Foreground Extraction of Underwater Videos via Sparse and Low-rank Matrix Decomposition*

Hongwei Qin[1,2], Yigang Peng[3], and Xiu Li[1,2]

1. Department of Automation, Tsinghua University, Beijing 100084
2. Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055
3. National Computer Network Emergency Response Technical Team Coordination Center of China (CNCERT or CNCERT/CC), Beijing 100029
qhw12@mails.tsinghua.edu.cn, pengyigang@cert.org.cn, li.xiu@sz.tsinghua.edu.cn

Abstract. In this paper, we propose a new method for foreground extraction of underwater videos based on sparse and low-rank matrix decomposition. By stacking the underwater video frames as columns of a matrix, principal component pursuit algorithm is used for decomposing the matrix into a low-rank matrix representing the stationary background and a sparse matrix representing the activities in the foreground. Then, the sparse matrix is processed with adaptive threshold to extract objects in the foreground. We evaluate our method quantitatively on various underwater videos. Our method is robust to various scenarios like blurred videos, illumination variations in the background, and crowded foreground objects. The experimental results demonstrate the promising performance of our proposed method.

Keywords: foreground extraction, underwater videos, principle component pursuit, adaptive threshold.

1 Introduction

The advancements in underwater imaging system, such as the NEPTUNE and VENUS observatories[1], have resulted in a proliferation of underwater video data and static images. On the one hand, it provides ocean scientists and biologists a new powerful way to monitor the complex underwater environments and marine species. On the other hand, it imposes a series of great challenges for underwater imagery and video analysis.

In terms of monitoring marine species, we often need to identify activities that stand out from the background from a sequence of monitoring underwater video frames. However, some special properties of underwater videos impose

* This work is supported by National Natural Science Foundation of China (Grant No. 71171121/61033005) and National 863 High Technology Research and Development Program of China (Grant No. 2012AA09A408).

[1] http://www.oceannetworks.ca/

great challenges for background modeling and foreground extraction of underwater videos. For instance, the smoothed and low contrasted images often affect the performance of algorithms that use texture information. The background of some videos may be featured by complex textures. Sometimes, the video may contain transient and abrupt luminosity changes. Absorption and scattering may cause light attenuation, limiting the visibility. Furthermore, underwater objects like fish behave erratically and fast in 3-degree. All of these properties make foreground extraction of underwater video a very challenging research problem. Methods performing well in normal scenarios are often difficult to generalize to underwater ones. In [1], they use adaptive Gaussian Mixture Model (GMM) [2] to detect fish. In [3], they use Video Background Extraction (ViBe) [4] algorithm which utilizes a list of most recent values of each pixel to decide the foreground. In [5], the authors present a texton based kernel density estimation method to build background and foreground models. A comparative study on some object detection algorithms in the task of fish detection is implemented in [6].

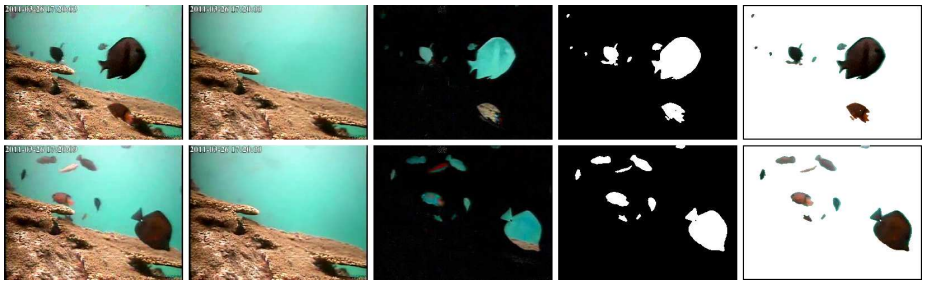


Fig. 1. One example of our method on foreground extraction of underwater videos. The first column shows two original frames of an underwater video. The second and third columns show the background part and foreground part obtained by the PCP procedure, respectively. The corresponding foreground masks calculated via the adaptive threshold procedure are shown in the forth column, and the fifth column shows the extracted foreground.

