Employing Feedback to Filter Caustic Waves in Underwater Scenes in Motion

Emanuel Trabes and Mario Jordán*

Argentinean Institute of Oceanography (IADO-CCT-CONICET). Florida 8000,
Bahía Blanca, ARGENTINA.
Dto. Ingeniería Eléctrica y de Computadoras- Univ. Nacional del Sur (DIEC-UNS)

Abstract. A real-time approach for removing sunlight flickers from sub-aquatic scenarios captured in videos is presented. For this end, a de-flickering filter is designed. The start point is a moving landscape scene. Essentially, the filtering approach is based on morphological characteristics of the caustic waves. It constructs an a-priori de-flickered image which is afterwards enhanced. The algorithm employs feedback of optical flow fields and brightness in order to predict a one-step-ahead value of the brightness.

Keywords: Underwater images, Image filter, Caustic waves, Underwater flicker, Feedback.

1 Introduction

Video surveys in shallow waters are occasionally affected by brightness fluctuations of a spatio-temporal scene radiance due to sun light refraction. Submerged objects on a dominantly textured sea-bottom are illuminated by a natural random pattern, which is spatially and temporally varying, following complex dynamics of the wavy water surface. These are referred to sunlight flickers or caustic waves and characterized by fast moving patterns, which can significantly degrade the quality of underwater imagery on sea-floor scenarios [13] [14].

This paper is motivated by specific problems in Computer Vision that arrive in autonomous and tele-operated navigation underwater. Therein, topics related to underwater robotics such as SLAM problems (Simultaneous Localization And Mapping), $3D$ shape reconstruction [11],[10] and $3D$ trajectory recovery of autonomous navigation [7], [2] stand out among the most representative ones that employs vision-based supports. Basically, a sequence of frames taken from variable viewpoints are employed to estimate a mapping metric of a sea-floor scenery, employing a $2D$ imagery provided by camera that moves along a $3D$ trajectory. Specifically, the metric is constructed from a sequence of estimated distances between static objects on the sea-floor and a moving camera. So, perturbed footage

---

* Corresponding Authors: M. Jordán: E-mail: mjordan@criba.edu.ar and E. Trabes: etrabes@iado-conicet.gob.ar. Address: CCT-CONICET. Florida 8000, B8000FWB, Bahía Blanca, ARGENTINA

2   Trabes and Jordán

with sunlight flickers will extremely reduce the ability of any SLAM algorithm to built up a metric of sea bottom surroundings. So, the de-flickering process emerges as a necessary step towards this goal.

When estimating camera motion with respect to the bottom structure, one basically searches for the traces of any physical point of the bottom by tracking its corresponding pixel from frame to frame [16]. Ideally, any physical point of a lambertian surface conserves the luminance isotropically during its relative motion. Perturbed points, mainly due to sunlight flicker and scattering, cause inconsistency on motion analysis. As the path of each object point can be tracked in the temporal sequence, it seems reasonable to employ temporal filtering to achieve a de-flickered video.

The removal of spatio-temporal noise passively has been studied, for example in [9], mainly to remove brightness error and fluctuation found in old film footage. The method employs motion estimation together with various other techniques in order to restore the footage. As reported, the footage must not be degraded in a manner that motion cannot be calculated with some degree of precision.

Caustic wave filtering in underwater imagery was studied in [12], [5]. In [12] the method proposed consists on using the spatial derivatives of the video frames, instead of the raw frames. The idea behind this implementation is that the derivatives of the brightness changes will have sharp edges. This, in combination of temporal averaging (median), would yield better results. This method requires a quasi-static landscape scene, because of the median averaging in time. Besides, the approach assumes that the brightness change found in the scene will have sharp edges. Nevertheless, it can be generally noticed that some caustic waves can be found to have smooth spatial derivatives.

In [5] the method developed employs a simple motion estimation though results in 2D horizontal motion are presented. It is based on computing the image difference between a given reference frame and the temporal median of a registered set of neighbouring images. A key observation therein is that this difference will have two components with separable spectral content. One is related to the illumination field (lower spatial frequencies) and the other to the registration error (higher frequencies). The illumination field, recovered by low-pass filtering, is used to correct the reference image. It is claimed the preservation of the image sharpness, even in the presence of registration inaccuracies.

In this work we address the problem of removing sunlight caustic waves in shallow water images obtained through a monocular camera in motion with no other mean available. A heuristics for a filter design is developed following certain morphological properties of the refracted patterns rather than the way of discerning band frequencies of the image [5], [15]. The method employs feedback of past optical flow fields and time averaging of brightness to perform predictions of both position and brightness value of any physical point along its trace in $2D$ footage. Design parameters are presented and their influence in the de-flickering process are analysed. The result is a clean imagery with a much more coherent optical flow of the true scene motion. The performance of the algorithm

is demonstrated in many case-studies involving film material of real subaquatic scenarios.

