A Two-Stage Reconstruction Approach for Seeing Through Water

Omar Oreifej[1], Guang Shu[1], Teresa Pace[2], and Mubarak Shah[1]

[1]Computer Vision Lab, University of Central Florida, Orlando, FL

[2]Night Vision and Electronic Sensors Directorate, Orlando, FL

oreifej@eecs.ucf.edu, gshu@eecs.ucf.edu, teresa.pace1@us.army.mil, shah@eecs.ucf.edu

Abstract

Several attempts have been lately proposed to tackle the problem of recovering the original image of an underwater scene using a sequence distorted by water waves. The main drawback of the state of the art [18] is that it heavily depends on modelling the waves, which in fact is ill-posed since the actual behavior of the waves along with the imaging process are complicated and include several noise components; therefore, their results are not satisfactory. In this paper, we revisit the problem by proposing a data-driven two-stage approach, each stage is targeted toward a certain type of noise. The first stage leverages the temporal mean of the sequence to overcome the structured turbulence of the waves through an iterative robust registration algorithm. The result of the first stage is a high quality mean and a better structured sequence; however, the sequence still contains unstructured sparse noise. Thus, we employ a second stage at which we extract the sparse errors from the sequence through rank minimization. Our method converges faster, and drastically outperforms state of the art on all testing sequences even only after the first stage.

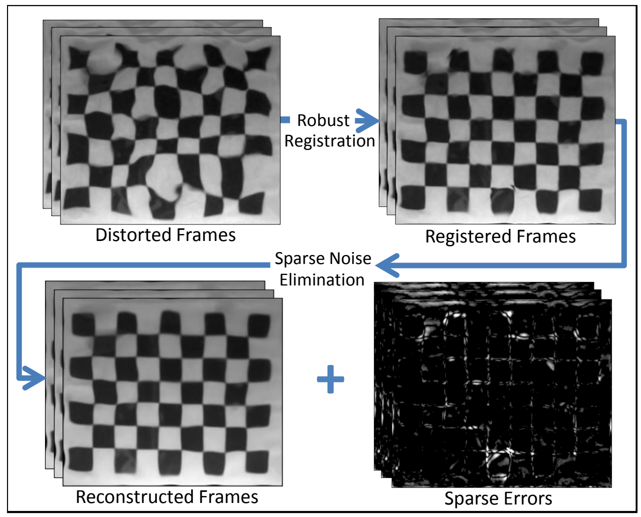

Figure 1. Two-Stage reconstruction of an underwater scene. Stage 1 (Robust Registration) aligns the sequence and overcomes the structured turbulence, while stage 2 (Sparse Noise Elimination) extracts the remaining unstructured random noise.

1. Introduction

Light rays reflected from objects go through several perturbations before being captured by the camera. If the rays pass through different media, they get affected by a complex series of reflections and refractions which can cause extreme distortion to the image. Imaging through water is an example of such a scenario, where reconstructing the original underwater scene still constitutes a big challenge. A fluctuating water surface poses significant difficulties to the process of image recovery mainly because the fluctuations tend to be high, random, and exhibit a complicated behavior especially near the edges of the container. In this context, traditional sparse or dense point correspondence and tracking methods [2, 14, 12, 16, 22], which could have been employed to learn the deformation function of the water, are in fact rendered unusable for three reasons: First, tracking over long periods is difficult in such a turbulent video; second, a noise-free template of the underwater scene is unavailable; third, even using a frame from the distorted video as a template for tracking will not capture the uninvertable distortion function of the water [18]. On the other hand, as a result of the high dimensionality and the embedded randomness of the waves, the techniques which attempt to model the distribution of intensity in the sequence usually suffer, starting from pixel-wise mean/median, to patch-wise clustering [4, 5, 6], and fourier-based averaging [20]. Fortunately, the latest advances in this area such as [19] provided evidence that a better solution for such a problem can be obtained using methods with combined generative and discriminative properties. In this work, we follow [19], and propose a generative-discriminative approach to robustly reconstruct the underwater sequence, however, without requiring the template as in [19].

Our technique is based on registering the frames to their temporal mean which is close to the original undistorted scene. However, the mean of an underwater sequence is blurry and noisy; therefore, a standard frame to mean non-rigid registration faces considerable challenges. The straightforward workaround is to deblur the mean; yet, the deconvolution operation involved in deblurring is generally ill-posed; therefore, even the latest deblurring techniques such as [21, 8] often fail and introduce undesirable edges. In this article, we show that registration-based reconstruction of an underwater scene could be greatly improved by a modified approach that we refer to as robust registration. The robust registration includes two advances: First, iterative refinement of the mean and the frames through iterative registration; second, aided mean-frame correlation step at which we blur the sharp frame instead of deblurring the blurry mean by estimating a blur kernel that brings the frame to the same blur level of the mean.

In a few iterations, the robust registration will converge to a reconstructed mean and a new sequence which are free from most of the structured turbulence of the water waves. However, many frames will still contain unregistered components caused by three types of random noise: First, light reflection on the water surface; second, occlusion of the underwater scene; third, the random behavior of the waves. Such unstructured random noise is however sparse and has a direct correlation with the rank of the matrix composed of the registered frames. Therefore, we employ a second stage of denoising at which we eliminate the sparse noise through convex rank optimization [11]. The result is a clear underwater sequence with significantly reduced noise. Figures 1 and 2 illustrate our two-stage framework which combines the powers of registration and sparse representation. The rest of the paper is organized as the following: In the next section, we discuss the most related works. Consequently, in section three, we present the first stage of our algorithm which is the robust registration, followed by the second stage which is the sparse noise elimination in section four. The experiments and the results are illustrated in section five. Finally, section six concludes the paper.