In this paper, we propose a new foreground extraction method for underwater videos based on sparse and low-rank matrix decomposition. If we stack the underwater video frames as columns of a matrix, the stationary background could be naturally modeled as the low-rank matrix component, and the moving objects in the foreground could be modeled as the sparse matrix component. In this way, the foreground and background separation problem is modeled as a sparse and low-rank matrix decomposition problem, which can be solved via a very convenient convex program called Principal Component Pursuit (PCP) [7]. To extract the objects in the foreground, we further process the sparse component with adaptive threshold on each frame, then we discard the small connected area so as to remove unwanted noise. In this way, the foreground mask and corresponding extracted foreground are obtained. We show the results obtained by our method in Fig. 1 as an example. The performance of our proposed method is evaluated on various underwater videos.

As a reminder, the paper is organized as follows. In Section 2, we introduce our proposed foreground extraction of underwater videos method based on sparse and low-rank matrix decomposition in details. We demonstrate our method on a variety of different underwater videos in Section 3. Finally, we conclude our work in Section 4.

2 Foreground extraction based on sparse and low-rank matrix decomposition

2.1 Foreground and background separation

Given a sequence of monitoring underwater video, we stack the video frames as columns of a matrix $\mathbf{M}$. As a result, for an n-frame video, each frame of which is of m-pixel, $\mathbf{M}$ is an $m \times n$ matrix.

Modeling stationary background. For monitoring the underwater environments, the camera is placed in a fixed place. The background of captured video is usually stationary except for smooth movements caused by water flowing and/or light changes. The model of background needs to be flexible enough to accommodate these smooth changes in the scene. In this sense, the background part of the monitoring underwater videos could be modeled as an approximate low-rank matrix $\mathbf{L}$ due to the correlation of frames.

Modeling foreground. The low-rank model will be violated due to the presence of moving objects in the foreground. In the monitoring underwater scenes, every frame may contain some moving marine species like fish in the foreground. Swimming fish generally occupies only a fraction of pixels of each video frame. In this sense, foreground objects could be presented as sparse errors. In other words, we could use a sparse matrix $\mathbf{S}$ to model the foreground of the monitoring underwater video.

Foreground and background separation via principal component pursuit. Putting all of these together, we model the background and foreground separation problem as a problem of decomposing the matrix of underwater video into a low-rank matrix of background component and a sparse matrix of foreground component. Mathematically, it is described as the following convex optimization problem:

$$\min_{\mathbf{L},\mathbf{S}} \quad \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1, \quad \text{s.t.} \quad \mathbf{M} = \mathbf{L} + \mathbf{S}, \tag{1}$$

which is dubbed as Principal Component Pursuit (PCP) [7], where $\|\mathbf{L}\|_*$ is the nuclear norm of matrix $\mathbf{L}$ (i.e. the sum of all its singular values), $\|\mathbf{S}\|_1$ is the ℓ_1-norm of matrix $\mathbf{S}$ (i.e. the sum of absolute value of all its entries), $\mathbf{M}, \mathbf{L}, \mathbf{S} \in \mathbb{R}^{m\times n}$, and $\lambda > 0$ is a weighting parameter. The sparse and low-rank matrix decomposition model and its variants have been applied in several image processing and computer vision problems successfully, such as robust batch image alignment [8], transform invariant low-rank texture [9], robust video restoration [10], and so on. Many practical algorithms are well developed to solve it efficiently [11,12,13].

Using principal component pursuit, the background and foreground of underwater videos could be well separated. The low-rank matrix captures the stationary background. It is also tolerated to smooth changes in the background caused by water flowing and global illumination changes. The sparse matrix represents the difference between the recorded underwater frame and the stationary background, which captures the objects like swimming fish in the foreground.

2.2 Foreground extraction via adaptive threshold method

Once the foreground and background is well separated using PCP, we deal with the sparse matrix obtained by PCP to further extract the objects in the foreground. In order to avoid the effect of small noises, we propose an adaptive threshold scheme. Specifically, for each obtained foreground frame $\mathbf{s}_i$ ($\mathbf{s}_i$ is the sparse component vector of the ith frame $\mathbf{m}_i$), we set the threshold as

$$T_i = \max\left(\alpha \text{median}(\text{abs}(\mathbf{s}_i)), \epsilon\right), \tag{2}$$

where median$(\cdot)$ is the median value of a vector, abs$(\cdot)$ is the absolute value, and α, ϵ are constant parameters. For each frame, the pixels of the sparse component $\mathbf{s}_i$ whose values are higher than T_i are marked as foreground. The threshold value T_i is related to the median value of the sparse component of each frame, which makes it robust and adaptive, and T_i is also restricted by a constant parameter ϵ so that it is not so small. Furthermore, in order to discard the noises, we follow [5] to remove the connected components containing area fewer than 15 pixels.

