NSCT-based fusion method for forward-looking sonar image mosaic

Jian Zhang^{1,2}, Ferdous Sohel³, Mohammed Bennamoun², Hongyu Bian^{1*}, Senjian An²

¹Acoustic Science and Technology Laboratory, College of Underwater Acoustic Engineering,
Harbin Engineering University, China

²School of Computer Science and Software Engineering, The University of Western Australia, Australia

³School of Engineering and Information Technology, Murdoch University, Australia

*Corresponding Author: bianhongyu1116@hrbeu.edu.cn

Abstract: To generate a seamless mosaic of a Forward-Looking Sonar (FLS) video sequence, this study proposes a novel fusion method for FLS image mosaic, which includes two main steps from coarse to fine. In the coarse fusion step, the source images are first decomposed into multi-scale sub-bands using the Non-Subsampled Contourlet Transform (NSCT), then the sub-bands of the source images are merged, based on the Gabor energy and the local contrast for low and high frequency sub-bands respectively, to generate the fused sub-bands which are used to produce the fused image by inverse NSCT. In the fine fusion step, a decision map is used to choose the pixels of the fine fused image from the coarsely fused image or one of the source images. This decision map is first constructed by measuring the similarity of the coarsely fused image to the source images, and then processed by a morphological post-processing technique to ensure its continuity and smoothness. Extensive experiments on FLS image fusion and mosaic have been conducted to demonstrate the effectiveness and the superiority of the proposed technique using both a subjective evaluation and objective metrics.

1. Introduction

Two-dimensional Forward-Looking Sonar (FLS) [1, 2] has become one of the most effective tools for the exploration of turbid waters, in which low visibility restricts the application of optical cameras. FLS can provide acoustic images at high frame rates similar to videos. However, due to the limited number of transducer beams, the azimuth field of view of an FLS image is too narrow. In some practical applications such as underwater structure inspection and wreckage detection, only a small portion of large objects (e.g. bridge and ship wreckage) can be detected in a single narrow frame. It is not convenient for visualization or computer techniques to deal with a partial view of an object in a frame or an object spanning across multiple frames. It is therefore important to integrate multiple frames, with overlapping regions, into one composite image that offers an overall view of the field of interest. This integration technique is referred to as image mosaic.

One of the earliest works on image mosaic can be traced back to [3]. The typical workflow of such a technique has three main blocks: image pre-processing, image registration with global alignment, and image fusion. This technique has been widely adopted in a number of image processing applications, including the FLS images [4, 5, 6, 7]. However, these works [4, 5, 6, 7] for FLS image mosaic have issues with sharp transitions along the boundaries between the overlapping images. These sharp transitions, caused mainly by the inhomogeneity (see Section 2) between the corresponding images, produce obvious seams and therefore adversely-affect the quality of the mosaic. To deal with this problem, image fusion techniques are applied to suppress the differences along these boundaries. In [8], a weighted average based fusion approach is used to eliminate the seams. To improve the Signal to Noise Ratio (SNR) of the mosaic, the weight values are distributed according to

the pixels'insonification intensities. However, because the mosaic is created by a small number of fully overlapping frames, the field of view of the mosaic is not enlarged. To solve this problem, Li et al [9] use the Scale Invariant Feature Transform (SIFT) features with the RANdom Sample Consensus (RANSAC) method for sonar image registration. After registration, the input frames are integrated to generate a panoramic image containing the integral object. Meanwhile, a fusion algorithm based on a weighted threshold and the Canny edge detector is used to distinguish between the image boundaries and the edges of the objects in the images. If a pixel is determined as the boundary of a source image, it will be excluded and replaced by the corresponding pixel from another source image to ensure smoothness in the transitions. Another mosaic based FLS image fusion scheme is proposed in [10]. A saliency mask obtained by computing the standard deviation of the local regions is used to remove the useless areas of an FLS image. Then, the remaining regions are smoothed using Gaussian convolution and fused by taking the average of the useful regions from all the images. These fusion methods improve the smoothness of the transitions by a simple averaging of the input images, which however, brings undesirable impacts such as a reduced contrast [11]. Moreover, neither the scale nor the directional information of the source images is taken into consideration by using the average calculation. To overcome these problems, multi-scale transform (MST) based fusion methods are proposed in [12, 13, 14, 15, 16].

The main steps of MST-based image fusion can be summarized as follows. First, the source images are decomposed into several sub-bands in different scales and directions using MST tools. Then the elements of the sub-bands are merged by the fusion rules, resulting in the calculation of fused sub-bands. Finally, the fused image is reconstructed by inverse MST using the fused sub-bands.

The image information with different scales and directions can be represented from the decomposed sub-bands. The MST-based fusion method originates from [12], in which the source images are decomposed and reconstructed using the Laplacian Pyramid Transform (LPT). Then the Gradient Pyramid and some Wavelets, such as the Discrete Wavelet Transform (DWT) are successively used for image fusion [13, 14, 15, 16]. However, these MST tools with limited directions cannot effectively represent the discontinuities along the edges and contours of the objects in the image [17]. Furthermore, due to sampling during image decomposition and reconstruction, the pseudo-Gibbs phenomenon which exhibits artefacts around the pixels with high gradients, such as object edges and image boundaries, affects the quality of the fused image. To overcome these problems, the work in [17] proposes the Contourlet Transform (CT), which employs the LPT to find the discontinuities and uses Directional Filter Banks (DFB) to represent the anisotropy in the image. But it fails in suppressing the pseudo-Gibbs phenomenon since it involves the up and down sampling as well. As an improved version of CT, the Non-Subsampled Contourlet Transform (NSCT) [18] successfully eliminates the pseudo-Gibbs phenomenon using non-subsampling for the decomposition and reconstruction, making it more suitable for MST-based image fusion [19]. However, MST-based image fusion has not been widely applied in FLS sonar image fusion. It is difficult to determine appropriate fusion rules due to the properties (such as the lack of features, high noise level and inhomogeneity) of FLS images. In addition, it is inevitable to lose useful information of the source image in the process of decomposition and reconstruction by MST [20]. Another group of image fusion methods, namely spatial-based fusion [21], can retain the complete information of the source images without using any decomposition and reconstruction. Instead, they divide the images into several blocks, and then select the blocks from the source images to the fused image according to some clarity measurements. These techniques are likely to produce blocky artefacts due to the inappropriate block division [22], thereby influencing the quality of the fused image. In this paper, a novel fusion method, combining the advantages of the spatial-based methods and the MST-based methods is proposed for FLS image fusion. The main contributions of this work are listed below.

