



A Centroid Algorithm for Stabilization of Turbulence-Degraded Underwater Videos

Kalyan Kumar Halder^{1*}, Manoranjan Paul², Murat Tahtali¹, Sreenatha G. Anavatti¹, and Manzur Murshed³

¹School of Engineering and Information Technology, The University of New South Wales, Canberra, ACT 2600, Australia

²School of Computing and Mathematics, Charles Sturt University, Bathurst, NSW 2795, Australia

³School of Information Technology, Federation University, VIC 3842, Australia

*Email: k.halder@unsw.edu.au

Abstract—This paper addresses the problem of stabilizing underwater videos with non-uniform geometric deformations or warping due to a wavy water surface. It presents an improved method to correct these geometric deformations of the frames, providing a high-quality stabilized video output. For this purpose, a non-rigid image registration technique is employed to accurately align the warped frames with respect to a prototype frame and to estimate the deformation parameters, which in turn, are applied in an image dewarping technique. The prototype frame is chosen from the video sequence based on a sharpness assessment. The effectiveness of the proposed method is validated by applying it on both synthetic and real-world sequences using various quality metrics. A performance comparison with an existing method confirms the higher efficacy of the proposed method.

Keywords—Image registration; entropy; shift maps; turbulence mitigation; video stabilization; water waves

I. INTRODUCTION

Light propagation in the water is similar to that through the atmosphere, although, unlike atmospheric imaging, there is the problem of higher light attenuation due to scattering and absorption [1]–[4]. Scattering is due to dissolved organic matter or small suspended biological organisms, which causes a change in the direction of the light beam. In addition, the absorption process causes a reduction in the beam energy. Even if the water is clear, i.e., free of the small observable suspended particles, it is still inhomogeneous due to random temperature and salinity fluctuations [5], [6]. These variations change the density of the water, causing a variation in the refractive index, a dominating factor for underwater turbulence. While imaging under the water, these turbulence effects impose random geometric distortions and non-uniform blur on the acquired images [7]–[11]. As a consequence, underwater imaging poses significant challenges at extended ranges when compared to similar problems in the air.

Several researchers in different fields, including computer vision, computer graphics, and optical engineering have attempted to address the general problems of recovering distorted images due to the water waves. Murase [12] proposed an algorithm to solve the image recovery problem by modelling a three-dimensional (3D) structure of the water waves. In the algorithm, the 3D shape of the water surface is reconstructed using the optical flow fields of the distorted frames. The algorithm assumes that the distance between the object and the water surface is known, and the surface waviness is mild enough so that the short-exposure images pose similar features. Researchers from adaptive optics (AO), such as Holohan

and Dainty [6], proposed an adaptive optical technology for a possible solution of the underwater turbulence problems. Although adaptive optical systems, which are comparatively popular in astronomical imaging, provide real-time corrections of images, the systems are very complex and the cost of full correction in real-time is very expensive. Taking these factors into account, Holohan and Dainty simplified the model and developed a low-cost low-order AO system, which is controlled by a standard personal computer. The complete system is developed based on the assumption that the degradations of the underwater images are predominantly low-frequency low-order aberrations and image motion.

Shefer *et al.* [13] assumed that each point in the object viewed through the wavy water surface oscillates around its true position, i.e., when the water surface is in stand-still condition. Also, the average of the displacements of each point adds up to zero when time tends to infinity. Based on this assumption, the authors proposed a simple method where the restoration of the image of a submerged object is done by taking the time-average of a large number of individual frames captured over a certain period of time. This method can provide a reasonable result when the waviness of the water is low, but with higher energy waves it fails to restore all features of the image, providing a blurry output. Based on the similar assumption, a graph-embedding method was proposed in [14] to reconstruct through-the-water images from a sequence of submerged objects. A shortest path algorithm, assuming the distribution of surface normals of the wavy water to be Gaussian, is employed to select the image having the shortest overall path to all the remaining images. A similar approach was presented in [15], [16] where the entire image is considered as a group of fixed small image patches. Unlike [14], the method compensates for the leakage problem, caused by false shortest-distances among similar but motion-blurred patches, by analysing the motion-blur and the refraction effects separately. Milder *et al.* [17] emphasized first the estimation of the shape of the distorted water surface and then reconstructing the scenes. The algorithm assumes that the water surface follows the linear wave dynamics. The radial surface slope, which is used to estimate the water surface, is deduced along extinction boundaries considering the transmission region and internal reflection region separately, and it is recursively improved by minimizing the quadratic measurement error. The slopes of the estimated surface are then employed to reconstruct the target image by an inverse ray tracing. In [9], a simulator of fluctuations of the water surface was first built using the forward Euler method and

then the original images were recovered using the estimated water surface.

