Underwater Image Sparse Representation based on Bag-of-Words and Compressed Sensing

Congcong Shi, Rui Nian*, Bo He, Yue Shen
School of Information Science and Engineering
No. 238 Songling Road
Ocean University of China, Qingdao
E-Mails: hd_shicc@163.com; nianrui_80@163.com
Tianhong Yan
Department of Mechanical & Electrical Engineering
China Jiliang University
Hangzhou, China,
E-Mails: thyan@163.com

Amaury Lendasse
Arcada University of Applied Sciences, 00550
Helsinki,Finland; Department of Mechanical and Industrial
Engineering and the Iowa Informatics Initiative, The
University of Iowa, Iowa City, IA 52242-1527, USA;
E-Mails: amaury lendasse@uiowa.edu;
lendasse@gmail.com

***Abstract*—In this paper, a novel approach based on bag-of-words and compressed sensing has been proposed for the sparse representation of the underwater images. Properties in the water, such as the limited range, non-uniform lighting, low contrast, diminished colors, has been taken into consideration for underwater image sparse representation. Our method is a simple and computationally efficient extension of an orderless bag-of-words model, and it shows consistently and significantly improved performance on underwater image representation tasks.**

Keywords—bag-of-words; compressed sensing; sparse representation;

I. INTRODUCTION

In the Era of Big Data, how to deal with the high-dimensional data turns out to be one central topic. In the ocean investigations, especially those tasks deployed by the remotely operated vehicles (ROVs) and the autonomous underwater vehicles (AUVs), there is a high demand on the real-time sparse representation of the underwater images in the high-dimensional space.

Sparse representation is a extension of traditional signal representation method (e.g. Fourier transform). Research on image sparse representation suggests that image patches can be well represented as a sparse linear combination of elements from an appropriately chosen over-complete dictionary. This image sparse representation model has been successfully applied to image processsing applications. such as denoising [13] and restoration [12] and so on. The model can be ralated to some existing model such as principal component analysis (PCA)[6], independent components analysis (ICA)[7].

In recent years, the Bag-of-Words (BoW) model [4], an image sparse representation through the occurrence counts of each visual word in the dictionary bag, has been widely used in some domain, such as sport image analysis [9], facial expression recognition [10], medical images [11], image quality assessment [5], and so forth.

However, in the underwater imaging system, poor visibility is still a major limitation. Properties in the water, such as the limited range, non-uniform lighting, low contrast, diminished colors, blur imaging and so on may bring the mismatching image feature extraction and limit the accuracy of the visual word acquisition with high quality in both efficiency and robustness. Recently, compressed sensing, has been brought as a sampling/reconstruction concept by Donoho [1] and Candes [2, 3], which could exploit the sparsity, the compressibility, the incoherence of the spatially sparse signals, and tolerate condensing at a sampling rate significantly below the Shannon-Nyquist theorem. Therefore, in this paper, a novel approach based on bag-of-words and compressed sensing has been proposed for the sparse representation of the underwater images.

In this paper, we further develop the underwater image sparse representation model and enhance the quality of the image understanding in the underwater system. The paper is organized as follows: section II describes the basic framework of underwater image representation in our system. Section III introduces the bag-of-words model. Section IV is about the compressed sensing and Section V refers to the experimental results and analysis. Section VI comes to the conclusions.

II. GENERAL FRAMEWORK

In this paper, we present a novel scheme for the underwater image sparse representation via bag-of-words and compressed sensing. A brief flow chart of our proposed sparse representation framework is shown in Figure 1, which is made up of several steps, including the underwater images acquirement, some preprocessing, the feature detection, the feature description, the dictionary generation, the sparse transformation, the measurement matrix, the histogram representation, etc. Figure 2 shows the whole sparse representation process of our proposed approach for the underwater images.

III. BAG-OF-WORDS

In this paper, the Bag-of-Words (BoW) can represent an underwater image through the occurrence counts of each visual word in the dictionary bag. Let one underwater image be $I(x, y)$ with the pixel coordinates (x, y). The size of $I(x, y)$ is $M \times N$. Suppose that the training set consists of n underwater images $X = [I_1(x, y), I_2(x, y), \ldots, I_n(x, y)]$.