- (1) A novel algorithm for the fusion of FLS images is introduced, which is composed of a coarse fusion using NSCT and a fine fusion based on the determination of the useful regions.
- (2) Our method achieves a higher performance with respect to metrics such as mutual information, edge based similarity measure, and signal to noise ratio. Higher quality can also be visually verified from our results in Figures 5-9.
- (3) A post-processing method based on morphology guarantees the continuity of the finely fused image. This also allows for a higher tolerance to image registration errors.

The rest of this paper is structured as follows. The motivation of FLS image fusion and the related background are presented in Section 2. Section 3 describes the methodology of the proposed fusion method. Experimental

results and analysis are reported in Section 4, while the conclusions are provided in Section 5.

2. Background

In an FLS imaging system, the sound waves are first emitted from the transducer arrays. After insonification and scattering, the echo signals are received by the receiving arrays. The received signals are then sampled accordingly with the spatial orientation in polar coordinates, and transformed into a fan-shaped image in Cartesian coordinates. Similar to optical images, the FLS images represent the reflection energy distribution in the image plane. However, the limited number of arrays restricts the field of view of the FLS images. For example, some of the FLS data in this work is recorded by the ARIS sonar [1]. There are 96 or 128 transducers where each beam has a width of 0.1-0.3 degree and 1500-2000 samplings. The horizontal field of view is approximately 28 degrees, which makes it difficult to detect large structures, such as bridges, dams and ship hulls, in a single frame. Therefore, it is necessary to integrate multiple overlapping frames to generate a mosaic that provides an overall field of view. Due to the high frame rate of FLS devices, a high overlap percentage is common between consecutive and near-consecutive frames. However, some inherent properties of FLS imaging may influence the quality of the mosaic.

As shown in Fig.1, FLS images suffer from inhomogeneity caused by the nonuniform sensitivity of the acoustic lens [4, 23]. This property leads to sharp transitions and seams along the boundaries of multiple frames in the mosaic. To improve the quality of the mosaic, the seams caused by the inhomogeneity of different frames need to be eliminated. One of the most commonly used solutions is to take the average of the overlapping regions. But this will lead to undesirable influences, such as reduced contrast since the pixels with high contrast usually contain useful information [24]. In addition, objects that are near the focused regions appear brighter and clearer while objects at a distance from the focused area appear darker and blurred. The change of viewpoint in detection leads to a phenomenon where the same object appears bright and clear in some frames but dark and unclear in some other frames. For further applications, the clear regions are expected to be included in the mosaic to improve the brightness and clarity of every object, for which a simple averaging is obviously not suitable. Based on the above analysis, a fusion method should pay more attention to the combination of the complementary information from the source images to generate a seamless mosaic.

CT consists of LPT and DFB. An image is decomposed into one low-frequency and several high-frequency sub-images, and each high-frequency sub-image can be also divided into different directional sub-bands, enhancing CT with a more comprehensive representation of the image. However, the up and down sampling operations during the decomposition and reconstruction cause the pseudo-Gibbs phenomenon, which translates as artefacts around the objects edges and the boundaries of the image, making it not perfect for image fusion. As a non-subsampling scheme of CT, NSCT is proposed to suppress these artefacts. NSCT is composed of Non-Subsampled

Fig.1 FLS images suffer from nonuniform illumination and clarity
a The background intensities of FLS images vary according to the beams position
b The background intensities of FLS images vary according to the sampling position
c A change of view leads to a clear object on the left but a blurred object on the right in the left frame, and a clear object on the right but a blurred object on the left in the right frame

Pyramid Filter Banks (NSPFB) and Non-Subsampled Directional Filter Banks (NSDFB). NSPFB is a two-channel non-subsampled filter bank. The input image can be decomposed into one low-frequency and one high-frequency sub-image by NSPFB. By iteratively using NSPFB to divide the low-frequency component k times, the image can be decomposed into $k+1$ levels of sub-images, including one low-frequency and k high-frequency sub-images. The bandpass of the low-frequency sub-image on the k 'th decomposition is $[-\pi/2^k, \pi/2^k]$, and the bandpass of the corresponding high-frequency component is between $[-\pi/2^{k-1}, \pi/2^{k-1}]^2$ to $[-\pi/2^k, \pi/2^k]^2$. The NSDFB is a two channel non-subsampled directional filter bank that is

composed of directional fan filter banks and DFB. With an l level decomposition, one high-frequency sub-band can be decomposed into 2^l directional sub-bands using NSDFB as shown in Fig.2a. These enable NSCT to represent effectively the multi-scale and anisotropy of the image. With the help of the non-subsampled structure, the pseudo-Gibbs phenomenon can successfully be overcome by avoiding the up and down sampling during the decomposition and reconstruction using NSCT. In addition, the non-subsampled scheme maintains the same size for all of the sub-bands with the input image. This makes it convenient to build fusion rules for the corresponding sub-bands of the source images and thereby, making the NSCT more suitable for image fusion. As shown in Fig.2b, for two source images A and B, the steps of NSCT-based fusion method [24] are summarized as follows:

- (1) Source images A and B are decomposed in K scales and L directions using NSCT, resulting the sub-bands of the source images, $\{L^A, H_{kl}^A\}$ and $\{L^B, H_{kl}^B\}$. L^A and L^B represent the low-frequency components of A and B, and H_{kl}^A and H_{kl}^B are the high-frequency sub-bands at scale $k \in [1, K]$ and direction $l \in [1, L]$, respectively.
- (2) The rules for the high- and low-frequency sub-bands fusion are established, and the fused sub-bands $\{L^F, H_{kl}^F\}$ are then obtained.
- (3) Perform the inverse of NSCT on the fused sub-bands to generate the fused image

3. Methodology

The proposed fusion method includes three steps: pre-processing, the coarse fusion and the fine fusion. The coarse fusion is based on NSCT. The coarsely fused result is then used to create a decision map which is used to select the pixels for the fine fused image from the coarsely fused image or one of the source images.

