Side Scan Sonar Image Super Resolution via Region-Selective Sparse Coding

Jaihyun PARK†, Bonhwa KU†, Youngsaeng JIN†, Nonmembers, and Hanseok KO†(a), Member

SUMMARY Side scan sonar using low frequency can quickly search a wide range, but the images acquired are of low quality. The image super resolution (SR) method can mitigate this problem. The SR method typically uses sparse coding, but accurately estimating sparse coefficients incurs substantial computational costs. To reduce processing time, we propose a region-selective sparse coding based SR system that emphasizes object regions. In particular, the region that contains interesting objects is detected for side scan sonar based underwater images so that the subsequent sparse coding based SR process can be selectively applied. Effectiveness of the proposed method is verified by the reduced processing time required for image reconstruction yet preserving the same level of visual quality as conventional methods.

key words: side scan sonar, super resolution, sparse coding, object detection

1. Introduction

Side scan sonar (SSS) efficiently acquires useful terrestrial images in underwater environment. The main issue for SSS usage is the tradeoff between operating range and resolution. SSS using low frequency is suitable for scanning large areas but cannot produce high quality images. In contrast, SSS using high frequency yields high resolution images over a narrow scanning range. Hence, there is a need to explore how to improve the image quality of low frequency SSS in practice. As SSS images are similar to optical images, a solution may be to apply conventional SR methods for optical images. Some studies have applied sparse coding to SSS images. The convolutional neural network (CNN) based method is a recent trend in optical image SR and its superiority has been proven [2]. Nevertheless, applying CNN based SR methods to SSS images is challenging due to the lack of a public database on which to train its model. Sparse coding not only works well without sufficient training data but its effectiveness for optical image SR has been proven by Yang et al. [3].

However, sparse coding based SR method requires huge computational load to accurately estimate the sparse coefficient of each image patch. As the sea speed decreases and the number of the sensors increases to search wide area in detail, the quantity of data size grows exponentially. As a result, the computational load of sparse coding makes it impractical to apply it for the SSS imaging system. One of the most important criteria by which to evaluate image quality is sharpness [4]. However, most regions in SSS images are flat sandy terrain, which contains few edges. As shown in Fig. 1, due to the lack of edge abundant objects, qualitative and quantitative difference between SR results for the sparse coding method and bicubic interpolation is imperceptible. Therefore, there is no need to apply sparse coding or other sophisticated SR methods to this or other such object-lacking regions since simple bicubic interpolation suffices.

In this paper, a region-selective sparse coding based SR system that emphasizes object regions is proposed to reduce processing time and preserve visual quality of output images. In order to distinguish object containing regions, we adopt scale invariant object detection algorithm suitable for underwater images. After detection, a sparse coding based SR algorithm is selectively applied to the object containing region. Experimental results prove the effectiveness of the proposed method.

2. Proposed Method

As shown in Fig. 2, the proposed method essentially consists of two steps: object-existent/nonexistent image classification using multi-scale features and super resolution via selective sparse coding. In Fig. 2, we slide a non-overlapping 240 × 240 pixels window across the whole SSS image and then each sub-image is categorized as an object-existent/nonexistent image by exploiting the object detection and image classification algorithms. For the object nonexistent sub-images, all pixels will be restored by only using bicubic interpolation. For object-existent sub-images, the background region will be processed by bicubic interpolation while those pixels that include objects will be recovered by a sparse coding algorithm.

2.1 Object-Existent/Nonexistent Image Classification

The proposed image classification method consists of five...
sequential steps, as illustrated in Fig. 3. First, $L$-layer image pyramid is made to cope with diverse scales of data. This practice is widely exploited in the computer vision field for scale robust object detection [5]. This image pyramid is constructed by scaling the input images using bicubic interpolation and each scaled image is processed separately until they are concatenated in pyramid as in step 4. Second, we use the Sobel edge detector [6] to calculate the threshold $T$, which is derived from the mean value of the gradient magnitude of the input image, and adjusted by the scale parameter $\delta$ as shown in Eq. (1).

$$ T = \sqrt{\frac{1}{MN} \sum_{i,j=1}^{N,M} (G_x^2 + G_y^2)}, \quad (1) $$

where $G_x$ and $G_y$ are the gradients in $x$ and $y$ directions, and $N$ and $M$ are the length of the image in $x$ and $y$ directions. Then the threshold $T$ is leveraged to acquire the binary image via the Sobel operator. If we just need to determine whether the sub-image is object-existent/nonexistent, the Sobel operation would be sufficient. However, segmenting object containing regions from background is needed to selectively apply the sparse coding based SR to the object regions only. For this reason, we dilate the binary image and fill the interior gaps using morphological operations [7], [8]. In this step, three types of masks, $[111]$, $[010]$, and $[111]^T$, are applied in order.