A lucky region technique, a variant of the lucky imaging technique, was proposed in [18] for reconstruction of underwater images. The algorithm divides the whole image frame into an ensemble of smaller image patches similar to [14]–[16], but in this case, instead of the whole image only the image patches of good quality are selected. The good image patches are then processed further to produce a high-quality image. The algorithm was tested for both through-the-water imaging and water-surface-reflection imaging. Kanaev *et al.* [11] presented an underwater image reconstruction approach which includes motion compensation, lucky patch selection and denoising. The motion compensation is performed by an image registration technique, whereas the lucky patches are selected by calculating the nonlinear gain coefficient for each pixel of the prototype-current frame pair by means of an image quality metric. The final reconstructed frame is obtained by fusing the lucky patches. In [10], a two-stage underwater image reconstruction approach was proposed. In the first stage, an iterative non-rigid image registration is employed to register all the input frames, after being pre-processed, against their temporal mean, which in turn, is used to produce a better-structured sequence. The second stage involves extracting the unstructured sparse errors from the sequence through rank minimization. Corrections for geometric distortions have been tackled reasonably successfully by previous efforts, but they still need improvements, especially in restoration accuracy.

In this paper, a centroid-based image restoration approach is proposed to correct the geometrically distorted frames of the underwater videos. An efficient image registration technique is employed to accurately estimate the motion vector fields, in other words, the pixel shift maps of all warped frames against a reference frame. A blind image quality metric is used to find out the sharpest frame from the given sequence, which is used as the reference frame in this study. The centroid of the estimated shift maps is then calculated and processed further to generate correct shift maps to restore all the input frames. A single iteration may not always be sufficient to restore the input frames to their original positions as the turbulence is generally strong in underwater imaging. Therefore, an iterative scheme is considered to achieve higher quality and stability of the dewarped sequence. The performance of the proposed method is compared against a state-of-the-art method [10] by applying them to both synthetic and real-world video sequences.

The rest of the paper is organized as follows: Section II provides the details of the proposed video stabilization method. In Section III, experimental evaluation of the proposed method and its performance on both synthetically distorted and real-world underwater videos are analysed using various quality metrics. Finally, Section IV concludes the paper and provides some future research directions.

II. PROPOSED VIDEO STABILIZATION METHOD

According to Cox-Munk law [19], in the case when the water surface is sufficiently large and the waves are stationary, the normals of the water surface, therefore, the displacements of the points on the object are approximately Gaussian distributed centred around their true locations. In the proposed

video stabilization method, this law is taken into consideration, thereby, it is assumed that the wander of each pixel of an image over a certain period of time oscillates around its true position, i.e., the mean wander of the pixel would be zero. In a different viewpoint, if the pixel's wander is measured with respect to any fixed position, the mean of its wander will represent its shift from the true position. This shift can be then used to restore the pixel to its true position.

Now, instead of a single pixel, if a whole frame is considered, there needs to have a reference frame to calculate the pixel shift maps of all the input frames against it. The simplest way could be calculating the mean of the warped frames and use it as the reference frame. Although a mean frame is somewhat geometrically stable, it contains motion-blur, thus, cannot provide a high-quality output [20], [21]. In this method, one of the warped frames from the whole video sequence is considered as the reference frame. There are three options to choose a reference frame in this approach: (1) first frame, (2) any random frame, and (c) the sharpest frame. Although the first two options do not involve too much processing time for the reference frame selection, the output video may not be of high-quality if the frame suffers from blur or noise [8]. Therefore in this study, a blind image quality metric is employed that calculates the sharpness of all input frames and use the values to choose the reference frame.