Fig.2 Illustration of NSCT and the framework of NSCT-based image fusion

a Illustration of NSCT

b The framework of NSCT-based image fusion

3.1. Pre-processing

It can be seen that FLS images suffer not only from inhomogeneity but also from low contrast [10, 25]. Their gray values are restricted to a small dynamic range. To enhance the contrast, the distribution of the intensity levels needs to be more expanded. In this work, a weighted threshold histogram equalization [26] scheme is used for contrast enhancement. The probability density function (PDF) of the input FLS frame is denoted as

$$P(k) = \frac{n_k}{N} \quad (1)$$

where $k \in [0, 255]$ is the gray level of the pixel, and n_k is the number of the pixel with a gray value equal to k . Then the equalization scheme is designed based on the following rule

$$P_e(k) = \begin{cases} P_u & \text{if } P(k) > P_u \\ \left(\frac{P(k) - P_l}{P_u - P_l} \right)^r \times P_u & \text{if } P_l \leq P(k) \leq P_u \\ 0 & \text{if } P(k) < P_l \end{cases} \quad (2)$$

where $P_u = v \times P_{\max}$, $v \in [0, 1]$, and $P_{\max}$ is the highest probability in the PDF of the input image. The role of v is to restrict the highest probability of new PDF to be less than P_u . P_l is the threshold and any value of the probability which is lower than P_l is set to 0. The parameter $r \in [0, 1]$ is used to enlarge the dynamic range of the gray values by relocating the original probability to the interval $[P_l, P_u]$. Then the new cumulative distribution function (CDF) is computed as

$$C_e(k) = \sum_{m=0}^k P_e(m) \quad k \in [0, 255] \quad (3)$$

Finally, using the CDF, the histogram equalization is given as

$$I_{\text{new}}(i, j) = W_{\text{out}} \times C_e(I_{\text{original}}(i, j)) + M_{\text{adj}} \quad (4)$$

The $I_o(x, y)$ and $I_e(x, y)$ are the original and equalized gray value of a pixel (x, y) . $W_{\text{out}} = \min(255, G_{\max} \times W_{\text{in}})$ and W_{in} are the dynamic ranges of the equalized and original images respectively. $G_{\max}$ is set to 2.5-3.5. M_{adj} is used to compensate for the difference between the images' mean value after equalization. An example of the results to illustrate the contrast enhancement is shown in Fig.3b. It can be seen that compared to the original image, the object, shadow and background become more distinguished due to a more enhanced contrast.

Before fusion, the source frames should be spatially aligned to avoid inconsistencies in the results. Two datasets are used in this paper. The first FLS video sequence is recorded by the ARIS series sonar rotating around the AR2 rotator [27], which precisely records the rotation. The second dataset consists of some frames in the presence of translation. This dataset is acquired using DIDSON series. Conventional GPS and differential GPS (DGPS) receivers do not work underwater and the localization equipment cannot provide accuracy up to a pixel level [28]. Therefore image registration is conducted to align these frames. For FLS images registration, the intensity based methods usually outperform the feature based methods since FLS images do not usually exhibit highly distinctive features and therefore

do not correctly match the features [4, 5, 6, 10, 29]. Although intensity based methods are sensitive to the spatially varying intensities between FLS frames [30], they can still perform satisfactorily in the registration of consecutive frames whose intensities can be approximately regarded as homogeneous. Then the near-consecutive frames can be aligned by accumulating the consecutive registration results without introducing significant errors. We use the Fourier based registration method [4], which calculates the rotation by first calculating the translation between the FLS frames in polar coordinates and then calculates the translation in the Cartesian coordinates. The result is shown in Fig. 3c and d.

3.2. Proposed Coarse FLS Image Fusion based on NSCT

The proposed coarse fusion is based-on NSCT and the flowchart of the coarse-fusion is shown in Fig.2b. In general, the elements of fused sub-bands are selected from the corresponding sub-bands of the source images through specific rules. Hence, a key step of the NSCT-based fusion method is the establishment of fusion rules, which should be designed according to the characteristics of the useful information in the source images. Typically, parameters such as the gradient and the spatial frequency, which measure the details in the image, are regarded as important for optical images [31], since they have rich features and low noise. In contrast, FLS images are usually featureless and noisy, which makes the mentioned parameters not suitable to represent the useful information of FLS images. Therefore, some other parameters need to be considered to set up the fusion rules for FLS images.

Fig.3 FLS image contrast enhancement and near-consecutive frames registration

- a The original FLS image
- b The contrast enhancement result
- c Overlay the source images by blending after registration
- d A difference image between the source images after registration

3.2.1 Fusion Rule of Low-Frequency Sub-band: An FLS image expresses the intensity distribution of the echo spatially and the clear regions usually represent a high reflected energy. Since most of the image energy is retained in the low-frequency sub-band, parameters which represent the image energy should be used to guide the fusion rule for the low-frequency sub-band.

Gabor energy is an effective tool to describe the energy distribution in an image [32]. It also exhibits a satisfactory performance in FLS image applications [33] due to the superior invariance to noise and transformation. Hence, the Gabor energy is suitable for the low-frequency sub-band fusion. Gabor transform can be originated from [34]. The low-frequency sub-band is convoluted with the Gabor filter banks to generate Gabor coefficients at different orientations and scales. The Gabor filter at the m th scale and n th orientation is defined as

$$g_{mn}(x, y) = a^{-m} g(x', y'), \quad a > 1,$$

$$m = 1, 2, \dots, m_0, \quad n = 1, 2, \dots, n_0,$$

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = a^{-m} \begin{bmatrix} \cos \theta_n & \sin \theta_n \\ -\sin \theta_n & \cos \theta_n \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \quad \theta_n = \frac{n\pi}{n_0}. \quad (5)$$

where a is the scale parameter, m_0 and n_0 are the total number of scales and orientations, respectively. $g(x, y)$ is the 2D Gaussian kernel function. Then the Gabor coefficients $C_g(x, y, m, n)$ at scale m and orientation n can be obtained as