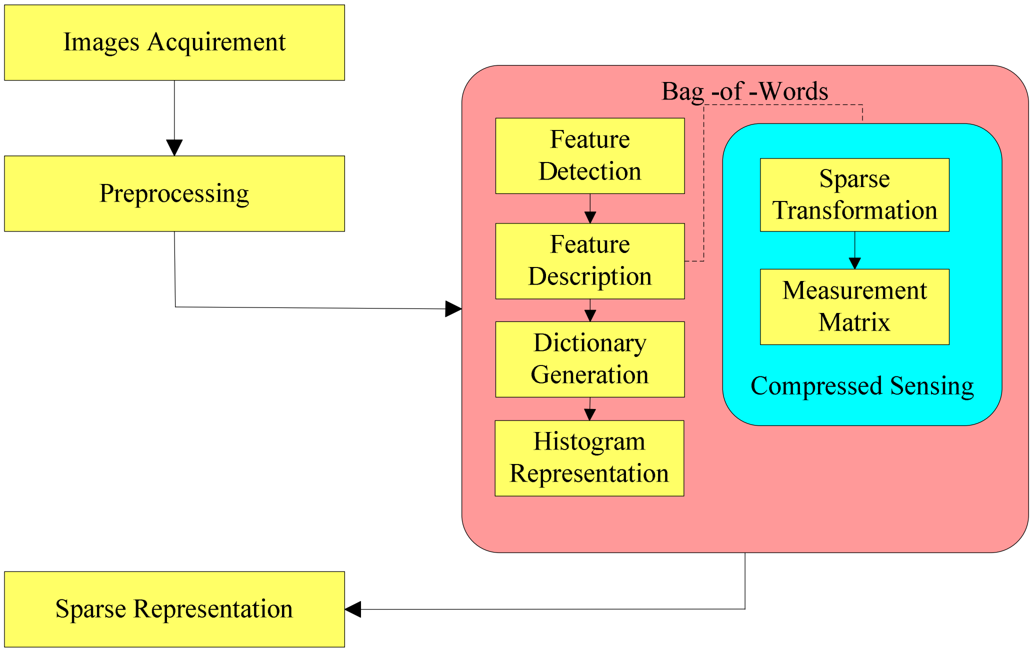


Fig. 1. The flow chart of underwater image sparse representation

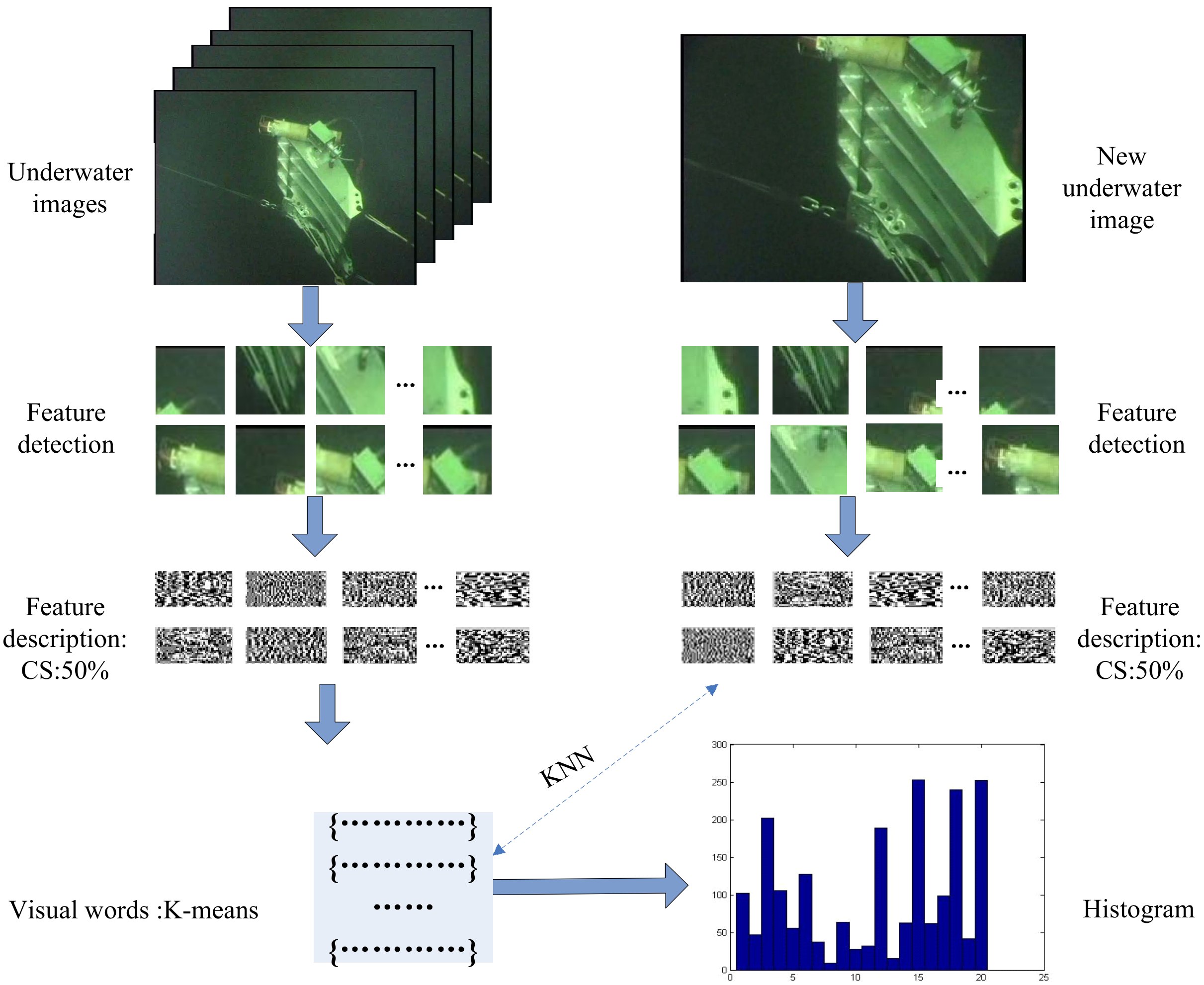


Fig.2. Sparse representation process for the underwater images

A. *Feature detection*

To detect the underwater image feature patches, we divide each underwater image $I(x,y)$ into m overlapping underwater image feature patches $r(x,y)=\{r_{i,j}(x,y)\}$ with the size $d\times d$, where $i=1:\mathrm{length}(1:d-l:M+1)-2$, $j=1:\mathrm{length}(1:d-l:N+1)-2$, l is the number of pixels in each overlapping underwater image feature patch.

B. *Feature description*

To describe each underwater image feature patch $r_{i,j}(x,y)$, the underwater image feature descriptor $y_{i,j}(x,y)$ can be acquired by compressed sensing.

C. *Dictionary generation*

Given the collection of underwater image feature descriptor $y_{i,j}(x,y)$ from the training set, we convert the underwater image feature descriptor into the visual words $m_\lambda, \lambda=1,2,\ldots k$ and