Suppose that $\bar{R}(t, \theta_s)$ is the expected value of entropy for image $t \in [1, 2, \dots, M]$, measured in directions $\theta_s \in [\theta_1, \theta_2, \dots, \theta_S]$. Here M and S represent the image size and the number of directions, respectively. The metric employed to select the sharpest frame is [22]

$$Q = \sqrt{\sum_{s=1}^S (\mu_t - \bar{R}(t, \theta_s))^2 / S}, \quad (1)$$

where μ_t is the mean of the values $\bar{R}(t, \theta_s)$. The frame that poses higher value of Q is deemed to be the sharpest frame.

In the next step, a non-rigid image registration technique [23] is employed to estimate the pixel shift maps (s_x and s_y for x - and y - directions, respectively) of all the input frames with respect to the reference frame. A simplified block diagram of the proposed video stabilization method is shown in Fig. 1. Using the individual shift maps, their centroid is calculated as

$$\begin{aligned} p_x(x, y) &= \frac{\sum_{k=1}^K s_x(x, y, k)}{K}, \\ p_y(x, y) &= \frac{\sum_{k=1}^K s_y(x, y, k)}{K}, \end{aligned} \quad (2)$$

where p_x and p_y are the centroid shift maps, k is the frame index, and K is the total number of frames. According to the initial assumption, this shift map represents the actual shift of the reference frame from its ground truth, therefore, it can be used to warp the frame back into its unwarped position through an interpolation technique. For all other frames, the actual shift maps are calculated using the centroid shift map and individual shift maps as

$$\begin{aligned} s_x^*(x, y, k) &= p_x(x, y) + s_x(x + p_x(x, y), y + p_y(x, y), k), \\ s_y^*(x, y, k) &= p_y(x, y) + s_y(x + p_x(x, y), y + p_y(x, y), k). \end{aligned} \quad (3)$$

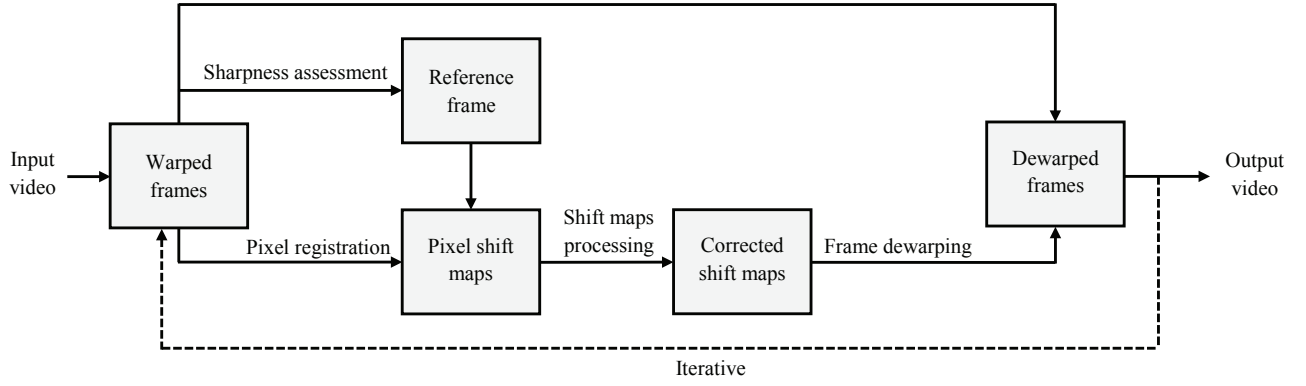


Fig. 1: Various steps of the proposed video stabilization method.

where s_x^* and s_y^* are the corrected shift maps. Using these shift maps, the warped frames can be restored as

$$I^*(x, y, k) = I(x + s_x^*(x, y, k), y + s_y^*(x, y, k), k), \quad (4)$$

where I^* is the dewarped version of the warped frame I .