$$C_g(x, y, m, n) = \iint L(x', y') g_{mn}(x - x', y - y') dx' dy' \quad (6)$$

where $L(x, y)$ is the input matrix, namely the low-frequency sub-band. The Gabor energy $E_g(x, y)$ is calculated by summing up the coefficients across all of the scales and orientations at pixel (x, y) :

$$E_g(x, y) = \sum_{m=1}^{m_0} \sum_{n=1}^{n_0} |C_g(x, y, m, n)|^2 \quad (7)$$

The fusion rule of the low-frequency sub-bands based on the Gabor energy is constructed as

$$L^F(x, y) = \begin{cases} L^A(x, y), & E_g^A(x, y) - E_g^B(x, y) > T, \\ L^A(x, y) \times w^A + L^B(x, y) \times w^B, & |E_g^A(x, y) - E_g^B(x, y)| \leq T, \\ L^B(x, y), & E_g^A(x, y) - E_g^B(x, y) < -T. \end{cases} \quad (8)$$

where $w^A = E_g^A / (E_g^A + E_g^B)$, $w^B = E_g^B / (E_g^A + E_g^B)$ is the weighted value, E_g^A and E_g^B are the Gabor energy of the low-frequency sub-bands of the source images. L^A , L^B and L^F are the low-frequency sub-bands of the source images A, B and the fused image F, respectively. T is a threshold which is set to 4 in our experiments. When there is a big difference between the sub-bands of the source images (i.e., when the absolute difference is larger than the threshold) the fused elements are directly chosen from the source images. When the difference is less than the threshold, the fused elements are calculated by the weighted averaging to ensure the stability of the fused image.

3.2.2 Fusion Rule of High-Frequency Sub-bands: The high-frequency sub-bands represent the sharp features in images, such as corners, edges and contours. Generally, the

rule of the maximum absolute value is used for the high-frequency sub-bands fusion to retain the details of the image. However, in addition to these image details, image noise is also enclosed in the high-frequency sub-bands. Especially for FLS images, the noise of the original images can be introduced to the fused high-frequency sub-bands using the maximum absolute value rule. In addition, the maximum absolute values rule does not consider the relationship between the low-frequency and high-frequency sub-bands. According to the human visual system (HVS) characteristics, humans are more sensitive to contrast than illumination [24, 35]. Hence, the fusion rule for the high-frequency sub-bands should only consider the contrast. The contrast R of the image is defined as [36]:

$$R = \frac{L - L_b}{L_b} = \frac{\Delta L}{L_b} \quad (9)$$

where L is the gray value of the local region and L_b is the brightness of the background, namely the low-frequency sub-band, and therefore ΔL denotes the high-frequency sub-bands. Then the local directional contrast $R_{kl}(x, y)$ of pixel (x, y) at k scale and l orientation is defined as [37]

$$R_{kl}(x, y) = \begin{cases} \frac{|H_{kl}(x, y)|}{\bar{L}(x, y)}, & \bar{L}(x, y) \neq 0, \\ |H_{kl}(x, y)|, & \bar{L}(x, y) = 0. \end{cases} \quad (10)$$

In Equation (10), the $H_{kl}(x, y)$ is the high-frequency sub-bands at k scale and l orientation at pixel (x, y) , and $\bar{L}(x, y)$ is the average of the low-frequency sub-bands in the local region:

$$\bar{L}(x, y) = \frac{1}{M \times N} \sum_{p=-(M-1)/2}^{(M-1)/2} \sum_{q=-(N-1)/2}^{(N-1)/2} L(x + p, y + q). \quad (11)$$

The window size $M \times N$ is commonly set to 5×5 or 7×7 . Then the fused high-frequency sub-bands $H_{kl}^F(x, y)$ are determined by the local directional contrast of the source images A and B:

$$H_{kl}^F(x, y) = \begin{cases} H_{kl}^A(x, y), & R_{kl}^A(x, y) \geq R_{kl}^B(x, y), \\ H_{kl}^B(x, y), & R_{kl}^A(x, y) < R_{kl}^B(x, y). \end{cases} \quad (12)$$

Here, the elements with a larger contrast are regarded as useful information and are selected for the fusion.

To suppress noise which is usually isolated and randomly located, a region consistency check is performed on the fused sub-bands. The consistency decision map M of the fusion is given by

$$M(x, y) = \begin{cases} 1, & R_{kl}^A(x, y) \geq R_{kl}^B(x, y), \\ 0, & R_{kl}^A(x, y) < R_{kl}^B(x, y). \end{cases} \quad (13)$$

According to this map, if the element at pixel (x, y) came from source image A with the majority of its neighbours being from image B, then the fused element will convert to the element from image B.

Once the fused elements of the low- and high-frequency sub-bands are obtained, the coarsely fused image can then be generated using the inverse NSCT on the fused sub-bands.

3.3. Proposed Fine Fusion Based on Useful Regions Determination

The proposed fusion method based on useful regions determination is described in detail in this section. The

fusion scheme can be divided into two steps: useful regions determination and image fusion based on the selected regions.

3.3.1 Useful Regions Determination: Although the coarse fusion has obtained the complementary information from the source images, it still loses some important information in the processing of the decomposition and reconstruction by NSCT. The most effective solution to address this problem is to directly copy the pixels from the source images to the fused image. Therefore, it is necessary to determine which regions between the source images are more useful in order to use these regions for the fused image. In addition, the fused results should not contain any seams or other artefacts that are introduced by differences between the source images.

