A Forward-Looking Sonar Progressive Coding Using Morphological Skeleton Representation

Danilo Avola[1], Marco Bernardi[2,3], Luigi Cinque[2], Gian Luca Foresti[1], Marco Raoul Marini[2], Chiara Petrioli[2,3]
[1]Department of Mathematics, Computer Science and Physics, University of Udine, Udine, Italy
{danilo.avola, gianluca.foresti}@uniud.it
[2]Department of Computer Science, Sapienza University, Rome, Italy
{bernardi, cinque, marini, petrioli}@di.uniroma1.it
[3]WSENSE s.r.l., Rome, Italy

***Abstract*—The Forward-Looking Sonar (FLS) is one of the most effective devices for underwater exploration. It provides high resolution images that can be used for several applications in marine research, oceanographic, and deep-sea exploration. However, due to underwater acoustic channels with limited bandwidth, FLS image compression is still an open problem. This paper presents a method for the progressive coding of underwater images acquired by FLS. The proposed method combines K-means clustering technique with Morphological Skeleton (MS). The MS approximates the shape information to provide the minimum amount of data to be transmitted. Moreover, K-means clustering is also used to reduce the number of distinct colours. Experimental results on real data acquired by FLS show that the proposed technique outperforms popular compression methods.**

***Index Terms*—forward-looking sonar, image compression, shape coding, k-means clustering, morphological skeleton**

I. Introduction

Several real applications in underwater context are carried out in turbid water or in highly-cluttered environments [12]. In these scenarios, Forward-Looking Sonar (FLS) offers the opportunity to perform exploration tasks regardless of the visibility conditions. In Fig. 1, an example of FLS image is shown. The automated FLS image processing is a complex problem due to the presence of interactions among visual cues and artifacts. In this setting, the transmission of FLS images assumes a great importance. In particular, one of the most critical issues is related to the limited bandwidth of the current acoustic communication channels. In the present literature, the discrete wavelet transform (DWT) [4] is often used to compress the sonar data. The problem of this approach is that it tends to smooth out the sharp edges in the images [21]. Another popular solution is based on the use of the Compressed Sensing (CS) [7], but it usually does not reach high results in terms of visual quality.

Recently, Haghighat et al. [9] showed that the segmentation of regions with a homogeneous intensity value (i.e., foreground regions) from background can be very useful for coding sonar images. These regions offer relevant shape information about the presence of various objects. Therefore, in this context, shape coding techniques can be considered one of the most suitable solutions to meet bandwidth constraints and the preservation of main information about foreground regions.

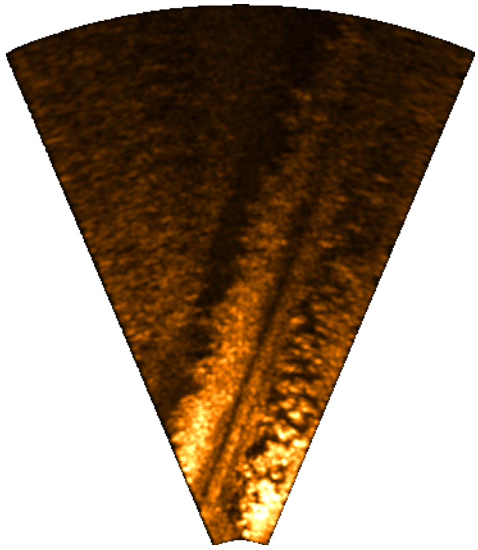


Fig. 1: An example of forward-looking sonar image.

In the last decades, shape coding has received a considerable interest from the scientific community [18], [20], [22], [23]. In particular, a very engaging approach, i.e., the Morphological Skeleton (MS) [11], [17], has shown remarkable results in different application contexts. The MS of binary images is calculated by iterating a set of morphological operations, i.e., dilation and erosion. By using this technique notable results can be also achieved in complex scenarios. In [8], Foresti et al. presented a representation method for coding of lossy binary images in surveillance applications based on an approximation of the statistical MS. The skeleton techniques are also used successfully in several medical applications [17]. For example, Chen et al. [6] used the skeleton reconstruction for the stenosis detection. Instead, Wang et al. [20] used the skeleton extraction for vascular models. Another interesting work is proposed by Xu et al. [22], where a generalized MS to combine the concepts of internal and external maximal disks into a unified framework is presented.