Then normalized feature values $f_l$, $l = 1, 2, \ldots, L$, are calculated by Eq. (2)

$$ f_l = \frac{1}{NM} \sum_{i,j=1}^{N,M} t_{i,j}^l, \quad (2) $$

where $t_{i,j}^l$ is a pixel value of the morphologically reconstructed image of $l$-th layer (i.e., 1 is assigned to object region and 0 is assigned to the rest) and these feature values are concatenated to form a feature vector $F = (f_1, f_2, \ldots, f_l)$. Due to the adaptive threshold based edge detection method, if noticeable objects are not present in the input image, a lower threshold $T$ would be assigned and many more edges will be detected. Hence, object-nonexistent images make the feature values high while object-existent images produce low feature values. Lastly, the feature vector $F$ is leveraged for the image classification. The boundary that delineates object-existent and -nonexistent images is derived by the linear support vector machine.

2.2 Super Resolution via Selective Sparse Coding

For the sparse coding based SR, we adopt the adaptive sparse domain selection (ASDS) method [9] and make it better suited to the SSS underwater imaging application. This method has shown superior performance compared with other SR methods. Furthermore, since SSS image patterns comprise relatively small variations, the number of patterns is more limited than those of the optical images. Hence, these images can be recovered effectively via specific sub-dictionaries derived by the domain selection algorithm.

When the bicubic-upsampled low resolution image patch is $y$ and the recovered high resolution image patch is $x$, $x$ can be derived by $x \approx D\alpha$ where $D$ is the pre-trained dictionary and $\alpha$ is the estimated sparse coefficient for the input $y$.

For the dictionary training, ASDS learns a series of compact sub-dictionaries $D_k$, $k = 1, 2, \ldots, K$, and adaptively applies it to each local patch instead of learning an over-complete single dictionary $D$. It reduces processing time while improves SR performance. Specifically, a training set of image patches $S$ passes through high-pass filter and then the resulting feature patches $S^f$ are clustered by $K$-means algorithm to divide $S$ into subsets $S_k$. Then, each sub-dictionary is trained by the corresponding subset as follows:

$$ D_k, \alpha_k = \arg \min_{D_k, \alpha_k} ||S_k - D_k\alpha_k||_2^2 + \lambda||\alpha_k||_1, \quad (3) $$

where $\lambda$ is a constant and $\alpha_k$ is $k$-th coefficient matrix. Equation (3) can be approximated by principal component analysis for the subsets $S_k$.

To recover high resolution image patch, the most appropriate sub-dictionary for each input patch is selected first as

$$ k = \arg \min_k ||y^h - \mu^k||_2, \quad (4) $$

where $y^h$ and $\mu^k$ are high-pass filtered input patch and centroid of subset $S_k$, respectively. In ASDS, eigenvectors of $y^h$ and $\mu^k$ are used to calculate the distance but it is shown not effective in this application. Then, coefficient $\alpha$ is estimated.
by the spatially adaptive regularization method, i.e.,
\[
\hat{\alpha} = \arg \min_{\alpha} \|y - S\Delta D_\alpha \alpha\|^2 + \gamma \| (I - A) D_\alpha \alpha \|^2 + \eta \| (I - B) D_\alpha \alpha \|^2 + \sum_{i=1}^N \sum_{j=1}^n \lambda_{i,j} |\alpha_{i,j}|, \tag{5}
\]
where \(\alpha_{i,j}\) is \(j\)-th atom of the coefficient for \(i\)-th patch, \(S\) is a down-sampling, \(H\) is a blurring operator, matrix \(A\) contains autoregressive model parameters, \(B\) is the weight matrix for non-local similarity, and Eq. (5) is solved by FISTA algorithm [10].

For Eq. (5), we propose region-selective sparse coding (RSSC) method, which updates the elements of regularization matrices \(A\) and \(B\) only for the objects-existent pixels which are classified by the morphological reconstruction image of the third layer in Fig. 3. As a result, the processing time of SR is reduced due to the use of more sparse regularization matrices.

In the ASDS method, non-local similarity regularization contributes to preserving edge sharpness and reducing noise. For each local patch, it searches similar patches in wide surrounding areas to regularize the local patch. However, object-existent and -nonexistent regions in SSS images do not share the same characteristics as mentioned in the introduction. As a consequence, this regularization is less effective for the application. For this reason, we introduce “region selective sparse coding–nonlocal (RSSC-NL)” which essentially reforms the non-local similarity regularization in RSSC. In particular, for each local patch of object-existent region, we search similar patches in the same region to improve effectiveness of the regularization.

3. Experimental Results

This section describes how the performance of the object-existent/nonexistent image classification and the performance of the image SR using the resulting region-selective sparse coding are verified. These experiments are carried out on a PC platform with a 2.1 GHz CPU and 64 GB RAM under the MATLAB R2017a programming environment.

3.1 Dataset and Experimental Settings

As shown in Fig. 4, a multi-scale SSS image dataset is used to verify the practical effectiveness of the proposed methods. In this dataset, the frequency range of SSS is 300∼780 kHz and distance (from the sensor to seabed) range is 25∼50 m. It consists of 90 object-existent images and 180 object-nonexistent images where the size of cropped image is 240 × 240. This dataset is divided into three sets for cross validation. These images are considered as original image and input image is generated by applying 7 × 7 Gaussian kernel with a standard deviation of 1.6 to the original image then down-sampling it by factors of 3, 4, or 5. We choose \(\delta = 0.24\), \(K = 165\), \(\gamma = 0.074\), \(\eta = 0.1386\), and \(\lambda_{i,j}\) is computed in the same method with previous work [9]. All the parameters are chosen via cross-validation. Data is labeled via a coarse-refine procedure. All patches are coarsely screened by comparing the mean intensity values against a threshold and then are manually corrected to refine the results.