For the source images, the gray values of the important pixels should be similar to the corresponding pixels in the coarsely fused image. The Root Mean Square Error (RMSE) metric [38] can be used to measure the similarity between the source images and the coarsely fused image pixel by pixel. The RMSE of each pixel between the source images and the image is calculated as follows

$$RMSE_A(x, y) = \left(\frac{1}{M \times N} \sum_{m=-(M-1)/2}^{(M-1)/2} \sum_{n=-(N-1)/2}^{(N-1)/2} (I_F(x+m, y+n) - I_A(x+m, y+n))^2 \right)^{1/2}$$

$$RMSE_B(x, y) = \left(\frac{1}{M \times N} \sum_{m=-(M-1)/2}^{(M-1)/2} \sum_{n=-(N-1)/2}^{(N-1)/2} (I_F(x+m, y+n) - I_B(x+m, y+n))^2 \right)^{1/2} \quad (14)$$

where I_A , I_B and I_F are the source image A and B, and the coarsely fused image F, respectively. The useful regions can be determined by comparing the RMSE. A binary matrix is used to represent the regions distribution. The mathematical expression of the map FMAP is constructed as follows

$$FMAP(x, y) = \begin{cases} 1 & \text{if } RMSE_A(x, y) \leq RMSE_B(x, y) \\ 0 & \text{if } RMSE_A(x, y) > RMSE_B(x, y) \end{cases} \quad (15)$$

When the $RMSE_A(x, y)$ is smaller than $RMSE_B(x, y)$, the value of $FMAP(x, y)$ will be set to 1, indicating that pixel (x, y) of image A is similar to the corresponding pixel in the coarsely fused image. It is then considered to be more useful than the corresponding pixels from image B. Otherwise the corresponding pixel in image B is considered to contain more important information. Therefore, the useful regions can be determined by this binary fusion decision map FMAP.

However, this decision map cannot be directly applied for the fine fusion. Because of the coarse quality of the FLS images, some narrow breaks and small spots appear around the boundaries of the black and white regions in the map, causing discontinuities in the fine fused image. To improve the fusion quality, we propose a morphological post-processing to remove these defects. It consists of the following steps:

Step1: Suppress the thin structures that are lighter than their surroundings and connect to the boundary. This step is first used to remove the small white spots and narrow white bands around the boundary, guaranteeing the continuity of the black regions of the binary map.

Step2: Erase the small black holes in the white region to improve the smoothness of the white region.

Step3: The morphological opening and closing operations are iteratively used to eliminate the narrow breaks and thin protrusions along the boundary, making the boundary as smooth as possible.

Finally, the ultimate decision map UMAP is made to show the useful regions of the source images.

3.3.2 Fine Fusion Based on Useful regions Determination: The proposed FLS image fusion method is, in essence, a spatial-based method. It takes advantage of the MST-based fusion method to generate a coarse fusion result to guide the partition of the source images. The fusion algorithm based on the useful regions determination consists of the following steps.

Step1: Perform coarse FLS image fusion based on NSCT. Meanwhile, the Gabor energy and the local contrast are used for the fusion rules to merge the low- and high-frequency sub-bands:

Step2: Following the coarse fusion, the similarity between the coarse fused image and the source images is measured using RMSE to generate a binary matrix. This matrix is then modified using morphological operations to guarantee the continuity of the partitioned regions.

Step3: The fine fused image is obtained using the pixels from the source images or the coarsely fused image according to the decision map UMAP. The process of generating the fine fused image $I_{FF}(x, y)$ is given as

$$I_{FF}(x, y) = \begin{cases} I_A(x, y), & \text{if } T(x, y) = (2M+1)(2N+1), \\ I_B(x, y), & \text{if } T(x, y) = 0, \\ I_F(x, y), & \text{otherwise.} \end{cases}$$

$$T(x, y) = \sum_{m=-M}^M \sum_{n=-N}^N UMAP(x, y) \quad (16)$$

where $T(x, y)$ is the sum of the neighbours surrounding the pixel (x, y) in the decision map and the neighbourhood was set as 3×3 in our experiments. $T(x, y) = (2M+1)(2N+1)$ implies that the value of the neighbouring pixels of (x, y) are all equal to 1, indicating that the pixel from I_A can be regarded to be more useful. The corresponding pixel value $I_{FF}(x, y)$ is accordingly set to $I_A(x, y)$. $T(x, y) = 0$ illustrates that the pixels (x, y) from I_B contain more useful information and should be selected to be included in the fine fused image. Besides, $0 < T(x, y) < (2M+1)(2N+1)$ represents that some values of the neighbours of pixel (x, y) are 1 and the others are 0. Therefore, these pixels can be considered to be located around the boundary of the regions, and their values are assigned by the coarse fused image I_F to keep a smooth transition region.

4. Experimental Results and Discussions

4.1. Evaluation Metrics

Image fusion should not only transmit the complementary information from the source images, but also effectively suppress any artefacts in the fused image. The adverse-effects of artefacts can be observed and subjectively measured by a subjective observation, while the amount of information transmitted to the fused image is

measured using objective measures. We use three popular metrics for objective evaluations of the results.

Mutual Information (MI) [39]: MI measures the similarity of the gray level distributions between the source and the fused images. It reflects how much information the source images deliver to the fused images. Therefore, a larger value of MI denotes better fusion results.

Edge based Similarity Measure ($Q^{AB/F}$) [40]: $Q^{AB/F}$ assesses the fusion effects by measuring the edges similarity between the source and fused images. Therefore, a larger $Q^{AB/F}$ denotes a better fusion result.

Peak Signal to Noise Ratio (PSNR) [41]: PSNR is the ratio which measures the maximum possible power of a signal, and the power of the noise which affects the fidelity [19]. A larger PSNR of an image means a better endurance to noise. So the larger PSNR is, the better the fusion result is.

4.2. Experimental Results of FLS Frames Fusion

Experiments were carried on two conditions, namely the fusion with perfect and imperfect registration of FLS images. Examples of two pairs of the source images are shown in the Fig.4. To show the advantages of the proposed FLS image fusion method, the proposed method is compared with three fusion schemes, LPT with maximum absolute value rule, NSCT with Sum-Modified-Laplacian (SML) rule [31, 42] (NSCT_1), NSCT with the proposed rules (NSCT_2) in Section 3.2. SML method extracts the image details using the second derivatives, making it popular for optical fusion.

Fig.4 Source images for the fusion with perfect and imperfect registration

a The overlapping regions in frames 338 and 368 used as the source images A and B for fusion with perfect registration;
b The overlapping regions in frames 255 and 285 used as the source images A and B for fusion with imperfect registration;

4.2.1 FLS Image Fusion with Perfect Registration: The FLS video used in the experiment was recorded by ARIS series [1]. The sonar was mounted on a tripod and the rotation was obtained from the rotator pan. So the source images can perfectly be aligned spatially. Fig.4*a* shows the overlapping regions of 338th and 368th frames, which act as the source images A and B. It can be seen that the left object is clear while the right objects are blurred in source image A. Conversely, the left object looks unclear while the right object appears clear in source image B. An image of which the two objects are both clear should be generated by the fusion of the source images.