In underwater context, most of the traditional coding approaches are unable to detect the relevant information to be

ISBN 978-90-827970-1-5 © EURASIP 2018

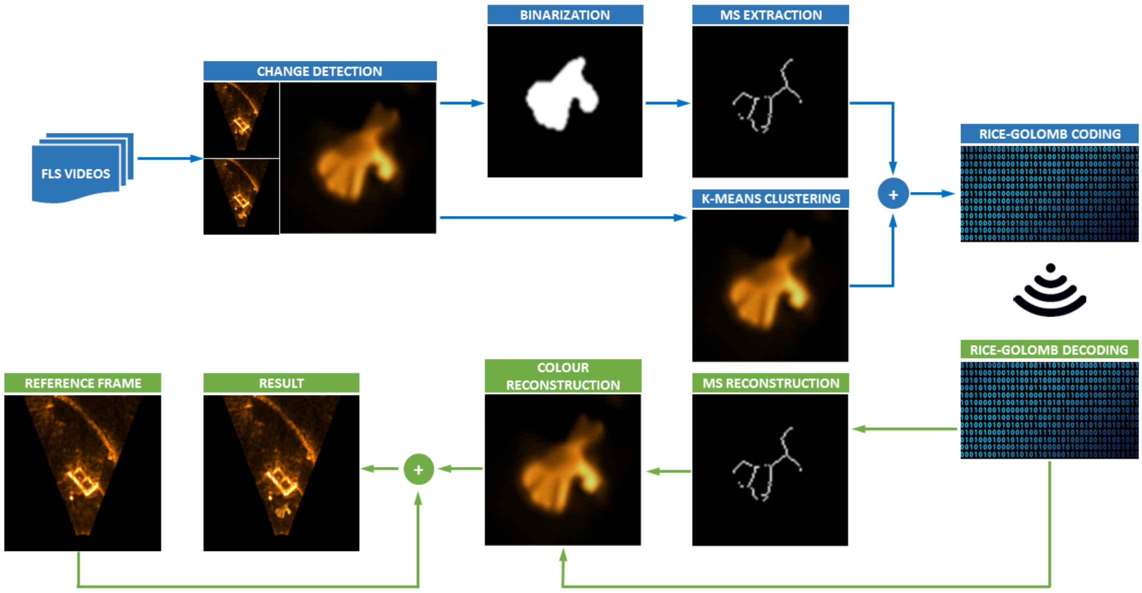


Fig. 2: Logical architecture of the proposed method.

transmitted. Inspired by the work proposed in [8], in this paper a novel method to encode foreground objects in FLS images is proposed. The aim of this work is to define a shape coding technique for encoding only the relevant information provided by the FLS, thus reducing the amount of significant data to be transmitted. The proposed method consists of three main steps. First, from each FLS image the MS is extracted. Then, on the same images, a K-means clustering technique to reduce the number of distinct colours is applied [5]. Finally, this computed information is encoded by the Rice-Golomb (RG) coding algorithm [19]. The reported experimental results, carried out on several real FLS videos, have shown that the proposed technique, applied to FLS images, outperforms popular compression methods in the current state-of-the-art.

The main contributions of the present work can be summarized as follows:

1) The application of a skeleton shape description and a K-means colour quantization on FLS data;
2) The definition of a light approach, i.e., based on simple algorithms, for coding and decoding of the FLS data;
3) Unlike several approaches in the current literature, applied on FLS data, the definition of a strategy able to detect the relevant information to transmit.

The rest of the paper is structured as follows. Section II provides an overview of the logical architecture. Section III illustrates the proposed method. Section IV reports the experimental results. Finally, section V concludes the paper.