2   Working Hypothesis and Heuristics

Kinematically, wavy fringes can be clearly recognized from specks due to their relative smooth motion in contrast to the quick twinkling behaviour of the scattering. So, sunlight wavy flickers have an optical flow field while specks do not. Also morphologically, sunlight flickers have characteristic strip and ring-like outlines in contrast with the randomly dispersive location of the specks.

On the contrary, morphological characteristics of the caustic waves, like the connectivity existing among many brilliant points, is much more discriminating than the spectra of the brightness gradient. Connected bright pixels differentiate from the underlying background scene by the fact that the scene generally looks rather matte textured, but not dominantly brilliant. Furthermore, the brilliant points of the scattering are relatively small, well spread and generally not connected each other.
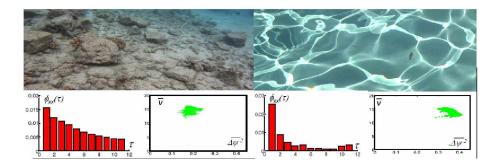

Fig. 1. Comparison of statistic and kinematic characteristics of a moving unperturbed scene (*left*) and a sunlight flicker on a flat landscape (*right*)

Before facing the filter design, it is appropriate to contrast some caustic wave and scene characteristics. They are depicted in Fig. 1. Herein, a video containing a moving scene (at left) and another video of a sunlight flicker evolution over an uniform background (at right) are taken as an illustrative example in order to cut off certain statistic, kinematic and morphological characteristics (see link to video [6]). These frames were arbitrarily picked up from their own videos.

Even when no tool is employed so far to recognize morphologically a flicker, it is totally intuitive that ring-like aureoles might simply be extracted by binarization of the brightness in picture, while the landscape scene does not have a so sharp contrast.

The functions $\phi_{xx}(\tau)$ at bottom represent the autocorrelations of the variable brightness of the central pixel calculated over 500 frames in each video, with

4   Trabes and Jordán

τ being the shift variable. The sharp profile of the autocorrelation stresses the stochastic nature of the flickers. In contrast, the smooth decreasing of the $\phi_{xx}(\tau)$ in the case of the scene points out a mayor predictability of this motion. In both cases, the pixel brightness represents two stochastic variables with spectra spanning a wide and a narrow frequency bands, respectively.

Finally, at bottom, one appreciates the plots of the variables $\overline{v}(x,y)$ vs. $\overline{\Delta\psi^2}(x,y)$, where $\overline{v}$ is the mean value of the module of the optical flow for each pixel (x,y) separately, calculated over the whole video in each case. Equivalently, $\overline{\Delta\psi^2}$ is the mean value of the direction change of the flow at (x,y) during the video in each case, with $\overline{\Delta\psi^2} = \Delta[\Delta(\psi/\Delta t)]\Delta t$, i.e., the second derivative approach of the flow passing through the fixed pixel (x,y) at time t. Employing optical flow field one can depict the loci $[\overline{v}, \overline{\Delta\psi^2}]$, which, in the case of the flicker dynamics, it is shifted to right in the direction of $\overline{\Delta\psi^2}$, and to the left in the case of the movement of the scene. The strong change of direction of the optical flow field is typically for flickers, which evolve randomly in time.

3   Algorithm Description

A block diagram of the algorithm is presented in Fig. 2. One begins the description with the direct path and then the feedback.

A frame of the original image sequence with brightness function $I(x,y,t)$ is compared with an a-priori estimation of the de-flickered image $\widehat{I}(x,y,t)$. The difference yield 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 which have their origin in scattering and high-frequency noise originated in the image processing. Afterwards the image is smoothed with a spatial Gaussian pre-filter with a standard deviation of pixels σ. Here, the small specks will normally disappear. For filtering the specks also an erosion pre-filter will work in like manner. Hereafter, $\Delta I(x,y,t)$ is binarized, choosing for this purpose a threshold value h. In this way the flicker contours enclose relatively large areas.

The a-posteriori estimation of the de-flickered image is based in the following reconstruction

$$\widehat{I_d} = \overline{(I(x,y,t) \cap \Delta I(x,y,t))} \cup \left(\widehat{I}(x,y,t) \cap \Delta I(x,y,t)\right) \qquad (1)$$

it is, the reconstructed frame inherits the sharpness of the original image in the places where there are no fringes. In turn, the places of estimated flickers are filled with the a-priori estimations $I(x,y,t)$.