3 Experiments and Analysis

Fig. 2. Selected frames of the underwater video dataset. From top left to bottom right: (1) Blurred (smoothed and low contrasted), (2) Complex Background, (3) Crowded, (4) Dynamic Background (background moving: e.g. water waving), (5) Hybrid (various features), (6) Camouflage Foreground Object (background and objects having similar colors), (7-8) Luminosity Variations.

We evaluate the effectiveness of our proposed method on Underwater Benchmark Dataset For Target Detection Against Complex Background[2] [14]. This

[2] All of the videos we used in our experiments are downloaded from http://f4k.dieei.unict.it/datasets/bkg_modeling/.

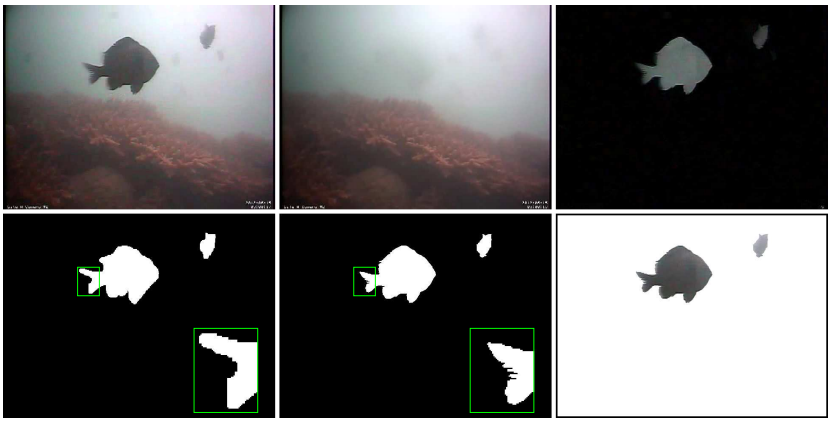
underwater benchmark dataset consists of seven categories of real-life underwater videos (spatial resolution ranging from 320×240 to 640×480, and frame rate ranging from $5 frames/s$ to $24 frames/s$), representing complex challenges in underwater video background modeling. The seven categories are Blurred, Complex Background, Crowded, Dynamic Background, Hybrid, Camouflage Foreground Object, and Luminosity Variations, respectively. Examples of these videos are shown in Fig. 2. In the experiments, to solve the principal component pursuit (PCP) problem, we use the inexact augmented Lagrangian multiplier (IALM) algorithm[3], and the weighting parameter λ is fixed to be $1/\sqrt{\max(m,n)}$. Finally, in all of the examples, we choose $p = 15$, i.e. we discard connected area which is smaller than 15-pixels.

Subjective evaluation. We apply our method on different categories of underwater videos, and some results are shown in Fig. 3. As can be seen from Fig. 3, the moving objects in the foreground are well extracted in all of the Blurred, Crowded, and Luminosity Variations cases. First, the PCP procedure makes background and foreground separated to a satisfied extend. Second, the adaptive threshold procedure can well identify the real moving objects from the sparse component obtained by PCP. We compare our method with the ground truth masks. Notice that the ground truth masks are labeled on a collaborative web-based platform [14], and they may be of low quality. In the Blurred cases with smoothed and low contrasted images as shown in Fig. 3(a), compared with the ground truth mask, our obtained mask contains even sharper edge such as in the fish tail. In a recent report from Fish4Knowledge project [4], blurred and highly blurred videos occupied 47.4%, while the normal ones only 12.9% among about 350,000 videos. In the Crowed cases as shown in Fig. 3(b), the extracted foreground by our method contains more fish including those hidden behind rocks. Our method is also robust to illumination variations as shown in Fig. 3(c).