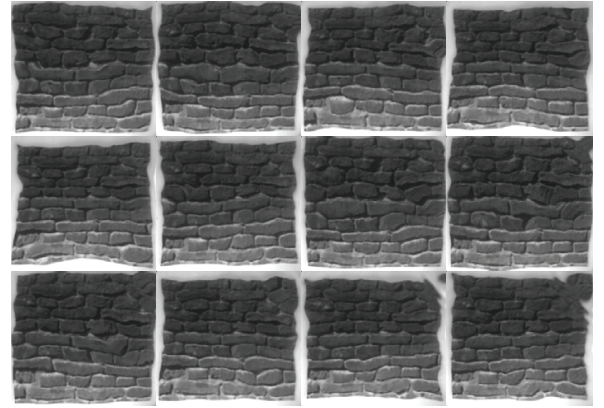
The dewarped sequence, after the first iteration, is more stabilized than the original one. However, it may require few more iterations to gain higher accuracy and to be more stabilize. The dewarped sequence is, therefore, given as input to the next iteration and repeated generally 2 to 4 times. The dewarped version of the reference frame after an iteration is used as the reference frame in the next iteration.

III. SIMULATION EXPERIMENTS

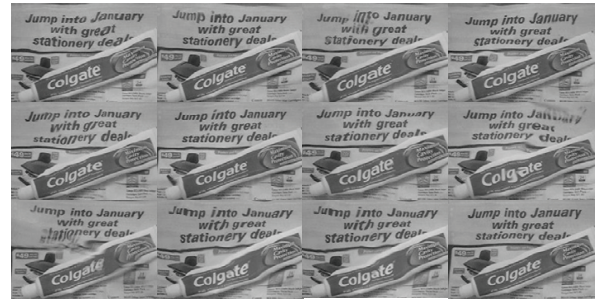
Experimental evaluation was carried out to verify the performance of the proposed method and to compare it against a state-of-the-art method, namely the Oreifej method [10]. Both the proposed and the Oreifej methods employ image registration to calculate the pixel shift maps that are later used to dewarp the frames. The main difference between the two methods is in the way of calculating the pixel shift maps. The Oreifej method uses the mean of the input frames as the reference frame, blurs the input frames with the same blur presented in the reference frame, and registers them against the reference frame to get the pixel shift maps. On the other hand, the proposed method uses the sharpest warped frame as the reference frame, registers the input frames directly against it, and uses the mean of the individual pixel shift maps to calculate the accurate ones. In short, both methods rely on averaging - the Oreifej method does the averaging of the frames, whereas the proposed method does the averaging of the pixel shift maps avoiding the blurring operation, thus provides more accurate estimation of the pixel shift maps.

The two methods were first applied on a synthetically distorted video sequence, called the *brick* sequence. The sequence consists of 60 frames, each of 268×292 pixels. The underwater turbulence effect was added to the video using a spatial distortion model, which is developed based on Snell's law of refraction and the wave equation [9]. For a fair comparison, the same image registration was used for both methods. Along with visual comparison, various image quality metrics are used to quantitatively evaluate the performance of the methods on the synthetic sequence. Later on, both methods

were applied on a real-degraded video sequence, namely the *colgate* sequence [18]. This sequence was captured in the laboratory by a video camera fixed above the water surface. A still object was laid on the planner ground under the clean water. Waves were generated by wind produced by a fan, while additional waves were created by dropping an object into the water. The sequence contains 60 frames, each of 192×288 pixels. A sampling of the two test sequences is shown in Fig. 2. The synthetic sequence contains brick textures and the real-world sequence contains different sizes of text fonts, which provides a good means of demonstrating underwater turbulence effects clearly, as can be seen from Fig. 2.



(a) brick sequence [9]



(b) colgate sequence [18]

Fig. 2: A sampling of the test video sequences (one in every 5 frames).

A. Quality Metrics

While the final arbiter of image quality is the human viewer, efforts have been made to create objective measures of quality. These objective measures require the existence of a distortion-free copy of an image, called the ground-truth image, which can be used for comparison with the image whose quality is to be measured.

1) *Mean square error (MSE)*: The MSE is a frequently used measure of how close the restored frame is to the ground-truth frame. It can be defined as

$$\text{MSE} = \frac{1}{m \times n} \sum_{x=1}^m \sum_{y=1}^n (R(x, y) - G(x, y))^2, \quad (5)$$

where R and G are any restored frame and the ground-truth frame, respectively. $m \times n$ represents the size of each frame.