2. Related Work

Earlier work in reconstructing an underwater scene without a template focused on finding the center of the distribution of patches in the video through clustering [4, 5], and manifold embedding [6], or employing the bispectrum to average the frames in the fourier domain [20]. The state of the art in this area, however, is the model-based tracking [18], where the characteristics of the water waves were employed to estimate the water basis using PCA. In such work, the optimal number and size of the basis remain vague; therefore, we argue that the estimated basis can be underfitted or over-fitted depending on the selected basis. Other than requiring an orthographic camera and a given water height, the basis are additionally obtained by simulation using a single parameter differential equation for the waves, with the assumption that the surface fluctuations are small compared to the water's height. Such a simple model with low parameter space is quite limited, and does not fully represent the actual scenario which can be much more complicated and dependant over several other factors such as the container size, external forces, and camera calibration; hence, the results from [18] are not quite satisfactory. Later in [19], Tian and Narasimhan proposed that the large nonrigid distortion caused by effects such as the water waves cannot be overcome through traditional B-spline registration, but rather through a pull-back operation that utilizes several images with known deformations. While the distorted video is typically the only available piece of information in such a problem, [19] assumes additional given training images with predefined deformation, or a template from which we can generate samples; therefore, their work is considered out of the scope of comparison with our method.

Sparse representation-based video denoising has recently flourished with several successful works reported, from which we only discuss the most related articles. Robust PCA was proposed in [3], where a low rank matrix was recovered from a small set of corrupted observations through convex programming. Similar concepts were later employed in [7] for video denoising, where serious mixed noise was extracted by grouping similar patches in both spatial and temporal domains and solving a low-rank matrix completion problem. Additionally, in [13], linear rank optimization was employed to align faces with rigid transformations, and concurrently detect occlusions. Such works provided enough evidence that low rank optimization can be successfully applied to extract the errors from a sequence as long as the errors are sparse. Therefore, in this paper we show that sparsity-based denosing can also be applied to recover the underwater scene if water turbulence was first "sparcified" through registration.

In the context of blur kernel estimation which is used in our robust registration stage, the latest advances focused on deblurring a single image [21, 8], or a motion blurred video [10, 1]. However, our problem layout is different in that the underwater sequence is not blurry, but its mean is extremely blurry and noisy such that it cannot be deblurred. Thus, we are interested in rather blurring the frames in order to aid the registration. For those reasons, we propose to estimate a spatially varying blur kernel which encodes the difference in blur between the mean and the frames by using the motion estimated at each iteration of our algorithm. Our robust registration is not very sensitive to the frame blurring operation; therefore, in principal, other good blur estimation algorithms could be employed.

We propose a generic, simple, and robust method which in contrast to the pervious methods, does not require a known template such as [19], the camera's height [18], or a special illumination [9]. Our method is similar to the state the art [18] in that it works on a short sequence (61 frames) rather than 800 in [6] and 120 in [20], but more importantly, superior to [18] in performance and processing time.

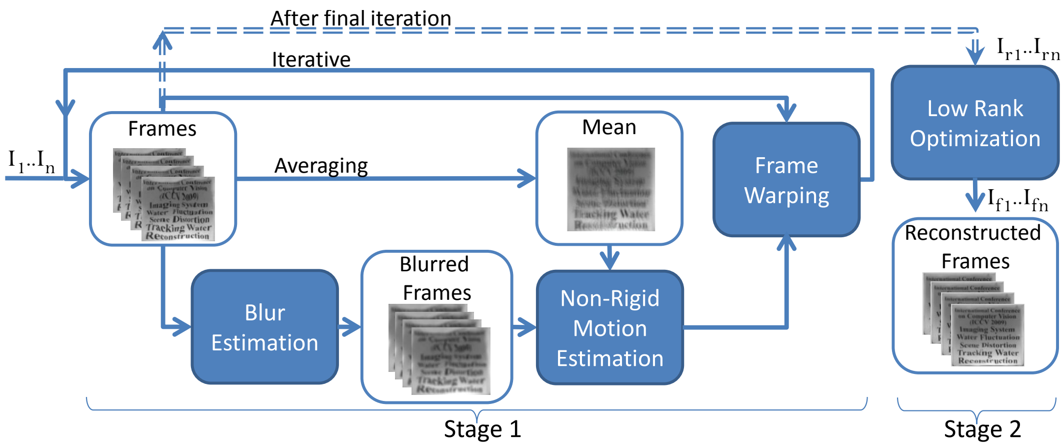

Figure 2. The various steps for seeing through water.

3. Robust Registration

For a video sequence $V \in \mathbb{R}^{w \times h \times n}$ of an underwater scene which consists of the frames $\{I_1...I_n\}$, where each frame I_i is distorted by an unknown water deformation $\mathbf{\Gamma}(\mathbf{x})$, our goal in this paper is to recover a new wave-free sequence $V_f = \{I_{f1}...I_{fn}\}$.