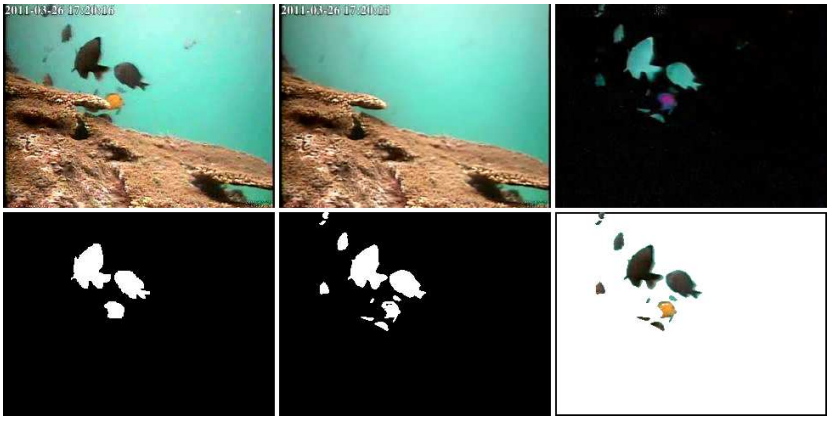
PR curves with respect to varying thresholds. We follow Achanta *et al.*'s two methodologies [15] to evaluate the accuracy of our extracted foreground. In the first evaluation, we calculate the average Precision-Recall (PR) curves with respect to varying thresholds for different categories of videos. Once the sparse component matrix is obtained by PCP, we vary the threshold T_i from 0 to 128 with a step of 1, and compare the obtained foreground masks with the manually labeled ground truth masks to generate precision-recall pairs. The PR curves of seven categories of videos are shown in Fig. 4. In particular, our method achieves better performance on Blurred videos, Crowded videos and Luminosity Variations videos than that on Complex Background videos, Camouflage Foreground videos and Hybrid videos. The performance on Dynamic Background videos is not so satisfying. This is reasonable because when the background changes fast, the PCP may not well separate the moving objects in the foreground and the

[3] The code for solving PCP problem using IALM is downloaded from http://perception.csl.illinois.edu/matrix-rank/home.html.

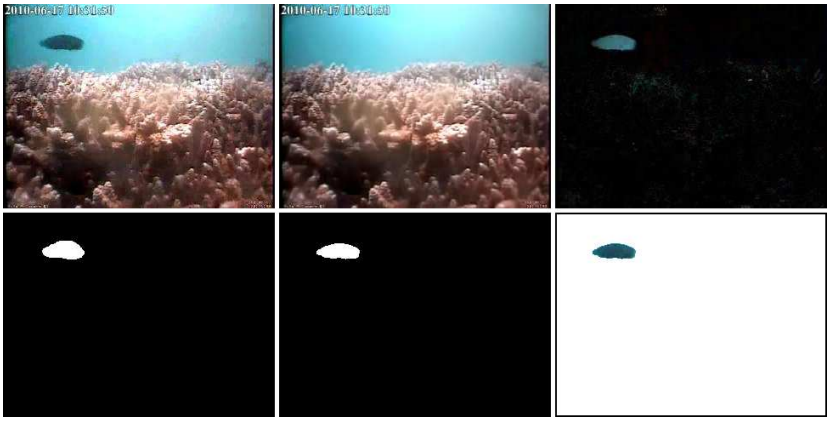
[4] http://groups.inf.ed.ac.uk/f4k/FINALPRES/F4KY3WP4Final.pdf



(a) Blurred videos ($\alpha = 12, \epsilon = 10$)



(b) Crowded videos ($\alpha = 11, \epsilon = 25$)



(c) Luminosity Variations videos ($\alpha = 4, \epsilon = 9$)

Fig. 3. Subjective evaluation of our proposed method on different videos. For each example (a,b,c) from top left to bottom right: (1) one frame of the original video, (2) the low-rank part and (3) the sparse part obtained by the PCP procedure, (4) the ground truth foreground mask, (5) the foreground mask and (6) the extracted foreground obtained by our method.

changes in the background, and thus the following adaptive threshold procedure cannot distinguish the moving objects from the obtained sparse component.

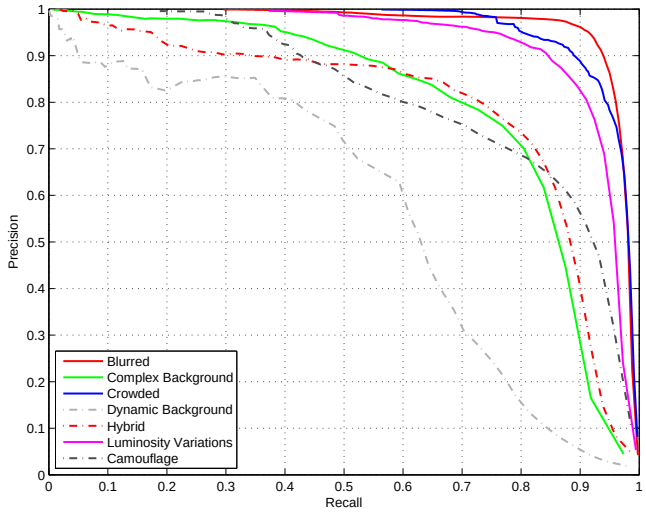


Fig. 4. Precision-recall curves of our method with the threshold ranging from 0 to 128 on the Underwater Benchmark Dataset.