In order to show the improvement, the fusion results are subtracted from the source image B, which results in the difference images. Fig.5*a-d* show the results, the difference images and the enlarged zoom-in around the right object generated by the LPT, NSCT_1, NSCT_2 and the proposed methods. The LPT methods produce extra artefacts, namely the pseudo-Gibbs phenomena at the edge of the objects and the fan-shaped boundaries of the FLS images, which are clearly visible in the difference images. Compared with the LPT results, the two NSCT-based methods effectively suppress the pseudo-Gibbs phenomena, so NSCT is more suitable for image fusion than LPT. However, due to the inappropriate fusion rules, the intensity is severely uneven. In addition, there remains a considerable amount of gray values at the object position in Fig.5*b*, illustrating that the useful information of source image B cannot be effectively conveyed into the fused image. NSCT with the proposed rules (NSCT_2) in Fig.5*c* and the proposed method as a whole in Fig.5*d* produce better subjective results compared to other techniques. There is only a little gray information left in the right region, while most of the remaining information is located in the left region in Fig.5*c*, which is corresponding to the condition that the right area is clear and the left is blurred in source image B. Therefore, the proposed fusion rules are more suitable for FLS image fusion. To avoid information loss during the decomposition and the reconstruction, the proposed method addresses the problem by directly copying the pixels from the source images to the fine fused image. There is no information left in the right region in the difference image in Fig.5*d*. For the proposed method, the coarse decision map is shown in Fig.7*a* and the final map after morphological processing is shown in Fig.7*b*. As shown in Fig.7*b*, the fine map has continuous regions and smooth boundary, ensuring that the fused image is visually stable and seamless.

The objective assessment is given in Tables 1. It can be observed that the metrics of the proposed method are the best compared to other methods. For instance, for the fusion of 338th and 368th frames, our proposed algorithm produced an MI value of 4.2034 which is superior to 2.2530, 3.5978 and 3.6834 generated by the other methods. Similarly, the proposed algorithm produced better results in terms of $Q^{AB/F}$ and PSNR. Therefore, the proposed method can obtain more useful information from the source images, leading to better fusion results. The algorithms are coded in Matlab and implemented using an Intel i7 at 2.60GHz. For the fusion of two 750×500 images, it takes approximately 1 min, including about 59s to decompose and reconstruct images by NSCT, 0.93s to implement the proposed rules, and 0.50s to implement the morphological post-processing

respectively. In total, we tested a video of 1020 pairs of frames and the average of the objective evaluations are reported in Table 2. As shown in the table, our proposed

algorithm consistently achieved the best fusion results for all three metrics.

Fig.5 (Left to right) Fusion result, difference image, and zoom-in of the highlighted (by red rectangle) part of the difference image
a LPT method
b NSCT_1 method
c NSCT_2 method
d The proposed method

Table 1 Objective Evaluation of difference methods fusion for 338 and 368 frames

Metrics	LPT	NSCT_1	NSCT_2	Proposed
MI	2.2530	3.5978	3.6834	4.2034
$Q^{AB/F}$	0.3411	0.5578	0.5790	0.6048
PSNR	14.5150	13.0392	16.9002	19.1666

Table 2 Average objective Evaluations of different fusion methods for 1020 pairs of frames

Metrics	LPT	NSCT_1	NSCT_2	Proposed
MI	2.2925	3.3027	3.6964	4.2897
$Q^{AB/F}$	0.3439	0.5249	0.5366	0.6079
PSNR	14.5146	12.8314	16.6764	18.0912

4.2.2 FLS Image Fusion with Imperfect Registration:

Fig.5b shows two mis-registration FLS images, which are subject to a small rotation. The fusion results, the difference images and local enlarged zoom-in are shown in Fig.6, and the decision maps are shown in Fig.7b. It can be seen that the MST-based fusion methods fail to handle this mis-registration. Since the identical object in the source images is not perfectly aligned, the object projections in the two source images are selected repeatedly as the useful information and conveyed into the fused image simultaneously, leading to these artefacts. Conversely, the proposed method successfully avoids the artefacts which appear in the fused image. As shown in the coarse decision map in Fig.7b, there are breaks due to the misalignment in the white region. When morphological operations are applied, the discontinuities in these regions are removed, eliminating the artefacts caused by mis-alignment.

The work in [43, 44] has demonstrated that the projection matrix $\mathbf{H}$ between two frame points $\mathbf{p}$ and $\mathbf{p}'$ follows an affine transformation. The elements of $\mathbf{H}$ vary across the image depending on the range and the unknown elevation angles. For the registration of consecutive and near-consecutive frames, the projection can approximately be simplified to a rigid transformation [4, 5, 6, 10], which is composed of a rotation along the z axis and translations along the x and y axes in the plane, since the influence caused by elevation, and roll-and-yaw motion of sonar can be neglected due to the large scope of FLS images. Although the simplified rigid model can satisfy most of the conditions for the case of consecutive and near-consecutive frames registration, it may lead to mis-registrations of FLS frames in some regions, which are subject to local non-rigid transformations. For example in Fig.8a and b, the initial

frames show an elongated shadow and as the sonar gets closer the shadow becomes shorter. This leads to a local mis-registration of some regions, such as the shadows when using a rigid transformation. Fig.8c and d correspond to the decision map and the fusion result respectively, using the proposed method. It can be seen that the object and its shadow are both from Fig.8a, which avoids the artefacts introduced by the mis-registration appearing around the shadow.