II. The Logical Architecture

Fig. 2 shows the logical architecture of the proposed FLS coding algorithm. The proposed method receives as input a FLS video sequence, whose frames are processed by a change detection (CD) module [2], [3]. The first frame is used as reference background and it is transmitted by the JPEG2000 lossless coding algorithm [15]. In particular, the CD module receives as input two gray-level FLS images and provides as output a binary image X, obtained by thresholding the intensity difference between the two images. The threshold is tuned according to different aspects, including sonar type, underwater environment, and illumination type. The aim of the CD step is to identify the different patterns, or blobs, (i.e., areas with significant changes) corresponding to possible interesting objects. For each area, two further elaborations are performed: MS extraction and colour quantization. The first is obtained by the iteration of morphological operators, i.e., erosion and dilation, on each binary blob, which is thus progressively shrunk. The second is got by using a K-means clustering technique to perform the colour quantization process. In the end, the information about the MS and colour quantization is coded by the RG algorithm. The decoding module performs the inverse operation with respect to that just illustrated. In particular, it receives the first frame and uses it as background reference. Then, it receives and reconstructs the shape information. Finally, it adds the colour data to the reconstructed image.

III. The Forward-Looking Sonar Coding Method

In this section, the proposed FLS coding method is described in detail. The pseudocode in Algorithm 1 is reported to summarize the main execution steps.

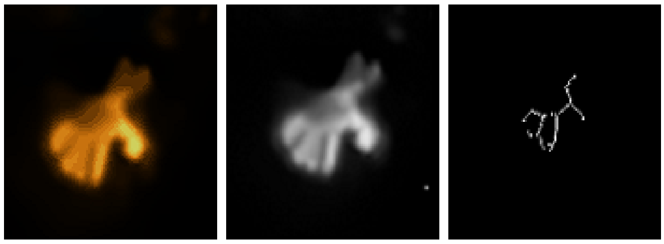
A. Change Detection

The CD module receives as input two FLS images, $I_{k-1}(x, y)$ and $I_k(x, y)$, acquired at the time instants $k - 1$ and k, respectively. The two images allow the method to detect areas corresponding to moving objects. In a first step, the intensity difference $D_k(x, y)$ between the two images is computed. Later, a threshold is applied to the $D_k(x, y)$ image

ISBN 978-90-827970-1-5 © EURASIP 2018

Algorithm 1: Pseudocode of the proposed method.

Input:
FLS video sequence: $I = \{I_0, .., I_T\}$
Clusters for colour quantization: v
Threshold for Change Detection: δ
Accumulator sets $eros1$,$eros2$ and $open$
Send I_0 with JPEG2000 lossless coding
for $k = 1$ *to* T **do**
  Compute change detection D_k between I_{k-1} and I_k
  Establish where each point $(x, y) \in D_k$ is a background point or a moving object point and obtain the binary mask X by using threshold δ
  Extract from X a collection M of n minimum bounding boxes
  for $m \in M$ **do**
    Compute morphological skeleton $S_k(X)$
    B = square $3x3$ structural element
    $eros1$ = m
    Number of iteration of MS $iter = 0$
    while *eros2 is empty* **do**
      $eros2$ = erode($eros1, B$)
      $open$ = dilate($eros2, B$)
      $Sk_{iter}(m) = eros1 - open$
      $eros1 = eros2$
      $iter = iter + 1$
    end
    $N = iter$
    $Sk_N(m) = eros1$
    Compute index table idx and colour palette
    $C = \{C_1, C_2, ..., C_v\}$ with K-means clustering
    $code$ =Rice-Golomb($Sk(X), idx, C$)
    Transmit $code$
  end
end



(a) Image (b) Mask (c) Skeleton

Fig. 3: Example of morphological skeleton extraction: (a) shows the original image, (b) shows the mask extracted by the CD module, and (c) shows the related skeleton.

to establish the points $(x, y) \in D_k(x, y)$ that belong to the foreground. If $D_k(x, y) > \delta$, for a fixed (x, y), then that point is marked as object, otherwise it is marked as background. Like reported previously, the threshold δ depends on several aspects (e.g., illumination changes) and it is empirically computed. At the end of the CD process, a binary image X is obtained, where sets of 1 are foreground elements, while sets of 0 are the background. Afterwards, from the image X, a collection of M Minimum Bounding Boxes (MBBs) is extracted. Each $m \in M$ represents an image area where significant changes are occurred [1]–[3], [8].

