇ low ◈ medium ◆ intense

Wind-gust induced fringes ///

Other disturbances ☁ darkness ⚡ glare ※ silt

⁘ bubbles ◕ shadow

Path ⟵ rectilinear ◎ closed

Altitude = stable ⌓ oscillating

Rate ☆ slow ✹ moderate

Camera tilt 30° or 60°

Initialization ✓ OK ✖ fault

Path estimation ☺ finish ☹ lost

Occlusions 0

Table 1 - Performance comparison of DSO and ORB-SLAM under subaqcuatic lighting perturbations with/without deflickering filter (F)

In loop paths in contrast, the ground truth was simply based on the coincidence of physical coordinates in 6DOF for the starting and ending points of the path. Initially the pose is achieved by launching the vehicle from inside a tight two-gate garage and recovering it after accomplishing a perfect alignment after entering the vehicle in the garage from the opposite direction. As no information to define the ground truth is available other than the extreme poses, the performance drop of the approach is evaluated by calculating the pose difference from the estimations. Since each approach operates with its own scale and every test video is common to both approaches, the pose gap is scaled according to the principle that the longer the estimated path the larger the deviation. So, the quantized pose gap is normalized with respect to the estimated path length in each approach, this results in the measure defined by D_0. Moreover, translation and rotational errors are evaluated separately.

In the table we summarized the results for both methods. The comparative characteristics of the approaches are directly exposed in the following.

6 Conclusions and future work

We the aim to bestow more robustness to novel and successful monocular visual odometry approaches in Computer vision for underwater navigation applications, we have proposed a visual system based on the conjunction of a deflickering filter and a one of such odometry techniques. The filter serves to wipe sunlight fringes off the footage and can on-line adapt its performance by itself to lighting changing conditions of the non-stationary lighting over the seafloor. For performance validation we have chosen a photometric-based technique like DSO on one side and a feature-based method like ORB-SLAM on the other side.

For the ground-truth experiment scheduling we have ensured a wide spectrum of lighting disturbances like sunlight caustics and turbidity levels, among the main ones. Also different navigation modes (in altitude and path geometry) and a variety of landscape of the seafloor have been complemented. For the result assessment, simple but resounding measures were employed, namely the deviations of the estimated path in physical rectilinear trajectories as well as in the falling coincidence of the extremes of the estimated paths in physical closed loops. In circa fifty registered footages we have chosen ten non-redundant ones for illustrating the results in this paper.

Despite the good resilience and robustness posted about direct methods in open air against blurriness, low-texture environments and highfrequency texture like asphalt for instance, DSO manifested problems to initialize the algorithm in datasets with medium and severe caustics together with some medium and high level turbidity. However, a clear improvement was observed when a filtering of fringes was put before the processing footage. Nevertheless, the mere presence of blurriness by low tilt pose of the camera is a serious cause for the pose lost. Perhaps the lack of merit accomplished in the performance of DSO might be partly related to a high-performance camera model and its sophisticated calibration that are particularly demanded by this method. In this sense, we have attempted to compensate rolling shutter effects and to carry out a conscientious setting of camera parameters of our deployed underwater cameras.

On the other side, ORB-SLAM have proved a better all-round performance, even in cloudy and averagely illuminance perturbed environments. The combination of ORB-SLAM with the filter has improved its resilience to spatiotemporal changes of the lighting.

Severe blurriness not only affects the success of the pose estimation but also the proper initialization in both methods alike, above all in low tilts of the camera in where the horizon is harshly affected. By all accounts, vehicle own shadow was not any cause of failure in clear waters.

Overall in the experiments, no matter the method applied, deflickering the footage previous to processing it, has manifested of contributing to better results in the experiments, above all, in caustics cases.

References

1. Lowry, S., Snderhauf, N., Newman, P. , Leonard, J. J., Cox, D. , Corke, P. and Milford, M. J.: Visual Place Recognition: A Survey,in IEEE Transactions on Robotics (TRO) (2016) 32(1):1–19

2. Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, N. Reid, I.D. and Leonard, J.J.: Past, Present, and Future of Simultaneous Localization And Mapping: Towards the Robust-Perception Age,in Cornell University Library, code of the digital open access arXiv:1606.05830v2 (2016)

3. Chahine, G. and Pradalier, C.: Survey of monocular SLAM algorithms in natural environments, in Proc. CRV (2018)

4. Wu, X. and Pradalier, C.: Illumination Robust Monocular Direct Visual Odometry for Outdoor Environment Mapping. HAL Id: hal-01876700 https://hal.archives-ouvertes.fr/hal-01876700, (2018)

5. Meilland, M., Comport, A. , Rives, P. and Mediterranee, I. S. A.: Realtime dense visual tracking under large lighting variations, in British Machine Vision Conference, University of Dundee (2011) vol. 29

6. Goncalves T. and Comport,A. I.: Real-time direct tracking of color images in the presence of illumination variation, in Robotics and Automation (ICRA), 2011 IEEE International Conference on. IEEE (2011) 4417–4422

7. Bloesch, M.S., Omari, S., Hutter, M. and Siegwart, R.: Robust visual inertial odometry using a direct ekf-based approach, in Intelligent Robots and Systems (IROS), IEEE/RSJ International Conference (2015)

on. IEEE, 2015, pp. 298–304

8. Greene, W. N., Ok, K., Lommel, P., and Roy, N.: Multi-level mapping: Real-time dense monocular slam,” in Robotics and Automation (ICRA), 2016 IEEE International Conference on. IEEE (2016) 833–840

9. S. Klose, P. Heise, and A. Knoll, Efficient compositional approaches for real-time robust direct visual odometry from rgb-d data,” in Intelligent Robots and Systems (IROS), IEEE/RSJ Int. Conf. on. IEEE (2013) 1100–1106.

10. Engel, J., Stueckler, J. and Cremers, D.: Large-scale direct slam with stereo cameras, in Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on. IEEE (2015) 1935–1942

11. Dai, A. Nießner, M., Zollhoefer, M., Izadi, S. and Theobalt, C.: Bundlefusion: Real-time globally consistent 3d reconstruction using on-thefly surface reintegration, ACM Transactions on Graphics (TOG), (2017), 36, no. 4, 76a

12. Klose, S.Heise, P. and Knoll, A.: Efficient compositional approaches for real-time robust direct visual odometry from rgb-d data, in Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on. IEEE, (2013) 1100–1106

13. Alismail, H., Browning, B. and Lucey, S.: Direct visual odometry using bit-planesm, arXiv preprint arXiv:1604.00990 (2016)

14. Raúl Mur-Artal, J. M. M. Montiel and Juan D. Tardós. ORB-SLAM: A Versatile and Accurate Monocular SLAM System, in IEEE Transactions on Robotics DOI: 10.1109/TRO.2015.2463671 (2015) vol. 31, no. 5, 1147-1163

15. Engel, J., Koltun, V. and Cremers, D.: Direct Sparse Odometry, In arXiv:1607.02565, (2016.)