remote sensing

MDPI

Article

A Statistical Evaluation of the Connection between Underwater Optical and Acoustic Images

Rebeca Chinicz [1] and Roee Diamant [1,2,*]

[1] Hatter Department of Marine Technologies, University of Haifa, Haifa 3103301, Israel; rchinicz@campus.haifa.ac.il
[2] Faculty of Electrical Engineering and Computing, University of Zagreb, 10000 Zagreb, Croatia
* Correspondence: roee.d@univ.haifa.ac.il

Abstract: The use of Synthetic Aperture Sonar (SAS) in autonomous underwater vehicle (AUV) surveys has found applications in archaeological searches, underwater mine detection and wildlife monitoring. However, the easy confusability of natural objects with the target object leads to high false positive rates. To improve detection, the combination of SAS and optical images has recently attracted attention. While SAS data provides a large-scale survey, optical information can help contextualize it. This combination creates the need to match multimodal, optical–acoustic image pairs. The two images are not aligned, and are taken from different angles of view and at different times. As a result, challenges such as the different resolution, scaling and posture of the two sensors need to be overcome. In this research, motivated by the information gain when using both modalities, we turn to statistical exploration for feature analysis to investigate the relationship between the two modalities. In particular, we propose an entropic method for recognizing matching multimodal images of the same object and investigate the probabilistic dependency between the images of the two modalities based on their conditional probabilities. The results on a real dataset of SAS and optical images of the same and different objects on the seafloor confirm our assumption that the conditional probability of SAS images is different from the marginal probability given an optical image, and show a favorable trade-off between detection and false alarm rate that is higher than current benchmarks. For reproducibility, we share our database.

Keywords: SAS object detection; SAS–optical multimodal; feature descriptors; entropy metrics; image matching

Citation: Chinicz, R.; Diamant, R. A Statistical Evaluation of the Connection between Underwater Optical and Acoustic Images. *Remote Sens.* **2024**, *16*, 689. https://doi.org/10.3390/rs16040689

Academic Editors: Jingchun Zhou, Wenqi Ren, Qiuping Jiang and Yan-Tsung Peng

Received: 7 January 2024
Revised: 8 February 2024
Accepted: 13 February 2024
Published: 15 February 2024

1. Introduction

1.1. Overview

Autonomous underwater vehicles (AUVs) are efficient tools for surveying the seabed for target detection, e.g., for identifying mines, exploring shipwrecks and monitoring marine fauna. In the past, sonar was the preferred sensor for these investigations. However, due to their higher resolution, optical cameras provide better information for target classification. The combination of the two modalities can therefore offer an advantage. For example, multimodal sonar and optical sensors have proven effective in underwater navigation [1], object classification [2] and coastal characterization [3]. However, this combination poses a challenge for the comparison of sonar and optical images.

The matching of objects between sonar and optical data can be seen as a kind of multimodal sensing, where the main challenge is to overcome the different physical properties of the two modalities when viewing the same object. Sonar images are generated by the reflection of acoustic signals from the object and its surroundings and contain highlights and shadow regions [4] for the direct reflections from the object and the area blocked by the object, respectively. The sonar image also contains reflections from the area around the object (the background) and has a relatively low pixel resolution. The sonar image provides

details about the height of the object (by shadow measurement), its material composition (by spectral analysis) and the general shape and size of the object. However, due to the low resolution (around 2 cm per pixel for high-precision Synthetic Aperture Sonar (SAS) systems [2]), no fine details of the object are captured. The optical image, on the other hand, provides millimeter resolution at short distances of a few meters, as well as texture and color data, but requires the observer to be close to the object and its quality is affected by the turbidity of the water. Differences between SAS and optical information can also be observed in the distribution of pixel intensity. For example, sonar reflections can be modeled as Beta- or K-distribution [5], while the pixels in optical images are considered as a Gaussian or Rayleigh distribution [6,7]. This is because the SAS image is created by beamforming and matched filtering of the received reflections, while optical images convert the light reflections into RGB values. Given this challenge, the combination of optical and SAS images is associated with object matching.