We refine both the mean and the sequence in an iterative registration operation, where at each iteration the frames are shifted closer toward to the correct mean (stage 1 in figure 2). We start by computing the temporal mean M; consequently, each frame is shifted to the mean through registration, thus generating a new sequence V_2. Since the mean is noisy, V_2 is not well registered; however, its mean M_2 is now better and shifted closer to the true image underwater. Therefore, we re-register V_2 to M_2 generating the sequence V_3 and its mean M_3. We keep on performing this process for a few iterations until we settle on a robust mean and a better registered sequence. The computed mean at each iteration is blurry; thus, to improve this operation we use the motion estimated from the previous iteration to compute a blur kernel that brings the frames to the same blur level of the mean. The details about that will be discussed in the following subsection. Once the frames are blurred, they are used to compute the motion; however, the computed motion is then applied to warp the original unblurred frames.

The core motion estimation is a standard B-spline non-rigid registration and is similar to [15]. A frame I is registered to the mean M by estimating a motion vector $\mathbf{\Gamma}$ for each point $\mathbf{x} = [x, y]^T$ in I using a grid of control points of size $s_x \times s_y$

$$\Gamma(\mathbf{x}) = \sum_{m=0}^{3} \sum_{l=0}^{3} B_m(v) B_l(u) \Phi_{i+l,j+m}, \qquad (1)$$

where $i = \lfloor x/s_x \rfloor - 1$, $j = \lfloor y/s_y \rfloor - 1$, $u = x/s_x - \lfloor x/s_x \rfloor$, and $v = y/s_y - \lfloor y/s_y \rfloor$. $\Phi_{i,j}$ is the motion vector of the ij-th control point on the grid, and B is a standard B-spline basis function which defines the weights of the control points contributing in the motion of $\mathbf{x}$

$$\begin{aligned} B_0(t) &= (1-t)^3/6, \qquad (2)\\ B_1(t) &= (3t^3 - 6t^2 + 4)/6,\\ B_2(t) &= (-3t^3 + 3t^2 + 3t + 1)/6,\\ B_3(t) &= t^3/6. \end{aligned}$$

The B-spline weights are pre-computed for every image location given the image size and the spacing between the control points. Φ in equation 1 is estimated similar to [15] by minimizing the difference in the normalized mutual information between the mean M and the frame I that is warped by Φ, in addition to a smoothness constraint. At each iteration of our algorithm, the first frame of the sequence is registered using three refinement levels; starting from a coarse 16×16 grid, then 24×24, and ending with 32×32, where the motion is hierarchically transferred between the levels. The rest of the frames are registered using one level of 32×32 grid as their motion parameters are initialized using the corresponding previous frames; for instance, Φ_2 is initialized as Φ_1. We found such grid sizes adequate for our sequences (frame size of $\sim 256 \times 256$), but they could also be set adaptively [15].

At their current condition, the mean and the frames are at different levels of sharpness since averaging the frames introduces severe blur to the computed mean. In the following, we discuss how we aid the mean-frame correlation through blur kernel estimation.

3.1. Frame Blurring

The frame to mean registration process is impeded by the severely blurred mean. However, we found in our experiments that since deblurring the mean does not work, blurring the frame can be employed instead. The intuition behind this is that once the blurry regions in the mean are also blurred in the frame, the registration process is accordingly guided to focus on the sharper regions of the mean rather than the corrupted blurry regions; thus, improving the overall performance. The required blurring depends on several factors such as the amplitude of the waves and the depth of the water. Such factors are hard to model precisely; however, their effect on the sequence is obvious which is the induced motion. Since the distribution of the positions which a certain tracked point $\mathbf{x}$ attains along the underwater sequence can be approximated with a Gaussian [6], the blur which is induced by the motion can also be approximated with a similar Gaussian. Therefore, without loss of generality, we assume that the blur kernel will be a 2D Gaussian centered at $\mathbf{x}$ with a covariance matrix equals to the covariance matrix of the tracked positions of $\mathbf{x}$ along the sequence

$$\Sigma(\mathbf{x}) = \begin{bmatrix} \sigma_x^2(\mathbf{x}) & \sigma_{xy}(\mathbf{x}) \\ \sigma_{xy}(\mathbf{x}) & \sigma_y^2(\mathbf{x}) \end{bmatrix}, \tag{3}$$

where

$$\sigma_x^2(\mathbf{x}) = \frac{\sum_{f=1}^{n}(x - x_f)^2}{n}, \quad \sigma_y^2(\mathbf{x}) = \frac{\sum_{f=1}^{n}(y - y_f)^2}{n},$$
$$\sigma_{xy}(\mathbf{x}) = \frac{\sum_{f=1}^{n}(x - x_f)(y - y_f)}{n},$$

where $(x_f, y_f)^T$ is the location of $\mathbf{x}$ at frame f. Since we have already estimated a dense motion field through registration, we do not need to further track the points, we can rather directly replace $(x - x_f)$ and $(y - y_f)$ with $\Gamma_x(\mathbf{x})$ and $\Gamma_y(\mathbf{x})$ at frame f, which are the x and y components of the estimated motion. Therefore, at each registration iteration k, we use the motion from the previous iteration $\Gamma^{k-1}(\mathbf{x})$ to compute the covariance of the blur kernel

$$\Sigma(\mathbf{x}, k) = \frac{1}{k^2 n} \begin{bmatrix} \sum_f \Gamma_x^2 & \sum_f \Gamma_x \Gamma_y \\ \sum_f \Gamma_x \Gamma_y & \sum_f \Gamma_y^2 \end{bmatrix}, \tag{4}$$

where we introduce an additional damping factor k^2 which will make sure that the blur reduces along the iterations. In the first iteration, we find $\Gamma^0(\mathbf{x})$ by an extra registration iteration without frame blurring. Once the kernel is computed, it is used to blur the frames using a fixed-size filter which is large enough to account for the maximum possible motion (we use a 20×20 filter). Figure 3 shows our blur estimation for two example sequences from [18]. It can be clearly seen from the figure that the determinant is high over the blurry regions of the mean; thus, the corresponding regions in the frame were blurred accordingly. Generally, the blur in the mean of an underwater sequence is less in the middle regions than the borders, such phenomenon is well captured by our spatially varying filter as illustrated in the figure.