Fig.6 (Left to right) Fusion result, difference image and zoom-in of the highlighted (by red rectangle) part of the fusion result
a LPT method;
b NSCT_1 method
c NSCT_2 method
d The proposed method

Fig.7 (Left to right) Coarse map and final map of the fusion with perfect and imperfect registration
a Coarse map and Final map of the fusion of Fig.5a
b Coarse map and Final map of the fusion of Fig.5b

Fig.8 Fusion result of a real scenario
a The shadow is long when the sonar is distant from the object
b The shadow becomes shorter as the sonar gets closer to the object
c The final map of the proposed method
d The fusion result of the proposed method

Finally, we provide the mosaic of multiple FLS images. The mosaic of an FLS video sequence is generated from its every 30th frames since they provide complementary information and they simultaneously enlarge the field of view. The mosaic is generated by fusing the frames incrementally one by one. For example, we use the 338th and the 368th frames to produce the first mosaic. Then the result of first mosaic and the 398th frame act as the source images to produce the next mosaic. In summary, the current mosaic is generated from the previous mosaic result and the subsequent 30th frame. Fig.9a and b are the mosaic results using the NSCT_1 and the proposed method. As shown in Fig.9a, the result of NSCT_1 is severely uneven, which influences the clarity of the objects and the smoothness of the mosaic. It can be demonstrated that the SML method is not suitable for FLS image fusion, since it cannot describe the properties of the useful information of the clear regions in FLS images. In contrast, Fig.9b is seamless and does not show any sharp transition, and the azimuth field of view is also widened. The objects and regions in the mosaic are all clear for visualization and can be readily used for further processing. In Figure 9b, the dark regions at the top and bottom of the mosaic cannot be eliminated since there is no obvious translation between frames. Another mosaic is produced to demonstrate that the proposed method is able to fuse the frames with translation. Fig.9c gives the mosaic result of 272 frames of a ship hull search on two different horizontal tracklines using the proposed method. It can be seen that the darkness in the bottom and top regions in each source frame has been effectively eliminated without leaving obvious seam. There are visible and dark un-illuminated regions in one frame and the visible regions are conveyed to the resulting mosaic by using the proposed method, which guarantees that the mosaic can avoid the dark regions. The metrics comparison is provided in **Table 3**, which shows that the proposed method achieves superior results compared to other methods. In addition, the proposed method is used to fuse the partly overlapping frames. In practical applications, the sonar is sometimes kept idle and the recorded frames are therefore totally overlapping. For these frames, their contrasts are almost the same while the noise is randomly located in each frame. In this case, we take the average of these totally overlapping frames as opposed to directly fusing the images one by one since taking the average can suppress the randomly located noise without reducing the contrast of the fused image. Then the averaged image can further be used in the fusion to produce a mosaic with the other partly overlapping frames.

Table 3 Average objective Evaluations of different fusion methods for 272 pairs of frames

Metrics	LPT	NSCT_1	NSCT_2	Proposed
MI	1.9750	2.6427	3.0867	4.4526
$Q^{AB/F}$	0.0927	0.2862	0.3287	0.4486
PSNR	11.8635	14.2003	15.0579	17.0950

a

b

c

Fig.9 *Mosaic of multiple FLS images*

a Mosaic of a video of bridge detection by NSCT_1 method

b Mosaic of a video of bridge detection by proposed method

c Mosaic result of multiple FLS images with y axis motion by proposed method

5. Conclusion

In this paper, a novel fusion method for FLS image mosaic is proposed. The overall algorithm involves coarse and fine fusion methods. The coarse fusion is based on NSCT, which can effectively suppress the artefacts introduced by the up and down sampling during the decomposition and reconstruction process. In addition, Gabor energy and local contrast have shown to be more suitable for the establishment of the fusion rules for low- and high-frequency sub-bands of FLS image fusion. Based on the coarse results, the importance of the source images can be measured by RMSE to generate a decision map, which determines the useful regions of the coarsely fused image and the source images. Finally, the useful regions are directly copied to the fused image. The proposed method produced superior fusion and mosaic results compared to existing methods.

6. Acknowledgments

This research is supported by National Natural Science Fund of China (No. 41376102 and 61633009), Fundamental Research Funds for the Central University of China (No. HEUCF140510), and Australian Research Council Research Grants (DP150104251 and DE1201029 60).