3.2 Object-Existent/Nonexistent Image Classification

The average results by the number of layers in pyramid are described in Table 1 where the scale factor is 5. Although the single layer shows same false positive rate (FPR) as the five layer pyramid, we can assess that the five layer pyramid obtains better performance considering that detection rate (DR) is more critical than FPR because the object region is our region of interest. As the dimensions of the feature increase according to the number of layers, processing time also increases. However, this impact is negligible. So applying the multi-layer pyramid is more practical classification method. Then the percent matched between manual and detected object regions is 91.3%.

3.3 Super Resolution via Selective Sparse Coding

The proposed RSSC and RSSC-NL are compared with four competing SR methods: bicubic interpolation (BC), SC [3], VDSR [2], and ASDS [9]. The 90 object-existent images in the dataset are used for the dictionary training (60 images) and image SR (30 images). The average ratio of manually labeled object regions is 73.5%. The degraded input image and the morphological reconstruction image of the third layer are up-sampled to the original image size and SR method is applied where patch size is 7 × 7 and stride is 2. Then, the output image is compared with the original image to evaluate SR performance. In this experiment, PSNR, SSIM, and processing time (second) are used as performance evaluation metrics.

The quantitative results are shown in Table 2. RSSC is clearly shown to outperform BC, SC and VDSR with regard to PSNR and SSIM. Although VDSR is very fast, image quality has the highest priority for the intended object-existent SSS image SR application, e.g., underwater exploration. RSSC-NL, which searches similar patches only in

Table 1 Results of object-existent/nonexistent image classification.

<table>
<thead>
<tr>
<th># of layers</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Used scales</td>
<td>1</td>
<td>1, 1/2</td>
<td>1, 1/2</td>
<td>1, 1/2</td>
<td>5/2, 1, 1/2</td>
</tr>
<tr>
<td>DR (%)</td>
<td>91.7</td>
<td>93.3</td>
<td>93.8</td>
<td>94.7</td>
<td>96.9</td>
</tr>
<tr>
<td>FPR (%)</td>
<td>2.50</td>
<td>1.67</td>
<td>1.73</td>
<td>2.01</td>
<td>2.49</td>
</tr>
</tbody>
</table>
Table 2  Average PSNR, SSIM, and processing time of the different methods for the object-existent images.

<table>
<thead>
<tr>
<th>Method</th>
<th>Scale</th>
<th>PSNR</th>
<th>SSIM</th>
<th>Time</th>
<th>PSNR</th>
<th>SSIM</th>
<th>Time</th>
<th>PSNR</th>
<th>SSIM</th>
<th>Time</th>
<th>PSNR</th>
<th>SSIM</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>BC</td>
<td>3</td>
<td>22.01</td>
<td>0.418</td>
<td>0.02</td>
<td>22.13</td>
<td>0.442</td>
<td>33.6</td>
<td>22.25</td>
<td>0.358</td>
<td>0.24</td>
<td>23.35</td>
<td>0.571</td>
<td>152</td>
</tr>
<tr>
<td>SC [3]</td>
<td>4</td>
<td>21.24</td>
<td>0.367</td>
<td>0.02</td>
<td>21.04</td>
<td>0.367</td>
<td>36.1</td>
<td>21.24</td>
<td>0.373</td>
<td>0.24</td>
<td>22.79</td>
<td>0.474</td>
<td>139</td>
</tr>
<tr>
<td>VDSR [2]</td>
<td>5</td>
<td>20.84</td>
<td>0.333</td>
<td>0.02</td>
<td>20.11</td>
<td>0.357</td>
<td>35.3</td>
<td>19.86</td>
<td>0.319</td>
<td>0.28</td>
<td>22.04</td>
<td>0.412</td>
<td>122</td>
</tr>
<tr>
<td>ADS [9]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RSSC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RSSC-NL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 5  Comparison of SR results using BC, SC [3], VDSR [2], ADS [9], and the proposed methods (RSSC, RSSC-NL) on SSS images at a scale of 3.

4. Conclusions

In this paper, we proposed a practical SR system for SSS images. The object-existent region, which is our region of interest, was effectively recovered by the proposed region selective sparse coding (RSCC) method. Simultaneously, the object-nonexistent region, which is less interesting, was restored by bicubic interpolation to save processing time. The proposed object-existent and -nonexistent image classification scheme enabled the system to selectively apply the SR methods according to region of interest. We also proposed RSSC-NL, which reformed the non-local similarity regularization in RSSC. The proposed RSSC-NL further improved SR performance for the SSS image. We experimentally demonstrated the effectiveness of our method, which reduced the processing time while preserving the same level of PSNR and SSIM as the conventional method.

Acknowledgments

This research was funded by the ADD of Korea (UD160015DD).

References