4. Sparse Noise Elimination

Applying the robust registration removes most of the distortion caused by the high fluctuations of the waves and generates a better mean image. However, the frames still include several unstructured and random noise caused by reflections and occlusions which the non-rigid registration cannot handle. Interestingly, after applying the first stage, the frames become generally aligned such that their difference can be considered as a sparse error. Therefore, through

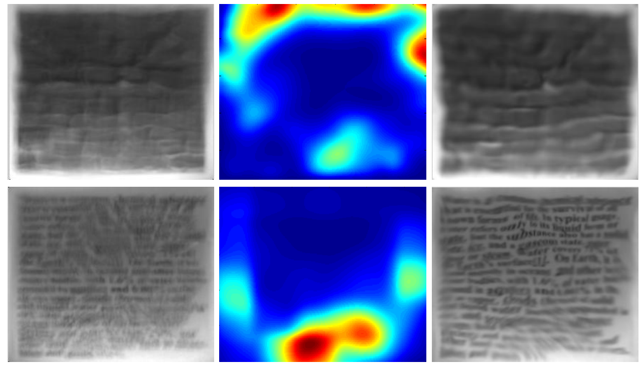

Figure 3. The estimated spatially varying filter at the first iteration of the robust registration for the brick sequence (top), and the small font sequence (down). Left to right: The mean, the per-pixel determinant for the motion covariance matrix, and a sample frame blurred using the computed spatially varying filter. Areas with high determinant correspond to areas with high amount of blur in the mean; therefore, the frames were blurred accordingly.

rank minimization, we can decompose the matrix which has its columns as the registered frames from the previous stage $F = \{vec\{I_{r1}\}...vec\{I_{r1}\}\}$ into two components; the noise-free matrix A, and the sparse error matrix E

$$\arg\min_{A,E} rank(A) \text{ s.t. } F = A + E, ||E||_0 \leq \beta, \tag{5}$$

where β is a constant that represents the maximum number of corrupted pixels expected across all images. Using the Lagrangian form, equation 5 can be written as

$$\arg\min_{A,E} rank(A) + \lambda ||E||_0 \text{ s.t. } F = A + E, \tag{6}$$

where λ is a parameter that trades off the rank of the solution versus the sparsity of the error, and we always set it to $1/\sqrt{(w \times h)}$ following the theoretical considerations in [3]. Consequently, we apply convex relaxation to the problem by replacing $rank(A)$ with the nuclear norm or sum of the singular values $||A||_* = \Sigma_i(\sigma_i)$, and replacing $||E||_0$ with its convex surrogate ℓ_1 norm $||E||_1$

$$\arg\min_{A,E} ||A||_* + \lambda ||E||_1 \text{ s.t. } F = A + E. \tag{7}$$

Equation 7 is convex and can be solved with convex optimization methods such as the Augmented Lagrange Multiplier (ALM) algorithm [11] which we found robust and fast in our scenarios. The final output matrix A comprises in its columns the final reconstructed frames $\{I_{f1}...I_{fn}\}$.

4.1. From Stage One to Stage Two

At the first sight, our combination of stages might seem adhoc, while in fact the two stages are tightly connected and complement each other in recovering the underwater scene. As discussed earlier in this article, the water turbulence induces two noise components in the underwater sequence; a local miss-alignment, and a random noise. The random noise can only be unveiled by rank minimization if it is reasonably sparse [3]. However, in the raw video, the local-miss alignment is dominant, and conceals the random noise. For that reason, the sparse noise elimination can only be utilized after first aligning the sequence through the robust registration. On the other hand, the robust registration itself fails to overcome the sparse errors, and that is where the second stage comes in handy. Figure 4 shows the result of applying the stages separately and then combined. It is clear that each stage plays an essential role in robustly reconstructing the underwater scene.

One issue arises under this formulation: When can we consider the sequence well-aligned such that the current errors are reasonably sparse? In other words, when can we move from stage 1 to stage 2? Our experiments indicate that the answer to such a question is still open-ended, because computing an exact parameter quantifying the current sparsity of error intrinsically requires the error to be first identified, which can only be available after stage 2. Interestingly, we found in our experiments that some quantities correlated to sparsity can be used as robust indicators of sparsity. Namely, we use the ℓ_1 difference between the frames and the mean of the current sequence V to make a stopping decision for stage 1

$$\ell_1(V) = \frac{\sum_i \sum_{\mathbf{x}} |M(\mathbf{x}) - I_i(\mathbf{x})|}{w \times h \times n}. \quad (8)$$

After normalizing the images to $[0-1]$, a fixed sparsity threshold of .025 on ℓ_1 worked quite well in all sequences. (Refer to our website for additional details about this). Technically, there is absolutely no penalty for reducing the threshold other than a possible waste of processing time. Such a threshold is directly related to the parameter λ in equation 7 since both are limits for the number of expected corrupted pixels. The investigation of such a relation goes beyond the scope of this paper; thus, we leave it for future work. The complete "Seeing through water" procedure is summarized in algorithm 1.