7. References

- [1] Sound Metrics Corporation, <http://www.soundmetrics.com/>, accessed on 16/06/2016.
- [2] Blueview Corporation, <http://www.blueview.com/>, accessed on 16/06/2016
- [3] Burt, P.J., Adelson, E.H., 1983. A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics*, 2(4), 217-236.
- [4] Hurtos, N., Ribas, D., Cufi, X., Petillot, Y., Salvi, J., 2014. Fourier-based Registration for Robust Forward-looking Sonar Mosaicing in Low-visibility Underwater Environments. *Journal of Field Robotics*, 32(1), 123-151.
- [5] Hurtos, N., Cufi, X., Petillot, Y., Salvi, J., 2012. Fourier-based registrations for two-dimensional forward-looking sonar image mosaicing. 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, 5298-5305.
- [6] Hurtos, N., Nagappa, S., Palomeras, N., Salvi, J., 2014. Real-time mosaicing with two-dimensional forward-looking sonar. *IEEE International Conference on Robotics and Automation*, 601-606.
- [7] Negahdaripour, S., Aykin, M.D., Sinnarajah, S., 2011. Dynamic scene analysis and mosaicing of benthic habitats by FS sonar imaging – Issues and complexities. *MTS/IEEE OCEANS 2011, KONA*, 1-7.
- [8] Kim, K., Neretti, N., Intrator, N., 2005. Mosaicing of acoustic camera images. *IEE Proceedings – Radar, Sonar and Navigation*, 152 (4), 263-270.
- [9] Li, H.Y., Dong, Y., He, X.D., Xie, S.R., 2014. A sonar image mosaicing algorithm based on improved SIFT for USV. 2014 IEEE International Conference on Mechatronics and Automation, 1839-1843.
- [10] Hurtos, N., Cufi, X., Salvi, J., 2013. A Novel Blending Technique for Two-Dimensional Forward-Looking Sonar Mosaicing. 2013 OCEANS, San Diego, 1-7.
- [11] Chai, Y., Li, H.F., Zhang, X.Y., 2012. Multifocus image fusion based on features contrast of multiscale products in nonsampled contourlet transform domain. *Optik - International Journal for Light and Electron Optics*, 123(7), 569-581.
- [12] Burt, P.J., Adelson, E.H., 1983. The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on Communications*, 31(4), 532-540.
- [13] Petrovic, V.S., Xydeas, C.S., 2004. Gradient-based Multiresolution Image Fusion. *IEEE Transactions on Image Processing*, 13(2), 228-237.
- [14] Yang, Y., Huang, S.Y., 2014. Multi-focus Image Fusion Using an Effective Discrete Wavelet Transform Based Algorithm. *Measurement Science Review*, 14(2), 102-108.
- [15] Kanan, K., Perumal, S.A., Arulmozhi, K., 2010. Area level fusion of Multi-focused Images using Multi-Stationary Wavelet Packet Transform. *International Journal of Computer Applications*, 2(1), 88-95.
- [16] Zhou, Z.Q., Li, S., Wang, B., 2014. Multi-scale weighted gradient-based fusion for multi-focus images. *Information Fusion*, 20, 60-72.
- [17] Do, M.N., Vetterli, M., 2005. The contourlet transform: An efficient directional multiresolution image representation. *IEEE Transactions on Image Processing*, 14(12), 2091-2106.
- [18] Cunha, A.L., Zhou, J., Do, M.N., 2006. The Nonsampled Contourlet Transform: Theory, Design, and Applications. *IEEE Transactions on Image Processing*, 15(10), 3089-3101.
- [19] Yang, Y., Tong, S., Huang, S.Y., Lin, P., 2015. Multifocus Image Fusion Based on NSCT and Focused Area Detection. *IEEE Sensors Journal*, 15(5), 2824-2838.
- [20] Li, S., Yang, B., 2010. Hybrid multiresolution method for multisensor multimodal image fusion. *IEEE Sensors Journal*, 10(9), 1519-1526.
- [21] Aslantas, V., Kurban, R., 2010. Fusion of multi-focus images using differential evolution algorithm. *Expert Systems with Applications*, 37(12), 8861-8870.
- [22] De, I., Chanda, B., 2013. Multi-focus image fusion using a morphology-based focus measure in a quad-tree structure. *Information Fusion*, 14(2), 136-146.
- [23] Ayin, M.D., Negahdaripour, S., 2012. On feature extraction and region matching for forward scan sonar imaging. 2012 Oceans, 1-9.
- [24] Zhang, Q., Guo, B.L., 2009. Multifocus image fusion using the nonsampled contourlet transform. *Signal Processing*, 89(7), 1334-1346.
- [25] Kim, K., Intrator, N., Neretti, N., 2004. Image registration and mosaicing of noisy acoustic camera. 2004 11th IEEE International Conference on Electronics, Circuit and Systems, 527-530.
- [26] Wang, Q., Ward, R.K., 2007. Fast Image/Video Contrast Enhancement Based on Weighted Threshold Histogram Equalization. *IEEE Transactions on Consumer Electronics*, 53(2), 757-764.
- [27] Sound Metrics Corporation, <http://www.soundmetrics.com/products/aris-sonars/Accessories/X2-Rotator>, accessed on 16/06/2016
- [28] Reed, S., Ruiz I. T., Capus C. and Petillot Y., 2006. The fusion of large scale classified side-scan sonar image

- mosaics. *IEEE Transactions on Image Processing*, 15(7), 2049-2060.
- [29] Zhang, J., Sohel, F., Bian, H.Y., Bennamoun, M., An, S.J., 2016. Forward-Looking Sonar Image Registration Using Polar Transform. *MTS/IEEE Oceans 2016*, Monterey, 1-6.
- [30] El-Baz, A., Farag, A., Gimelfarb, G., Experiments on Robust Image Registration Using a Markov-Gibbs Appearance Model. *Structural, Syntactic, and Statistical Pattern Recognition*, Vol. 4019 of the series *Lectures notes in Computer Science*, 65-73.
- [31] Huang, W., Jing, Z., 2007. Evaluation of focus measures in multi-focus image fusion. *Pattern Recognition Letters*, 28(4), 493-500.
- [32] He, C., Zheng, Y.F., Ahalt, S.C., 2002. Object Tracking Using the Gabor Wavelet Transform and the Golden Section Algorithm. *IEEE Transactions on Multimedia*, 4(4), 528-538.
- [33] Chen, J.B., Gong, Z.B., Li, H.Y., Xie, S.R., 2011. A detection method based on sonar image for underwater pipeline tracker. *2011 IEEE Second International Conference on Mechanic Automation and Control Engineering*, 3766-3769.
- [34] Gabor, D., 1946. Theory of communication. Part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, 93(26), 429-441.
- [35] Gao, G.R., Xu, L.P., Feng, D.Z., 2013. Multi-focus image fusion based on non-subsampled shearlet transform. *IET Image Processing*, 7(6), 633-639.
- [36] Toet, A., Ruyven van, J.J., Valetton, J.M., 1989. Merging thermal and visual images by a contrast pyramid. *Optical Engineering*, 28(7), 789-792.
- [37] De Valois, R.L., Yund, E.W., Hepler, N., 1982. The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research*, 22(5), 531-544.
- [38] Chai, Y., Li, H., Li, Z. F., 2011. Multifocus image fusion scheme using focused region detection and multiresolution. *Optical Communication*, 284(19), 4376-4389.
- [39] Yang, B., Li, S., 2012. Pixel-level image fusion with simultaneous orthogonal matching pursuit. *Information Fusion*, 13(1), 10-19.
- [40] Xydeas, C.S., Petrovic, V., 2000. Objective image fusion performance measure. *Electron Letters*, 36(4), 308-309.
- [41] Miao, Q.G., Shi, C., Xu, P.F., Yang, M., Shi, Y.B., 2011. A novel algorithm of image fusion using shearlets. *Optical Communication*, 284(6), 1540-1547.
- [42] Chai, Y., Li, H.F., Zhang, X.Y., 2012. Multifocus image fusion based on features contrast of multiscale products in nonsubsampling contourlet transform domain. *Optik - International Journal for Light and Electron Optics*, 123(7), 569-581.
- [43] Negahdaripour, S., Sekkati, H., 2007. 3-D Motion estimation for positioning from 2-D acoustic video imagery. In *Proceedings of 3rd Iberian Conference on Pattern Recognition. Image Analysis*, 80-88.
- [44] Negahdaripour, S., 2012. Visual motion ambiguities of a plane in 2-D FS sonar motion sequences. *Computer Vision and Image Understanding*, 116(6), 754-764.