compose the dictionary bag by performing k-means algorithm. Suppose that $Y=\left\{y_\eta\right\}_{\eta=1}^{m\times n}$ contains all underwater image feature descriptors $y_{i,j}(x,y)$ from training set X, we want to partition the dataset $Y=\left\{y_\eta\right\}_{\eta=1}^{m\times n}$ into disjoint clusters $\left\{w_\lambda\right\}_{\lambda=1}^{k}$, k-Means [8] minimizes the sum of the intra cluster variances ε_{sum}:

$$\varepsilon_{sum}=\sum_{\lambda=1}^{k} v_\lambda=\sum_{\lambda=1}^{k}\sum_{\eta=1}^{m\times n}\delta_{\eta\lambda}\parallel y_\eta-m_\lambda\parallel^2 \tag{1}$$

Where $v_\lambda=\sum_{\eta=1}^{m\times n}\delta_{\eta\lambda}\parallel y_\eta-m_\lambda\parallel^2$ is the variance, $m_\lambda=\sum_{\eta=1}^{m\times n}\delta_{\eta\lambda}y_\eta/\sum_{\eta=1}^{m\times n}\delta_{\eta\lambda}$ are the centers of the k-th clusters, $\delta_{\eta\lambda}$ is a cluster indicator variable, $\delta_{\eta\lambda}=\begin{cases}1 & y_\eta\in w_\lambda\\ 0 & y_\eta\notin w_\lambda\end{cases}$.