5. Experiments

We extensively experimented on the proposed ideas using several standard underwater sequences from [18]. Each frame is 256×256 with 61 frames per sequence. The robust registration stage converges in $3-6$ iterations, while the sparse noise elimination stage quickly converges in $40-43$ ALM iterations. Overall, our algorithm takes about half of the time required by [18]. Figure 5 shows our output compared to the state of the art [18]. Our algorithm is able to reconstruct all the underwater scenes and generate superior high quality means. It can be noticed from the figure that the mean of the sequence does not considerably improve after the sparse noise elimination, this essentially indicates that the majority of the remaining errors after the registration step belong to a zero mean additive noise. The effect of the sparse noise elimination, however, can be clearly observed

Algorithm 1: Seeing through water

input : Distorted image set
$V \in \mathbb{R}^{w \times h \times n} = \{I_1, ..., I_n\}$
output: Undistorted image set
$V_f \in \mathbb{R}^{w \times h \times n} = \{I_{f1}, ..., I_{fn}\}$

$M \leftarrow TemporalMean(V)$;
Stage 1: Robust Registration begin
  while $\ell_1(V) > Sparsity\_Threshold$ **do**
    $B \leftarrow ComputeBlurKernel(\Gamma)$;
    $V_b \leftarrow$ Convolve the frames V with B;
    $\Gamma \leftarrow$ Compute the warping from each frame in V_b to the mean M;
    $V_w \leftarrow$ Warp the unblurred set V using Γ;
    $V \leftarrow V_w$;
    $M \leftarrow TemporalMean(V)$;
  end
  $V_r \leftarrow V$;
end
Stage 2: Sparse Noise Elimination begin
  $F \leftarrow vec\{V_r\} = \{vec(I_{r1}), ..., vec(I_{rn})\}$;
  $A, E \leftarrow \arg\min_{A,E} ||A||_* + \lambda ||E||_1$ s.t. $F = A + E$;
  $\{I_{f1}, ..., I_{fn}\} \leftarrow$ Reshape A to $w \times h \times n$;
end

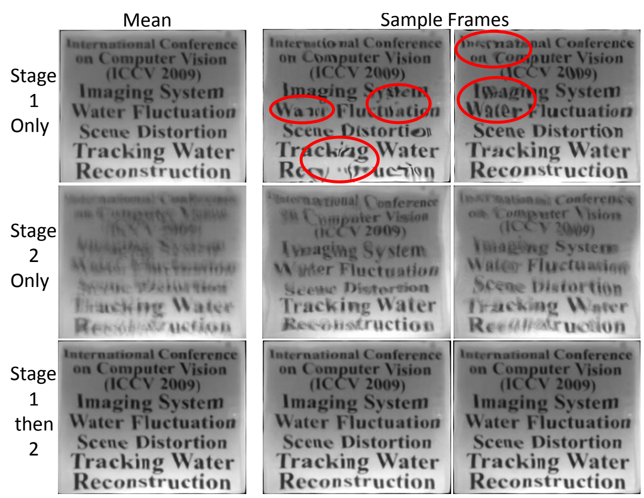

Figure 4. Mean image and sample frames after applying the robust registration without the sparse noise elimination (top), the sparse noise elimination without first applying the registration (middle), and both stages applied (bottom). The first stage itself is capable of obtaining a robust mean; however, the frames still contain several sparse errors (highlighted in red). Applying the second stage to the raw video fails for the reason discussed in the text. Combining the stages clearly achieves the best results.

on the frames as illustrated in figure 6. Additionally, the role of each stage is better observed in the video stabilizing results provided on our website.