2) *Point spread function (PSF)*: The PSF describes, as the name implies, the response of an imaging system to a point source or point object. It is basically the image of a bright spot. The sharper and narrower the PSF of an imaging system the better it is at resolving finer detail. In order to use PSF as a quality measure for the restored frames, it is calculated as

$$\text{PSF}(x, y) = \mathcal{F}^{-1} \left(\frac{\mathcal{F}(R(x, y))}{\mathcal{F}(G(x, y))} \right), \quad (6)$$

where \mathcal{F} denotes the Fourier transform and \mathcal{F}^{-1} its inverse.

B. Performance Analysis and Comparison

The performance of the proposed method is compared against the earlier Oreifej method [10] using both qualitative and quantitative analyses. The ground-truth frame, which was used to synthetically generate the video sequence, is shown in Fig. 3(a). The mean of all the frames of the sequence is shown in Fig. 3(b). This mean frame suffers from extreme motion-blur since the frames are distorted by the water waves that cause random shifting of the pixels. The means of the restored frames using the Oreifej method and the proposed method are shown in Fig. 3(c) and 3(d), respectively. A mean frame is used here to show the video stabilization performance as it is a good measure of how well the frames are stabilized and close to each other. The more the mean frame is sharp, the more the frames are geometrically aligned. It is evident from Fig. 3(c) and 3(d) that both methods compensate for the geometric distortions to some extent. However, the Oreifej method is less effective at doing so, which is also reflected in its difference image (with respect to the ground truth frame) in Fig. 3(e) where the lighter regions indicate higher residuals. For the proposed method, the restored frames are more geometrically aligned, therefore, provides a mean frame closer to the ground-truth. Also, its difference image in Fig. 3(f) has lesser artefacts than Fig. 3(e), which is also indicative of the better geometric distortion-correction capability of the proposed method. The MSEs of the restored mean frames for the Oreifej method and the proposed method are calculated as 331 and 195, respectively. In order to show the performance of the methods on the whole sequence, the MSEs of all the frames before and after restoration are also plotted in Fig. 4. As can be seen, the proposed method provides better corrections for each individual frame with a consistently lower MSE throughout the sequence.

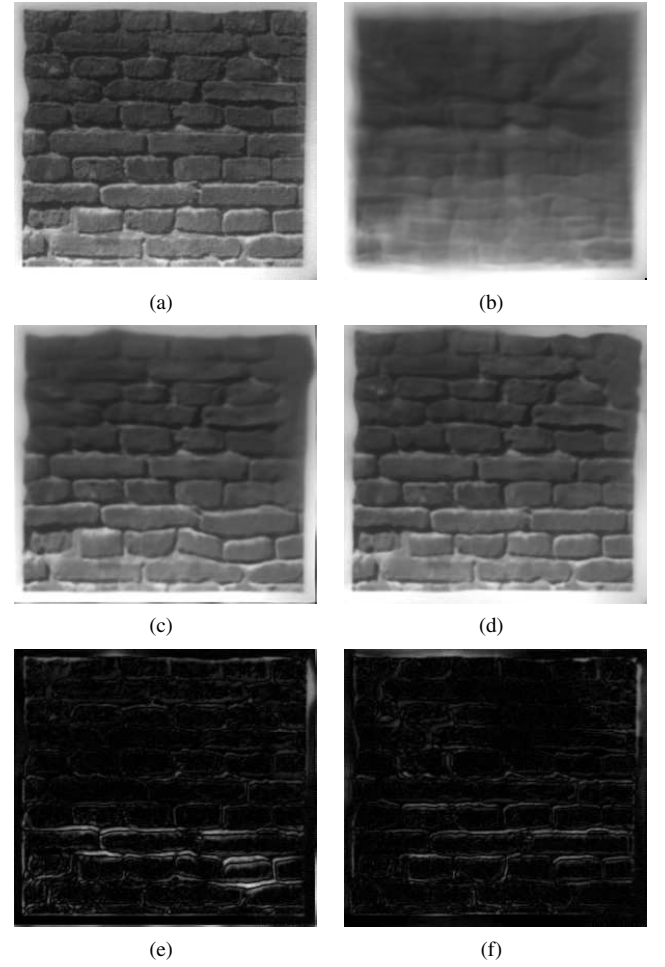


Fig. 3: Restoration results on *brick* sequence: (a) ground truth frame, (b) mean of warped frames, (c) mean of restored frames using Oreifej method, (d) mean of restored frames using proposed method, (e) difference image between (a) and (c), and (f) difference image between (a) and (d).