Object matching refers to the characterization of a complete mapping process between images acquired by different sensors or at different times. The available methods can be divided into two main approaches: area-based or feature-based. In the area-based approach, a template is matched to the different images using a similarity metric, such as cumulative cross entropy or residual entropy [8]. In feature-based approaches, relevant features are extracted from both images and matched, such as the Scale-Invariant Feature Transform (SIFT) algorithm [9] and its inheritors. In this method, the image is transformed into a collection of features that are fully scale-invariant, as well as translationally and rotationally invariant. In this transformation, a scale space is constructed and Gaussian differences are computed to form common feature descriptors for both images. Both the area-based and feature-based methods have found applications in areas such as optical and Synthetic Aperture Radar (SAR) registration [10] and matching between different optical, SAR and Light Detection and Ranging (LiDAR) images [11], with the primary goal of object detection and classification. However, recent works [2] showed that the above approaches fail when matching optical and sonar images due to the different pixel distribution and structure. Therefore, feature extraction is still an essential part of most object matching methods.

1.2. State of the Art

In the case considered in this paper, full registration is not required to align SAS and optical images. Instead, we focus on image matching: the aim is to check whether the same object is represented in both the SAS and the optical modality. In image matching, the features extracted from the object help to recognize the object, since it is assumed that the basic shape of the object is preserved in both modalities. One of the most influential feature descriptors for optical images is the Scale Invariant Feature Transform (SIFT) [9]. However, matching optical and SAS images using SIFT is more challenging than matching images of the same modality or even specifically the multimodal case of matching optical and SAR images, since the appearance of the objects is different in the two modalities. An example of this can be found in Figure 1, where the structural similarities between optical and SAR images are more apparent than in a pair of optical and SAS images of an object on the seafloor.

To overcome the challenges of matching between optical and SAS images, ref. [12] proposes the use of Convolutional Neural Networks (CNNs) to extract content and style features from sonar and optical images, respectively, and attempts to build a conversion scheme from sonar and optical images of a submerged panel of numbers and letters in an experimental pool. However, this approach is limited to a specific use case of character conversion between the two modalities, and requires a large amount of data. In an oceanic environment, it is more difficult to obtain optical and especially SAS images of different objects, especially image pairs, so training CNNs to match such images is more challenging. Recently, geometric features have been proven to be efficient to describe both SAS and optical images [2]. Notable geometric features for both optical and SAS modalities include contour-based features, such as the solidity of an object, eccentricity and normalized

central moments, and region-based features, such as roughness and compactness [13]. With the absence of a large database, feature-based matching is a common approach for multimodal combination.

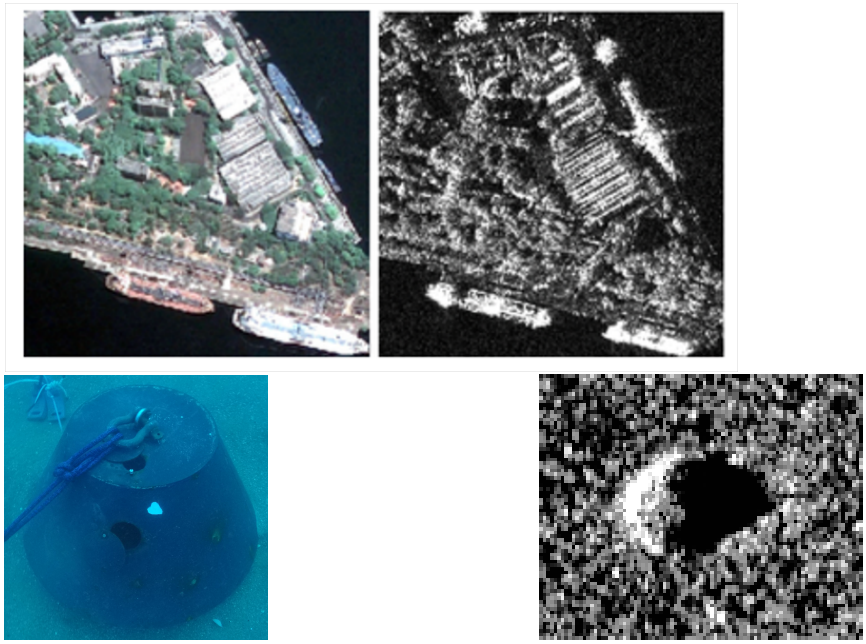


Figure 1. Optical and SAR images of the same structure, from [11] (**top panels**), and, for comparison, optical and SAS images of an underwater object (**bottom panels**). Similarities can be observed between the land images, whereas the marine-based images appear very different.