Figure 7 shows an example for the evolution of the mean

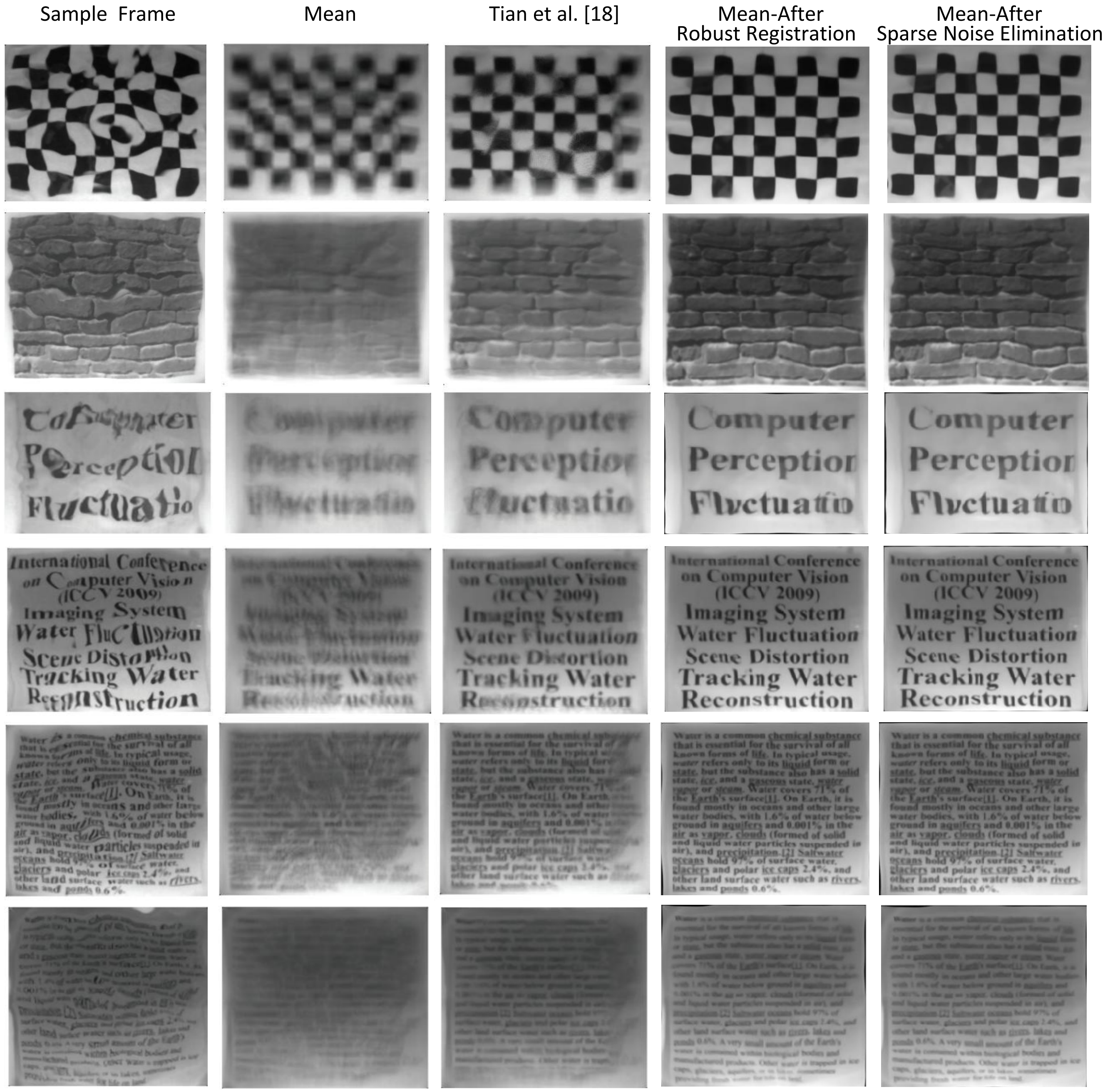

Figure 5. Image restoration results on standard sequences from [18]. The first column shows a sample frame from the input video, which is severely distorted. The second column shows the temporal mean of the sequence. The third column is the result from [18]. Finally our results are shown in the last two columns, after the first and second stages respectively. Results from our method clearly outperform [18] on all sequences even after the first stage. (Please zoom in to see the details).

Table 1. Performance of our method compared to [18].

	SSD	NMI	LNMI	SSDG
Sequence	Tian et al. / Ours	Tian et al. / Ours	Tian et al. / Ours	Tian et al. / Ours
Brick	0.019 / 0.009	1.083 / 1.101	1.116 / 1.138	0.007 / 0.004
Middle Font	0.027 / 0.011	1.072 / 1.178	1.147 / 1.223	0.012 / 0.005
Small Font	0.021 / 0.008	1.046 / 1.118	1.100 / 1.168	0.018 / 0.006
Tiny Font	0.017 / 0.011	1.088 / 1.134	1.091 / 1.133	0.006 / 0.004