B. Morphological Skeleton Extraction

For each image area $m \in M$, the extraction of the MS, defined by $sk(m)$, is performed. In general, the MS can be described as follows:

$$sk(m) = \{[x, y, n(x, y)] : (x, y) \in m, n(x, y) \in [1, .., N]\} \quad (1)$$

where, the function $n(x, y)$ associates at each skeleton point (x, y) the iteration in which it was detected. Instead, N represents the maximum number of iterations. Notice that, the MS preserves the structure of the shape, but, at the same time, it removes all redundant pixels. As a running example, Fig. 3 shows the processing on a real FLS image. More specifically, Fig. 3a represents an acquired human hand, Fig. 3b is the binary pattern (in this case, m), and Fig. 3c is the result obtained by the skeleton extraction. The MS is computed by applying iteratively the morphological operators as detailed in [8], [13].

C. Colour Quantization

Colour quantization process is a crucial step for obtaining limited-colour images with a high quality level for each blob. The proposed method adopts the K-means clustering technique described in [5]. It is one of the most widely used methods for colour quantization. This algorithm is designed to classify all pixels, within an image, in a fixed number v of colours. The value of v must be chosen to reduce the colour space and, at the same time, guarantee that the reconstructed image is qualitatively close to the original one. In Fig.4 the steps of the colour quantization are shown. Summarizing, given a set $Z = \{z_1, \ldots, z_N\} \in \mathbb{R}^D$ (where D represents the intensity values of the images), the objective of the K-means clustering is to split Z in v clusters $C = \{c_1, \ldots, c_v\}$, where $\cup_{i=1}^{v} c_i = Z$ and $c_i \cap c_j = \emptyset$ with $i \neq j$, by minimization of the Sum of the Squared Error (SSE) as follows:

$$SSE = \sum_{j=1}^{v} \sum_{z_i \in C_j} \|z_i - c_j\|^2 \quad (2)$$

The K-means clustering was chosen for several advantages, including easy implementation and linear complexity.

D. Rice-Golomb Coding

The RG coding is a lossless compression algorithm used in JPEG-LS [16]. In general, given a constant m, any symbol c can be represented as a quotient q and a remainder r, where:

$$c = qm + r \quad (3)$$

To encode a character c, first $q = \lfloor \frac{c}{m} \rfloor$ is computed. Then, a q-length string of 1 and a r-length string of 0 bits are written out. At the end, the last m bits of r are also write out. The decode module works in the same way. First, it determines q by counting the number of 1 before the first 0. Then, it calculates r reading the next m bits as a binary value and, finally, it decodes the character c with the formulation expressed in Equation 3. If c is small (relatively to m) then also q is small. The RG algorithm was used because, due to the large number of small values, it reduces the average length per encoded value compared to fixed-width encoding.

IV. Experimental Results

In this section, the experimental results are reported. Due to the lack of public datasets for FLS images, we have used, for the final tests, a small set of images that are however