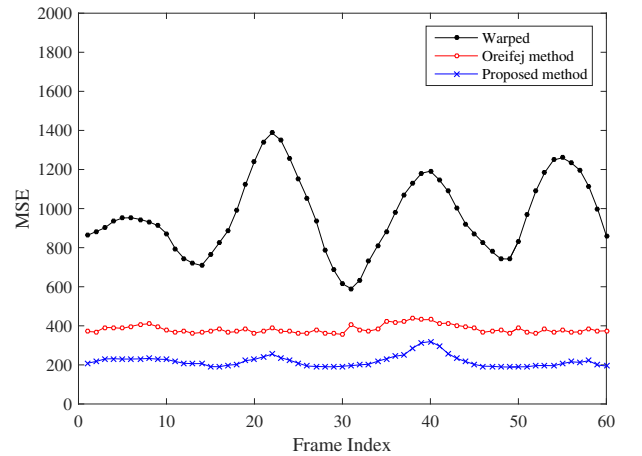


Fig. 4: MSE plot for all frames of *brick* sequence before and after applying stabilization methods.

Furthermore, Fig. 5 shows the PSF plot for the mean frames before and after restoration. As expected, the PSF of the mean of the warped input frames in Fig. 5(a) is indistinct without any notable peak at its centre. In the case of the Oreifej method, although it could provide a reasonable restoration of the warped frames, it is not very distinguishable from the PSF, as it only sports a weak raise at the centre of the PSF shown in Fig. 5(b). On the other hand, the proposed method's PSF as shown in Fig. 5(c) has a distinct peak at the centre, although not as sharp as to indicate perfect correlation, the peak is significantly higher with less activity away from the centre.

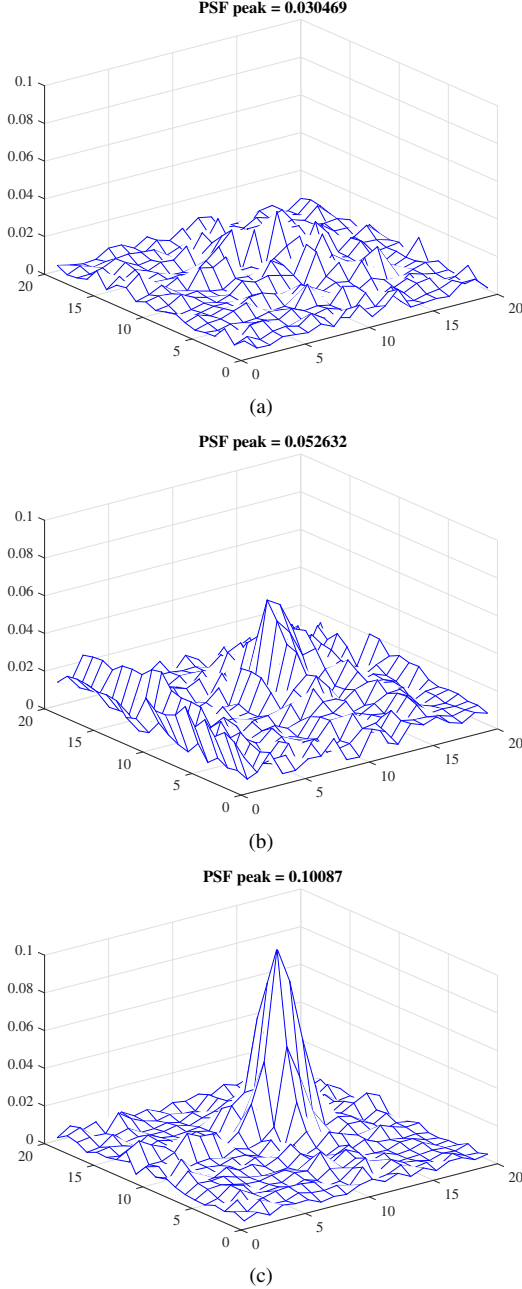


Fig. 5: PSF plot for *brick* sequence: (a) mean of warped frames, (b) mean of restored frames using Oreifej method, and (c) mean of restored frames using proposed method.