With the reconstructed image $\widehat{I_d}(x,y,t)$ and the previous one, i.e. $\widehat{I_d}(x,y,t-1)$, one calculates the optical flow field for the time t referred to as $I_{OFF}(x,y,t)$. Particularly in this paper, the algorithm of Farnebäck was employed to estimate the optical flow field (OFF) [4].

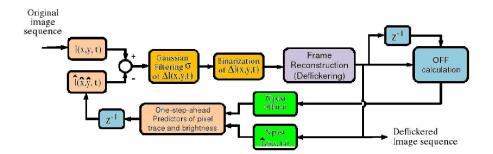With the help of Fig. 3, the steps and operations in the feedback path can be appropriately laid down.

Fig. 2. Real-time deflickering algorithm

The blocks in this path contain N former values of the fields $I_{OFF}(x, y, t-i)$ and of the a-posteriori estimations of the brightness $\widehat{I}_d(x, y, ti)$. On one side, this information is employed to calculate the a-priori estimation $\widehat{I}(x, y, t+1)$ for the next time $t+1$, and on the other side, to estimate the corresponding next link $[\widehat{x}(t+1), \widehat{y}(t+1)]$ of any pixel trace in the period $[t-N+1, t+1]$. To this end two one-step-ahead predictors of order N are put to use. To yield an estimation of $\widehat{I}$, a simple averaging procedure can be employed. Similarly, the last link concatenation for the specific pixel trace $[\widehat{x}, widehaty]$ is complete by obtaining the mean value of N vector coordinates.
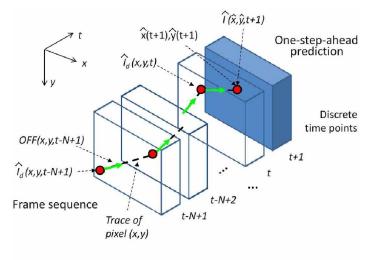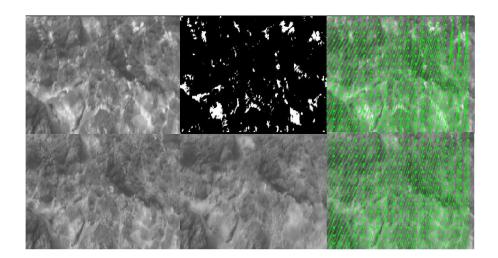

Fig. 3. Prediction of position and brightness

Clearly, the output of the algorithm is the a-posteriori image sequence $I_d(x, y, t)$. By processing the information based on predictions one step ahead, the deflickering algorithm will not produce any delay.

6   Trabes and Jordán

It is worth noticing that the difference between the output sequence $\widehat{I_d}$ and the a-priori image sequence $\widehat{I}(x, y, t)$ rests in the level of sharpness, i.e. $\widehat{I_d}$ largely inherits the sharpness of the original footage and $\widehat{I}$ does not. Nevertheless, as seen later in the case studies, the performance of $\widehat{I}$ seems not to drop to much with respect to $\widehat{I_d}$. However, the quality of $\widehat{I_d}$ shows a little more sensibility with respect to the design parameter h in the binarization, which must be properly tuned up for keeping the superior quality.

4 Case Study

With the goal to illustrate the performance of the method, a case study is investigated. Many of the conclusions exposed in this section were drawn out from the video rather than from the frames selected to perform the description. The specific video contains a moving scene of a camera over an irregular landscape on the sea-floor. In this scenario, only a frame is depicted. A video with the whole sequence can be appreciated in [6]. The video size was 320x240 pixels.

The figures are arranged as follows. On the first row from left to right, the original image I, the binarized error ΔI and the optical flow field I_{OFF} of I are depicted side by side. Down below from left to right, the a-priori and the a-posteriori images, namely I and $\widehat{I_d}$, respectively, and the optical flow field I_{OFF} of $\widehat{I_d}$.

Fig. 4. Case study I: irregular terrain underwater - Top: original image, binarized flickers and optical flow field of the original image. Bottom: a-priori and a-posteriori de-flickered images, and optical flow field of the a-posteriori de-flickered image

In Fig. 4, the illustration of results corresponding to the irregular landscape on the sea-floor are deployed. By comparing I and $\widehat{I_d}$ one notices a practically

complete removal of the flickers. Also the shapes in the binarization error ΔI seems to capture to a great extent the flickers from I in a reliable manner. This result is manifested throughout the video, see [Link to videos]. Besides, there is no appreciable difference between the optical flow fields. This is so because the kinematic of the caustic wave is filtered in the estimation of the function I_{OFF}. This is a typical situation where flickers are overlapped with a scene with many details as in this case. So, the track of points of a wavy flicker does not make sense. On the contrary, the pixels that are traceable are the physical points that belong to the scene only. The situation may be totally different when the scene is flat without texture as shown in the next case.