ISBN 978-90-827970-1-5 © EURASIP 2018

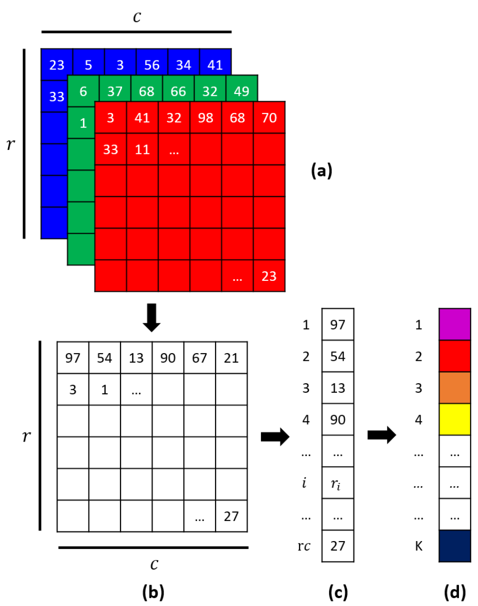


Fig. 4: Colour quantization: (a) composition of a quantized-colour image with r rows and c columns, (b) two-dimensional image pixel matrix with colour indexes, (c) one-dimensional data structure of the limited-colour image, and (d) the colours palette with the K corresponding indexes.

freely available. In particular, we have used real FLS sample data obtainable by the Teledyne Marine ProViewer[1]. In Table I, we summarize the frame rate and the video size used for the experiments. To speed up the entire process, in the

TABLE I: FLS videos used for testing the proposed method.

Name	Frames Rate (fps)	Size	Duration (s)
Airplane	3	1137×474	29
Diver	1	1137×474	199
Hand & Cinderblock	7	1137×474	19

pre-processing stage, all frames were resized by pixel area relation to 568×237. To evaluate the performance of the proposed method, we used two qualitative metrics [10]: Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). Furthermore, to show the effectiveness of the proposed approach, the bit per pixel information (bpp) is also provided. In the experiments, the threshold δ and the value of k have been empirically set to 1 and 32, respectively. We compared the proposed method with JPEG [24], progressive JPEG [14], and SPHIT [4]. To highlight the goodness of the MS representation, the just introduced techniques were used on the image areas extracted from the CD step. The Table II shows the results of the comparison and the performance of the proposed approach on the FLS videos.

The proposed method obtains remarkable results on each video. The best results are presented for the Airplane video, where the method achieves comparable results in term of bpp

[1]The videos are available on: http://www.teledynemarine.com/ProViewer.

TABLE II: Results of the proposed method (average values on the test FLS videos).

Method	Metric	Airplane	Diver	Hand & Cinderblock
JPEG	PSNR	30.1 dB	31.1 dB	29.8 dB
	SSIM	83%	83%	80%
	bpp	0.2	0.1	0.1
Progressive JPEG	PSNR	30.1 dB	31.1 dB	34.7 dB
	SSIM	83%	83%	80%
	bpp	0.18	0.06	0.08
SPHIT	PSNR	30.1 dB	32.3 dB	34.7 dB
	SSIM	84%	89%	86%
	bpp	0.2	0.1	0.05
Proposed	PSNR	30.8 dB	37.3 dB	38.8 dB
	SSIM	90%	93%	96%
	bpp	0.2	0.09	0.1

and outperforms other methods on PSNR and SSIM metrics. Also in the Diver video our approach presents a significant result with a comparable performance in bpp with respect to the progressive JPEG. In Hand & Cinderblock video, the proposed method obtains the better result. In terms of PSNR, the best result is of 34.7 dB, while in our method the metric value is of 38.8 dB. As previously reported, the output quality of the developed approach is completely comparable with different key works of the current literature (at the same bit rate). Anyway, we know that we are working in a niche context with specific images. However, this context can be considered very promising in the exploration of the underwater environments. Fig. 5 shows a visual comparison among the selected algorithms for a bit rate lower than 1 bpp. The JPEG images present blocking artifacts and are perceptibly worse than competitive techniques. The approach reported in this paper and the SPHIT algorithm produce comparable output, but SPHIT sacrifices some details in favour of the visual appearance. Unlike the other reported approaches, our method can preserve all the image details. Notice that, due to the noise present in the sonar image, some white dots can appear in the image reconstruction. Nevertheless, the quality of the reconstructed image is better than compared methods. Finally, thanks to the effectiveness of MS, the proposed method can process all videos in real-time.