D. *Histogram representation*

For an unknown underwater image $I_{n+1}(x,y)$, here we start with the calculation of underwater image feature descriptors $Y_{n+1}=\left\{y_\eta\right\}_{\eta=1}^{m\times n}$, and then use K Nearest Neighbors algorithm (KNN) to represent the underwater image by counting the visual words $m_\lambda, \lambda=1,2,\ldots k$ from the dictionary bag.

IV. COMPRESSED SENSING

Here we make use of the concept of compressed sensing which suggests that the underwater image feature patch $r_{i,j}(x,y)$ can be represented by the measurement matrix into a number of the non-adaptive linear projection combinations in a low-dimensional space. So here, we try to acquire underwater image feature descriptor $y_{i,j}(x,y)$ to represent the underwater image feature patch $r_{i,j}(x,y)$.

Let $\left\{r_\eta\right\}_{\eta=1}^{m\times n}$ be the collection of $m\times n$ underwater image feature patches from training set X, we plan to acquire underwater image feature descriptors $Y=\left\{y_\eta\right\}_{\eta=1}^{m\times n}$ to represent the underwater image feature patches.

We suppose that every underwater image feature patch $r_\eta(x,y)$ has a sparse representation in some orthonormal basis, such as Fourier, wavelet, sinusoids, B-splines, and so on. In practical instances, the sparse matrix $x_\eta(x,y)$ may be the coefficients of the underwater image feature patch $r_\eta(x,y)$ in an orthonormal basis :

$$x_\eta(x,y)=\Psi r_\eta(x,y) \tag{2}$$

where Ψ is the d by d matrix with the waveforms ψ_i as rows or equivalently, $r_\eta(x,y)=\Psi^* x_\eta(x,y)$.

Further, let $\Phi_\eta(x,y)$ (Φ is the $q\times d$ matrix, $q<d$) be denoted as measurement matrix, then the above sparse matrix $x_\eta(x,y)$ can be transformed into underwater image feature descriptors :

$$\begin{aligned}y_\eta(x,y)&=\Phi_\eta(x,y)r_\eta(x,y)=\Phi_\eta(x,y)\Psi^* x_\eta(x,y)\\&=Ax_\eta(x,y)\end{aligned} \tag{3}$$

The sensing matrix A should obey a "restricted isometry property (RIP)". For each integer $s=1,2,\ldots$, define the isometry constant δ_s of the sensing matrix A as the smallest number such that

$$(1-\delta_s)\parallel x\parallel_{l2}^2\le\parallel Ax\parallel_{l2}^2\le(1+\delta_s)\parallel x\parallel_{l2}^2 \tag{4}$$

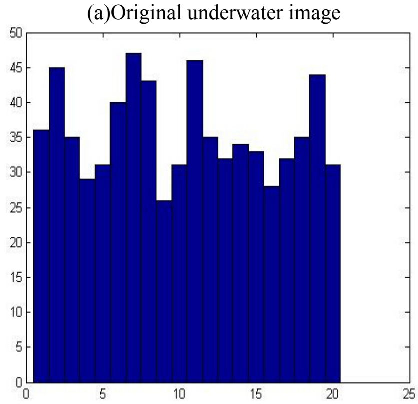
Holds for all sparse matrix $x_\eta(x,y)$. Where $\parallel x\parallel_{l2}^2=(|x_1|^{l2}+|x_2|^{l2}+\cdots+|x_d|^{l2})^{2/l2}$. We will loosely say that the sensing matrix A obeys the RIP if δ_s is not too close to one. Then we will get underwater image feature descriptors $Y=\left\{y_\eta\right\}_{\eta=1}^{m\times n}$ to describe underwater image feature patches.