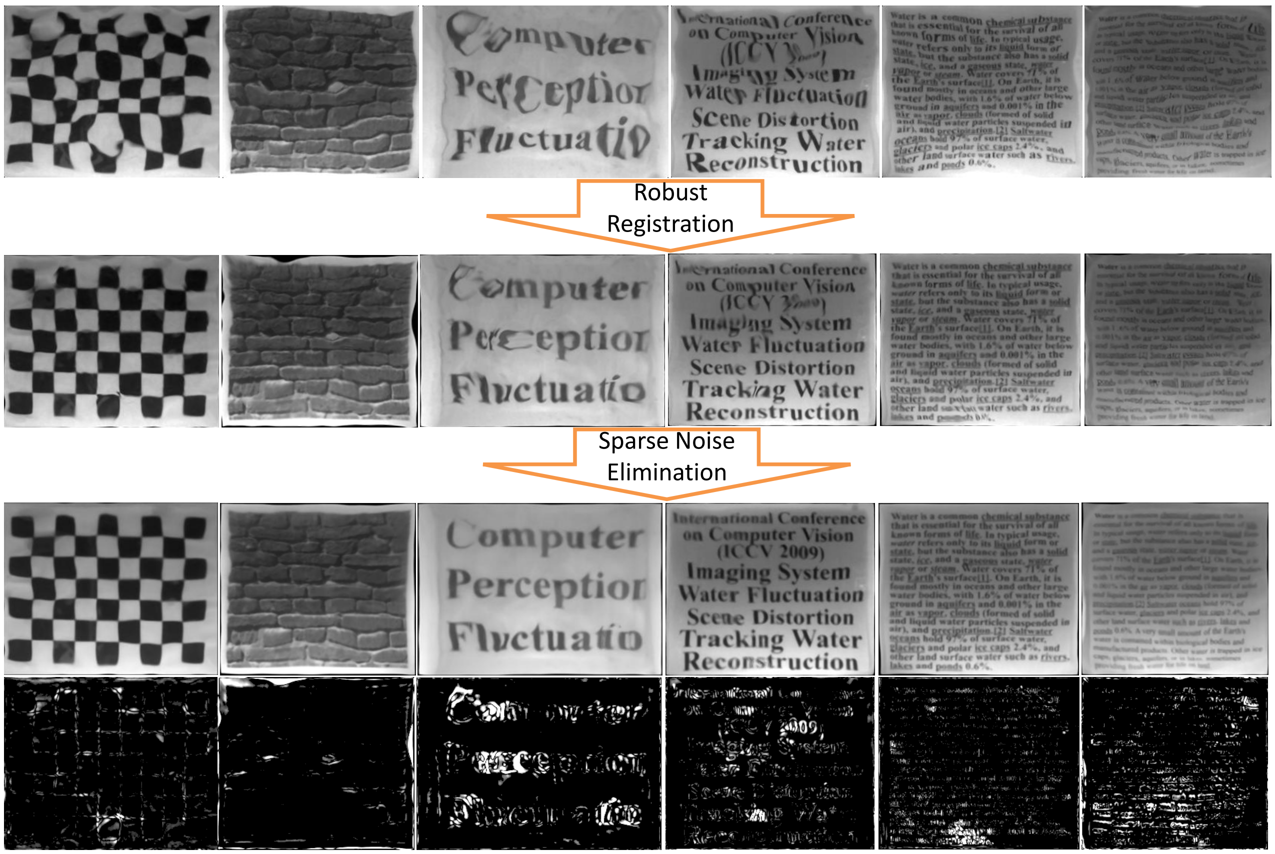

Figure 6. A sample frame from each sequence after applying each stage of our algorithm. The first stage overcomes most water turbulence; however, sparse errors are only eliminated after the second stage. The final two rows show the reconstructed images and the sparse errors respectively after rank minimization. (Please zoom in to see the details, and refer to our website for the complete videos).

for the middle font sequence from [18] during stage 1 of the algorithm under three cases: direct registration without blurring or deblurring, with mean deblurring using [21], and finally with frame blurring. It is clear from the figure that our robust registration algorithm is capable of finding a high quality mean in a few iterations in all cases. However, frame blurring evidently achieves the best results with all underwater words correctly reconstructed.

We use the reconstructed mean M and the template T provided from [18] to quantitatively compare with the results from [18] using four standard performance metrics:

- Sum of squared difference (SSD):

$$SSD(M,T) = \frac{\sum_{\mathbf{x}}(M(\mathbf{x}) - T(\mathbf{x}))^2}{w \times h}. \quad (9)$$

- Normalized Mutual Information (NMI):

$$NMI(M,T) = \frac{(H(M) + H(T))}{H(M,T)}, \quad (10)$$

  where $H(M)$, $H(T)$ are the entropies of M and T, and $H(M,T)$ is the joint entropy of M and T. H is calculated from the histograms of the gray values of the images. This measure is heavily used in image registration such as [15]. Therefore, here it is a strong indication of how well the two images are aligned.

- Local Normalized Mutual Information (LNMI): This is similar to NMI except that it is computed for every patch of a 10×10 grid, and then normalized. LNMI captures the spatial relations among the image parts.

- SSD in Gradient (SSDG):

$$SSDG(M,T) = SSD(M_x, T_x) + SSD(M_y, T_y), \quad (11)$$

where M_x, T_x, M_y, T_y are the horizontal and vertical gradients for M and T. Gradient-based features were proved to be crucial in text recognition [17]. Since many of our testing sequences contain underwater text, we use this measure to compare text recognition accuracy.

Table 1 summarizes the obtained results for each of the sequences with an available template after normalizing the images to $[0-1]$. It is clear that our method drastically outperforms [18] in all sequences in terms of all measures. It is important to note that our proposed method should not be compared to [19] since it assumes a different formulation where the template of the sequence is used.