V. Conclusion

This paper presents a novel compression method for forward-looking sonar data in underwater environments. The proposed approach presents different contributions, including shape description and colour quantization. Compared with popular algorithms, the proposed method, on underwater FLS data, outperforms the current state-of-the-art on the basis of well-known qualitative metrics. Both code and encode processes are not time-consuming, thus allowing their implementation in different hardware solutions. The latter is a crucial aspect of the deep-sea exploration.

Acknowledgment

This work was supported in part by the MIUR under grant “Dipartimenti di Eccellenza 2018-2022” of the Department of Computer Science of Sapienza University and by the European Union within the Project ARCHEOSUb “Autonomous

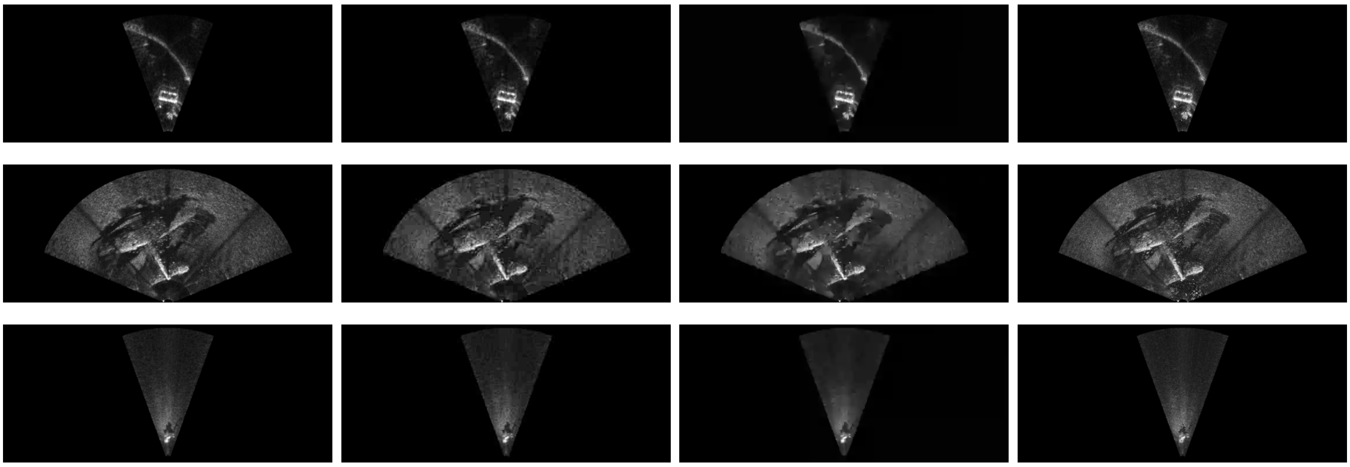


Fig. 5: Compression result for test images. The first column denotes the original images. The second, third, and fourth columns denote the images obtained by the progressive JPEG, SPHIT, and proposed algorithm, respectively.

underwater Robotic and sensing systems for Cultural Heritage discovery Conservation and in situ valorization".

REFERENCES

[1] D. Avola, M. Bernardi, L. Cinque, G. L. Foresti, and C. Massaroni. Adaptive bootstrapping management by keypoint clustering for background initialization. *Pattern Recognition Letters*, 100(C):110–116, 2017.

[2] D. Avola, L. Cinque, G. L. Foresti, C. Massaroni, and D. Pannone. A keypoint-based method for background modeling and foreground detection using a ptz camera. *Pattern Recognition Letters*, 96:96–105, 2017.

[3] D. Avola, G. L. Foresti, N. Martinel, C. Micheloni, D. Pannone, and C. Piciarelli. Aerial video surveillance system for small-scale uav environment monitoring. In *Proceedings of the 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6, 2017.

[4] L. Bin and M. Qinggang. An improved spiht wavelet transform in the underwater acoustic image compression. In *Proceedings of the 2nd International Conference on Measurement, Information and Control (ICMIC)*, pages 1315–1318, 2013.