F-measure evaluation. In the second evaluation, we use F-measure to evaluate the performance of our approach, where F-measure is defined as $F = \frac{2PR}{P+R}$, and P, R are precision value and recall value, respectively. We calculate an average F-measure of the video frames for each category, and the results are listed in Table 1.

Table 1. The F-measure evaluation on different video categories

	Blurred	Complex	Crowded	Dynamic	Hybrid	Luminosity	Camouflage
F-measure(%)	93.75	74.65	89.36	55.88	78.1	86.27	73.35

4 Conclusions and future work

In this paper, we propose an effective foreground extraction method for underwater videos based on sparse and low-rank matrix decomposition. By modeling the background as low-rank component and the foreground as sparse component, the principal component pursuit algorithm is adopted for background and foreground separation. Then, we design an adaptive threshold scheme to deal with the obtained sparse matrix so as to extract foreground. Experiments on various underwater videos verify the effectiveness of our method. It could be further used for underwater video analysis such as fish counting, tracking, recognition, etc. As foreground extraction is among the most sophisticated image processing tasks, solutions that work well on as many categories as possible are expected. We believe sparse and low-rank matrix decomposition could be an option to achieve that goal.

References

1. Nadarajan, G., Chen-Burger, Y.H., Fisher, R.B., Spampinato, C.: A flexible system for automated composition of intelligent video analysis. In: Image and Signal Processing and Analysis (ISPA), 2011 7th International Symposium on, IEEE (2011) 259–264
2. Stauffer, C., Grimson, W.E.L.: Adaptive background mixture models for real-time tracking. In: Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on. Volume 2., IEEE (1999)
3. Beauxis-Aussalet, E., Palazzo, S., Nadarajan, G., Arslanova, E., Spampinato, C., Hardman, L.: A video processing and data retrieval framework for fish population monitoring. In: Proceedings of the 2nd ACM international workshop on Multimedia analysis for ecological data, ACM (2013) 15–20
4. Barnich, O., Van Droogenbroeck, M.: Vibe: A universal background subtraction algorithm for video sequences. Image Processing, IEEE Transactions on **20**(6) (2011) 1709–1724
5. Spampinato, C., Palazzo, S., Kavasidis, I.: A texton-based kernel density estimation approach for background modeling under extreme conditions. Computer Vision and Image Understanding **122** (2014) 74–83
6. Kavasidis, I., Palazzo, S.: Quantitative performance analysis of object detection algorithms on underwater video footage. In: Proceedings of the 1st ACM international workshop on Multimedia analysis for ecological data, ACM (2012) 57–60
7. Candès, E., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? Journal of the ACM **58**(3) (2011) 1–37
8. Peng, Y., Ganesh, A., Wright, J., Xu, W., Ma., Y.: RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) **34**(11) (2011) 2233–2246
9. Zhang, Z., Ganesh, A., Liang, X., Ma, Y.: TILT: Transform-invariant low-rank textures. International Journal of Computer Vision **99**(1) (August 2012) 1–24
10. Ji, H., Huang, S., Shen, Z., Xu, Y.: Robust video restoration by joint sparse and low rank matrix approximation. SIAM Journal on Imaging Sciences **4**(4) (2011) 1122–1142
11. Toh, K.C., Yun, S.: An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems. Pacific Journal of Optimization **6** (2010) 615–640
12. Lin, Z., Chen, M., Ma, Y.: The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. arXiv preprint arXiv:1009.5055 (2010)
13. Yuan, X., Yang, J.: Sparse and low-rank matrix decomposition via alternating direction methods. Pacific Journal of Optimization **9**(1) (2013) 167–180
14. Kavasidis, I., Palazzo, S., Di Salvo, R., Giordano, D., Spampinato, C.: An innovative web-based collaborative platform for video annotation. Multimedia Tools and Applications (2013) 1–20
15. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE (2009) 1597–1604