After verifying the performance of the two methods on the synthetic sequence, they were applied on the real-world *colgate* sequence [18]. Since the ground truth frame is not available in this case, a sample warped frame is shown in Fig. 6(a) in order to aid the visual comparison. The mean of the input warped frames is shown in Fig. 6(b), which is also blurry for this sequence with a loss of details. The mean of the restored frames using the Oreifej method and the proposed method are shown in Fig. 6(c) and 6(d), respectively. It can be noted that the Oreifej method improves the geometric alignments at a cost of considerable blurring in the restored frame. The proposed method's output seems fully restored with no significant blurring and distortions, which is further confirmed by the better legibility of the word 'stationery' in it compared to the other frames in Fig. 6(a) to 6(c). Finally, an object comparison is made for these results using the blind image quality metric Q , defined in (1). The value of Q for the mean frame restored using the Oreifej method is calculated as 0.0018. In the case of the proposed method, Q is significantly higher with a value of 0.0056, indicating its higher geometric correction fidelity.

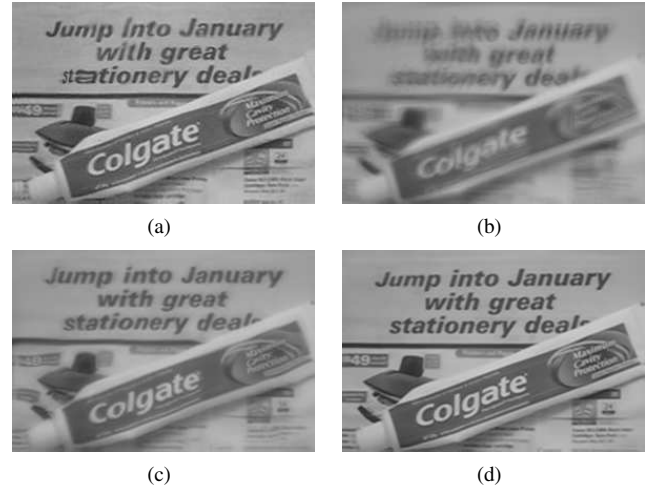


Fig. 6: Restoration results on *colgate* sequence: (a) sample frame, (b) mean of warped frames, (c) mean of restored frames using Oreifej method, and (d) mean of restored frames using proposed method.

IV. CONCLUSION

An efficient geometric distortion-correction method using a highly accurate motion estimation technique has been implemented to stabilize video sequences degraded by underwater turbulence. The proposed scheme demonstrates a convincing way of calculating the nearly accurate pixel shift maps of the input warped frames that could be used to restore the frames into their truth position. The proposed method outperforms a state-of-the-art method both for synthetic and real-world videos, especially, in terms of accuracy. In its current form, only geometrical correction of the frames is of concern, thus, the future aim would involve the consideration of the presence of other imaging problems, such as localized blur and CCD noise, and extend the method to address those problems efficiently.