[5] M. E. Celebi. Improving the performance of k-means for color quantization. *Image and Vision Computing*, 29(4):260–271, 2011.

[6] S. Y. J. Chen, J. D. Carroll, and J. C. Messenger. Quantitative analysis of reconstructed 3-d coronary arterial tree and intracoronary devices. *IEEE Transactions on Medical Imaging*, 21(7):724–740, 2002.

[7] W. Chen, F. Yuan, and E. Cheng. Adaptive underwater image compression with high robust based on compressed sensing. In *Proceedings of the IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pages 1–6, 2016.

[8] G. L. Foresti and C. S. Regazzoni. Statistical morphological skeleton for representing and coding noisy shapes. *IEE Proceedings - Vision, Image and Signal Processing*, 146(2):85–92, 1999.

[9] M. Haghighat, Xiuying Li, Zicheng Fang, Yang Zhang, and S. Negahdaripour. Segmentation, classification and modeling of two-dimensional forward-scan sonar imagery for efficient coding and synthesis. In *OCEANS MTS/IEEE Monterey*, pages 1–8, 2016.

[10] A. Hore and D. Ziou. Image quality metrics: Psnr vs. ssim. In *Proceedings of the 20th International Conference on Pattern Recognition (ICPR)*, pages 2366–2369, 2010.

[11] C Lantuejoui. Skeletonization in quantitative metallography. *Issues of Digital Image Processing*, 1980.

[12] Y. Li, H. Lu, J. Li, X. Li, Y. Li, and S. Serikawa. Underwater image de-scattering and classification by deep neural network. *Computers & Electrical Engineering*, 54(C):68–77, 2016.

[13] P. Maragos and R. Schafer. Morphological skeleton representation and coding of binary images. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(5):1228–1244, 1986.

[14] M. Mody, V. Paladiya, and K. Ahuja. Efficient progressive jpeg decoder using jpeg baseline hardware. In *Proceedings of the 2nd IEEE International Conference on Image Information Processing (ICIIP)*, pages 369–372, 2013.

[15] M. Rabbani and R. Joshi. An overview of the jpeg 2000 still image compression standard. *Signal processing: Image communication*, 17(1):3–48, 2002.

[16] S. D. Rane and G. Sapiro. Evaluation of jpeg-ls, the new lossless and controlled-lossy still image compression standard, for compression of high-resolution elevation data. *IEEE Transactions on Geoscience and Remote Sensing*, 39(10):2298–2306, 2001.

[17] P. K. Saha, G. Borgefors, and G. S. di Baja. A survey on skeletonization algorithms and their applications. *Pattern Recognition Letters*, 76:3–12, 2016.

[18] Z. Shen, M. R. Frater, and J. F. Arnold. Quad-tree block-based binary shape coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(6):845–850, 2008.

[19] A. Vasilache. Order adaptive golomb rice coding for high variability sources. In *Proceedings of the 25th European Signal Processing Conference (EUSIPCO)*, pages 1789–1793, 2017.

[20] S. Wang, J. Wu, M. Wei, and X. Ma. Robust curve skeleton extraction for vascular structures. *Graphical Models*, 74(4):109–120, 2012.

[21] L. S. Wong, G. E. Allen, and B. L. Evans. Sonar data compression using non-uniform quantization and noise shaping. In *Proceedings of the 48th Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pages 1895–1899, 2014.

[22] J. Xu. A generalized morphological skeleton transform using both internal and external skeleton points. *Pattern Recognition*, 47(8):2607–2620, 2014.

[23] J. W. Yoon. An efficient geometric shape coding and representation approach to obtain both skeleton and contours. In *Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, pages 1–8, 2016.

[24] X. Zhang, S. Wang, K. Gu, W. Lin, S. Ma, and W. Gao. Just-noticeable difference-based perceptual optimization for jpeg compression. *IEEE Signal Processing Letters*, 24(1):96–100, 2017.