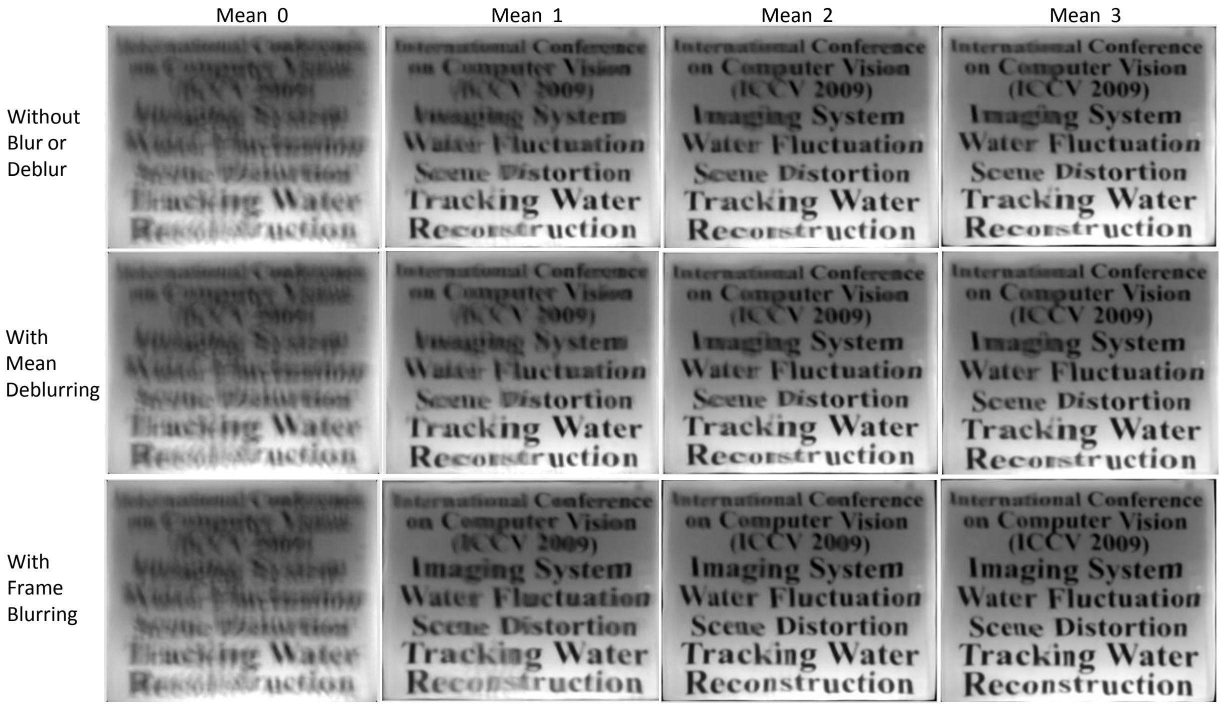

Figure 7. Evolution of the mean in stage 1. Left to right: The mean after each iteration of registration. Top to bottom: Stage 1 applied without blurring or deblurring, with mean deblurring, and with our frame blurring. After three iterations, the mean is significantly enhanced in all cases. However, underwater words on the left part of the image like "Imaging", "Water", "Scene", "Tracking", and "Reconstruction" are only correctly reconstructed using the frame blurring. (Please zoom in to see the details).

6. Conclusion

We presented a novel, simple, and easily implementable method to reconstruct a sequence distorted by water waves. An iterative registration algorithm is first employed to recover a well aligned sequence and an enhanced mean. Consequently, sparse errors are extracted through rank minimization. We showed by experiments that the proposed method robustly recovers several sequences and highly outperforms state of the art. Our method is a general dewarping technique; therefore, in the future, we will investigate its application to further types of noise and turbulence.

References

[1] A. Agrawal, Y. Xu, and R. Raskar. Invertible motion blur in video. In *SIGGRAPH*, 2009. 2

[2] S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework. In *IJCV*, 2008. 1

[3] E. J. Candes, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? In *Preprint*, 2009. 2, 4, 5

[4] A. Donate, G. Dahme, and E. Ribeiro. Classification of textures distorted by waterwaves. In *ICPR*, 2006. 1, 2

[5] A. Donate and E. Ribeiro. Improved reconstruction of images distorted by waterwaves. In *International Conference on Computer Vision Theory and Applications*, 2006. 1, 2

[6] A. Efros, V. Isler, J. Shi, and M. Visontai. Seeing through water. In *NIPS*, 2004. 1, 2, 4

[7] H. Ji, C. Liu, and Y. X. Zuowei Shen. Robust video denoising using low rank matrix completion. In *CVPR*, 2010. 2

[8] N. Joshiy, C. L. Zitnicky, R. Szeliskiy, and D. J. Kriegman. Image deblurring and denoising using color priors. In *CVPR*, 2009. 2

[9] I. M. Levin, V. V. Savchenko, and V. J. Osadchy. Correction of an image distorted by a wavy water surface: laboratory experiment. In *Applied Optics*, 2008. 2

[10] Y. Li, S. Kang, N. Joshi, S. Seitz, and D. Huttenlocher. Generating sharp panoramas from motion-blurred videos. In *CVPR*, 2010. 2

[11] Z. Lin, M. Chen, L. Wu, and Y. Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. In *UIUC Technical Report*, 2009. 2, 4

[12] I. Matthews and S. Baker. Active appearance models revisited. In *IJCV*, 2004. 1

[13] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma. Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. In *PAMI*, 2010. 2

[14] J. Pilet, V. Lepetit, and P. Fua. Fast non-rigid surface detection, registration and realistic augmentation. In *IJCV*, 2008. 1

[15] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. In *Medical Imaging*, 1999. 3, 7

[16] M. Shimizu, S. Yoshimura, M. Tanaka, and M. Okutomi. Super-resolution from image sequence under influence of hot-air optical turbulence. In *CVPR*, 2008. 1

[17] P. Shivakumara, T. Q. Phan, and C. L. Tan. A gradient difference based technique for video text detection. In *ICDAR*, 2009. 7

[18] Y. Tian and S. Narasimhan. Seeing through water: Image restoration using model-based tracking. In *ICCV*, 2009. 1, 2, 4, 5, 6, 7

[19] Y. Tian and S. G. Narasimhan. A globally optimal data-driven approach for image distortion estimation. In *CVPR*, 2010. 1, 2, 7

[20] Z. Wen, D. Fraser, A. Lambert, and H. Li. Reconstruction of underwater image by bispectrum. In *ICIP*, 2007. 1, 2

[21] L. Xu and J. Jia. Two-phase kernel estimation for robust motion deblurring. In *ECCV*, 2010. 2, 7

[22] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. In *ACM Computing Surveys (CSUR)*, 2006. 1