V. SIMULATION EXPERIMENT

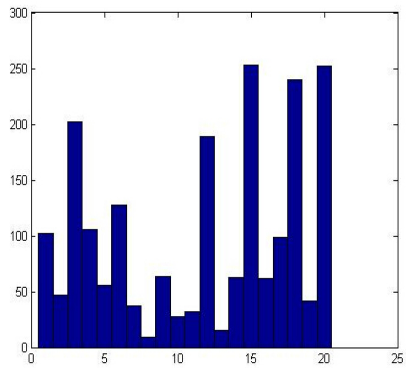
In the experiment simulations, we set up a Myring streamline AUV system with CCD camera with the horizontal resolution 480 TVL/PH and the minimum scene illumination 0.28 lux on board for the underwater visual tasks, as is shown in Figure 3. Figure 4 is one BoW histogram representation for an example underwater image, with both the classical SIFT algorithm as well as the compressed sensing in our model. It has been shown in the simulation experiments that our approach consistently and significantly achieves performance improvements over the orderless underwater image representation from the basic BoW.

Fig.3. The CCD camera on board for our underwater visual tasks

(a)Original underwater image



(b) BOW with SIFT



(c) BOW with compressed sensing

Fig.4. One example BoW histogram representation with both SIFT and compressed sensing

VI. CONCLUSION

In this paper, we have put forward a novel approach of the underwater image sparse representation based on bag-of-words and compressed sensing. The Bag-of-Words model has been first introduced to represent an underwater image through the occurrence counts of each visual word in the dictionary bag. Compressed sensing is then established with sparse matrix and measurement matrix, to improve the entire BoW model in underwater image sparse representation. The simulation results have shown the effectiveness and feasibility of the proposed approach, which is comparable to the direct BoW in underwater image representation.

ACKNOWLEDGMENT

This work is partially supported by the Natural Science Foundation of P. R. China (31202036), the National Science & Technology Pillar Program (2012BAD28B05), the National High-Tech R&D 863 Program (2014AA093410), the Key Science & Technology Project of Shandong Province (2013GHY11507), the Fundamental Research Funds for the Central Universities (201362030), and the Natural Science Foundation of P. R. China (41176076, 51379198 and 51075377).

REFERENCES

[1] D. L. Donoho, "Compressed Sensing," Technical Report, Stanford University, 2004.

[2] E. J. Candè, M. B. Wakin "An introduction to compressive sampling," IEEE Signal Processing Magazine,2008.

[3] E. J. Candès, "Compressive sampling," Proceedings of the International Congress of Mathematicians, Madrid, Spain, 2006.

[4] F. F. Li, P. Perona, "A bayesian hierarchical model for learning natural scene categories," IEEE Computer Science Society International Conference of Computer Vision and Pattern Recognition,2005.

[5] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebooks," IEEE Transactions on Image Processing, vol. 21, no. 7, pp. 3129–3138, 2012.

[6] Turk, Pentland, "Eigenfaces for ecognition," Journal of Cognitive Neuroscience, 1991, 3(1):1-86.

[7] P. Comon, "Independent component analysis- a new concept? " Signal Processing, 1994, 36:287-314.

[8] I. S. Dhillon, Y. Guan, B. Kulis, "Kernel k-means, spectral clustering and normalized cuts," ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004, pp. 551–556.

[9] P. Tirilly, V. Claveau, and P. Gros, "Language modeling for bag-of-visual words image categorization," in Proceedings of the International Conference on Image and Video Retrieval (CIVR '08), pp. 249–258, July 2008.

[10] B. Fasel, F. Monay, and D. Gatica-Perez, "Latent semantic analysis of facial action codes for automatic facial expression recognition," in Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR'04), pp. 181–188, October 2004.

[11] J. Wang, Y. Li, Y. Zhang et al., "Bag-of-features based medical image retrieval via multiple assignment and visual words weighting," IEEE Transactions on Medial Imaging, vol. 30, no. 11, pp. 1996–2011, 2011.

[12] J.Mairal, G.Sapiro and M.Elad, "Learning multiscale sparse representations for image and video restoration", Multiscale Modeling and Simulation, vol.7, pp. 214-241,2008.

[13] M. Elad, M.Aharon, " Image denoising via sparse and redundant representations over learned dictionsries," IEEE Transaction on Image Processing,vol. 2, pp. 3736-3745,2006.