REFERENCES

- [1] A. Galdran, D. Pardo, A. Picón, and A. Alvarez-Gila, "Automatic Red-Channel underwater image restoration," *Journal of Visual Communication and Image Representation*, vol. 26, pp. 132–145, 2015.
- [2] W. Hou, "A simple underwater imaging model," *Optics Letters*, vol. 34, no. 17, pp. 2688–2690, 2009.
- [3] M. Boffety, F. Galland, and A.-G. Allais, "Influence of polarization filtering on image registration precision in underwater conditions," *Optics Letters*, vol. 37, no. 15, pp. 3273–3275, 2012.
- [4] G. Wang, B. Zheng, and F. F. Sun, "Estimation-based approach for underwater image restoration," *Optics Letters*, vol. 36, no. 13, pp. 2384–2386, 2011.
- [5] R. Schettini and S. Corchs, "Underwater image processing: State of the art of restoration and image enhancement methods," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, p. 746052, 2010.
- [6] M. L. Holohan and J. C. Dainty, "Low-order adaptive optics: A possible use in underwater imaging?," *Optics & Laser Technology*, vol. 29, no. 1, pp. 51–55, 1997.
- [7] J. Gilles and S. Osher, "Wavelet burst accumulation for turbulence mitigation," *Journal of Electronic Imaging*, vol. 25, no. 3, p. 033003, 2016.
- [8] K. K. Halder, M. Tahtali, and S. G. Anavatti, "Simple algorithm for correction of geometrically warped underwater images," *Electronics Letters*, vol. 50, no. 23, pp. 1687–1689, 2014.
- [9] Y. Tian and S. G. Narasimhan, "Seeing through water: Image restoration using model-based tracking," in *Proc. IEEE International Conference on Computer Vision (ICCV)*, pp. 2303–2310, 2009.
- [10] O. Oreifej, S. Guang, T. Pace, and M. Shah, "A two-stage reconstruction approach for seeing through water," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1153–1160, 2011.
- [11] A. V. Kanaev, W. Hou, S. Woods, and L. N. Smith, "Restoration of turbulence degraded underwater images," *Optical Engineering*, vol. 51, no. 5, p. 057007, 2012.
- [12] H. Murase, "Surface shape reconstruction of a nonrigid transport object using refraction and motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 10, pp. 1045–1052, 1992.
- [13] R. Shefer, M. Malhi, and A. Shenhar, "Waves distortion correction using crosscorrelation," <http://visl.technion.ac.il/projects/2000maor/>.
- [14] A. Efros, V. Isler, J. Shi, and M. Visontai, "Seeing through water," in *Proc. Neural Information Processing Systems (NIPS)*, 2004.
- [15] A. Donate, G. Dahme, and E. Ribeiro, "Classification of textures distorted by water waves," in *Proc. International Conference on Pattern Recognition*, pp. 421–424, 2006.
- [16] A. Donate and E. Ribeiro, "Improved reconstruction of images distorted by water waves," in *Advances in Computer Graphics and Computer Vision* (J. Braz, A. Ranchordas, H. Araújo, and J. Jorge, eds.), vol. 4 of *Communications in Computer and Information Science*, pp. 264–277, Springer Berlin Heidelberg, 2007.
- [17] D. M. Milder, P. W. Carter, N. L. Flacco, B. E. Hubbard, N. M. Jones, K. R. Panici, B. D. Platt, R. E. Potter, K. W. Tong, and D. J. Twisselmann, "Reconstruction of through-surface underwater imagery," *Waves in Random Complex Media*, vol. 16, no. 4, pp. 521–530, 2006.
- [18] Z. Wen, A. Lambert, D. Fraser, and H. Li, "Bispectral analysis and recovery of images distorted by a moving water surface," *Applied Optics*, vol. 49, no. 33, pp. 6376–6384, 2010.
- [19] C. Cox and W. Munk, "Slopes of the sea surface deduced from photographs of sun glitter," *Bulletin of the Scripps Institution of Oceanography*, vol. 6, no. 9, pp. 401–479, 1956.
- [20] M. Tahtali, D. Fraser, and A. J. Lambert, "Restoration of non-uniformly warped images using a typical frame as prototype," in *Proc. IEEE Region 10 Conference (TENCON)*, pp. 1–6, 2005.
- [21] K. K. Halder, M. Tahtali, and S. G. Anavatti, "Simple and efficient approach for restoration of non-uniformly warped images," *Applied Optics*, vol. 53, no. 25, pp. 5576–5584, 2014.
- [22] S. Gabarda and G. Cristóbal, "Blind image quality assessment through anisotropy," *Journal of the Optical Society of America A*, vol. 24, no. 12, pp. B42–B51, 2007.
- [23] A. Myronenko and X. Song, "Intensity-based image registration by minimizing residual complexity," *IEEE Transactions on Medical Imaging*, vol. 29, no. 11, pp. 1882–1891, 2010.