In this case study, the design parameters were tuned up to the values: $h = 175$, $N = 3$, $\sigma = 9.14$.

5 Conclusions

A real-time approach for de-flickering videos is presented, in where a moving landscape scene was perturbed by sunlight caustic wave. No strong premises to the kinematics are imposed other than a smooth motion of the scene. The approach requires a monocular simple camera. The strategy for filter design was based above all on morphological characteristics of the caustic waves and, to a lesser extend, on differentiation of frequency bands among scene, scattering and processing noise. Basically, it departs from an a-priori de-flickered image to generate an enhanced image, termed as the a-posteriori image. Employing feedbacks of optical flow fields and past a posteriori images, the actual a-priori image is estimated by a one-step-ahead predictor. In this way, the predictor can trace accurately the displacements of any physical point and assign it an estimated bright value. In the direct path of the loop, areas in where the flickers are supposed to be, are determined by gaussianization and binarization processes. Finally, the enhancement of quality for the a-posteriori de-flickered image, result by allowing the brightness be inherited from the original image just in the complementary region of the estimated flicker areas.

The results have proven the quality of the flickering filter in three real videos containing sunlight flickered images. For them it is shown both the high performance and the stability reached by the algorithm in cases of simple as well as complex camera movements.

Though the main motivation of authors were the applications as 3D trajectory recovery for autonomous navigation and SLAM (simultaneous localization and mapping), it is believed that this method can be re-engineered to be useful in other areas of the Computer Vision such as aerial navigation and SLAM.

8   Trabes and Jordán

References

1. R. Chellappa, Qian, G., Srinivasan, S., "Structure from motion: sparse versus dense correspondence methods", in International Conference on Image Processing, 1999.
2. A.J. Davison, I.D. Reid, N.D. Molton, and O. Stasse, "MonoSLAM: Real-Time Single Camera SLAM", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 6, june 2007.
3. R. Dudek, F. Quintana, and C. Cuenca, "Using Optical Flow to Reduce Noise in Image Sequences", in International Conference on Image Processing, Computer Vision, 2009.
4. G. Farnebäck, "Very high accuracy velocity estimation using orientation tensors, parametric motion, and simultaneous segmentation of the motion field", in Proc. Eighth International Conference on Computer Vision, volume 1, pages 171177, Vancouver, Canada. IEEE Computer Society Press, 2001.
5. N. Gracias, S. Negahdaripour, L. Neumann, R. Prados, and R. Garca, "A motion compensated filtering approach to remove sunlight flicker in shallow water images", in Proc. MTS/IEEE Oceans, 2008.
6. Link to videos: https://www.youtube.com/channel/UC2g6YrWOhGcbDRCNeQvS9eg
7. Lourenco, P.; Guerreiro, B.J.; Batista, P.; Oliveira, P.; Silvestre, C., "3-D inertial trajectory and map online estimation: Building on a GAS sensor-based SLAM filter", in Control Conference (ECC), 2013 European, Issue Date: 17-19 July 2013.
8. A. Renyi. Probability Theory. North-Holland, Amsterdam, 1970.
9. P.M.B.V. Roosmalen, R.L. Lagendijk and J. Biemond, J., "Flicker reduction in old film sequences", in Time-varying Image Processing and Moving Object Recognition 4 , Elsevier Science, pp. 917, 1997.
10. F. Sadlo, T. Weyrich, R. Peikert, and M. Gross, "A practical structured light acquisition system for point-based geometry and texture", in Proc. IEEE Eurographics, 2005.
11. D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light", in Proc. IEEE CVPR,2003.
12. Y.Y. Schechner and N. Karpel, "Attenuating natural flicker patterns", in Proc. MTS/IEEE Oceans, 2004.
13. Y. Swirski, Y. Y. Schechner, B. Herzberg, and S. Negahdaripour, "Stereo from flickering caustics", in Proc. IEEE ICCV, 2009.
14. Y. Swirski, Y. Y. Schechner, B. Herzberg, and S. Negahdaripour, "Underwater stereo using natural flickering illumination", in IEEE OCEANS10, Sydney, 24-27 May 2010.
15. E. Trabes and M.A. Jordan, "On-line Filtering of Sunlight Caustic Waves in Underwater Scenes in Motion", in evaluation in 7th International Scientific Conference on Physics and Control. 19-22 August, 2015, Istanbul, Turkey.
16. Z. Zhang, P. Cui and H. Cui, "Recovery of Egomotion from Optical Flow with Large Motion Based on Subspace Method", IEEE International Conference on Robotics and Biomimetics, 2006. ROBIO 06, 17-20 Dec. 2006.