Recent work has focused on improving navigation methods by fusing visual and acoustic information. A fusion of acoustic DVL measurements and stereo imaging is offered in [14] for Simultaneous Localization And Mapping (SLAM) by formulating a joint utility function that also incorporates observations from a gyroscope and a depth sensor. In [15], a visual odometry system is proposes for underwater unmanned vehicles. The method filters out unreliable points from the 3D reconstruction of a scene provided by stereo cameras augmented by depth information from a single-beam sonar. Another multimodal SLAM is considered in [16], where monocular images are merged with images from a forward-looking sonar. The features from both modalities are extracted separately and matched in a maximum likelihood estimation model. However, this requires that the images from both sensors are acquired simultaneously and from the same vehicle, otherwise the matching will fail.

The aim of this work is to quantify the likelihood that an object identified in an SAS image is the same object identified in an optical image. We investigate the use of state-of-the-art feature descriptors, compute their distribution, and evaluate the statistical relationship between the descriptors of the two modalities in terms of a concept we call *mutual entropy*, which provides insights into the potential benefits of combining SAS and optical images. In other words, mutual entropy quantifies whether the information embedded in the optical image can be used to characterize the object in the SAS image and vice versa. This study provides metrics for the comparison of optical and SAS images, as well as statistical evidence that the multimodal combination of SAS and optical images is beneficial.

In the following, we refer to matched SAS–optical image pairs, i.e., pairs of images of the same object, as *positive* samples, and unmatched image pairs as *negative* samples. To test the performance of our entropy-based combination metric, we use a dataset of 1217 pairs of SAS and optical images we have collected in several sea experiments with our A18 ECA-robotics AUV, which contains a two-sided Kraken-based SAS and high-resolution optical cameras. To create a dataset of SAS–optical pairs, we compared the objects recognized in SAS images and optical images. The SAS images and the optical images were not acquired

at the same time, position and orientation. The reason for this is the different observation range of SAS and optical: SAS requires a distance of tens of meters above the seafloor, while optical images require a small distance of the camera to the object (about 5 m). Therefore, we matched our SAS and optical observations by: (1) location of the object and (2) manual recognition of the object. The dataset contains 185 positive samples and 1032 negative samples. Our results show that entropy metrics, which quantify the information embedded in SAS and optical images by some feature descriptors, are effective in detecting SAS–optical similarities and that there is indeed a statistical relationship between the two modalities. In particular, the results show a favorable trade-off between the detection and false alarm rates, which is much better than the chance level and represents an improvement compared to benchmark methods. To ensure reproducibility, we present our database of matched and unmatched SAS and optical files in: https://drive.google.com/file/d/1ggz8BA0W6CtRXH-48XTfy9eO23-Ag1Qt/view?usp=sharing, accessed on 14 February 2024.

Although we have made considerable efforts to collect the above database with several sea experiments involving AUV and scuba divers, its size is still small. For this reason, we refrain from offering a deep learning approach and focus on statistical analysis instead. Statistics-based approaches also benefit from more data, but are more intuitively understandable despite the size of the dataset and therefore help in this initial exploration of the relationship between optical and acoustic images. The main contributions of this work are:

1. A first demonstration of the statistical relationship between underwater optical and acoustic images.
2. A statistics-based method for validating optical and SAS image pairs as matching or not.
3. A shared database of manually reviewed and labeled underwater optical and SAS images in which objects have already been recognized and segmented.

The rest of the paper is structured as follows: In Section 2, we provide our system model with our main assumptions and preliminaries for feature descriptors and entropy measures. Section 3 presents the methodologies of our solution to quantify the similarities between SAS–optical pairs, and Section 4 describes the exploration of the statistical relations between SAS and optical images. Results are discussed in Section 5, and conclusions are drawn in Section 6.

2. System Model

Referring to the illustration in Figure 2, the scenario under consideration is an AUV searching an area for a target object with its SAS. As soon as a target of interest is detected, e.g., a wreck, gas sip or a submerged mine, the AUV approaches the detected target to obtain optical images for target verification. This process can involve (1) matching an object found in the optical image with the object in the SAS image for target verification, or (2) matching an object found in the optical image with an object found in the SAS image as a navigation aid. The two scenarios considered are illustrated in Figure 2, and the latter scenario is also explained in detail in [17]. As a central tool for both objectives, we focus in this paper on the comparison of an object from an optical image with an object from an SAS image. We clarify that our solution is not specifically tailored for, e.g., navigation or target tracking, but rather focus on the backbone of the task of determining if an object found in an SAS image and an object found in an optical image are the same.

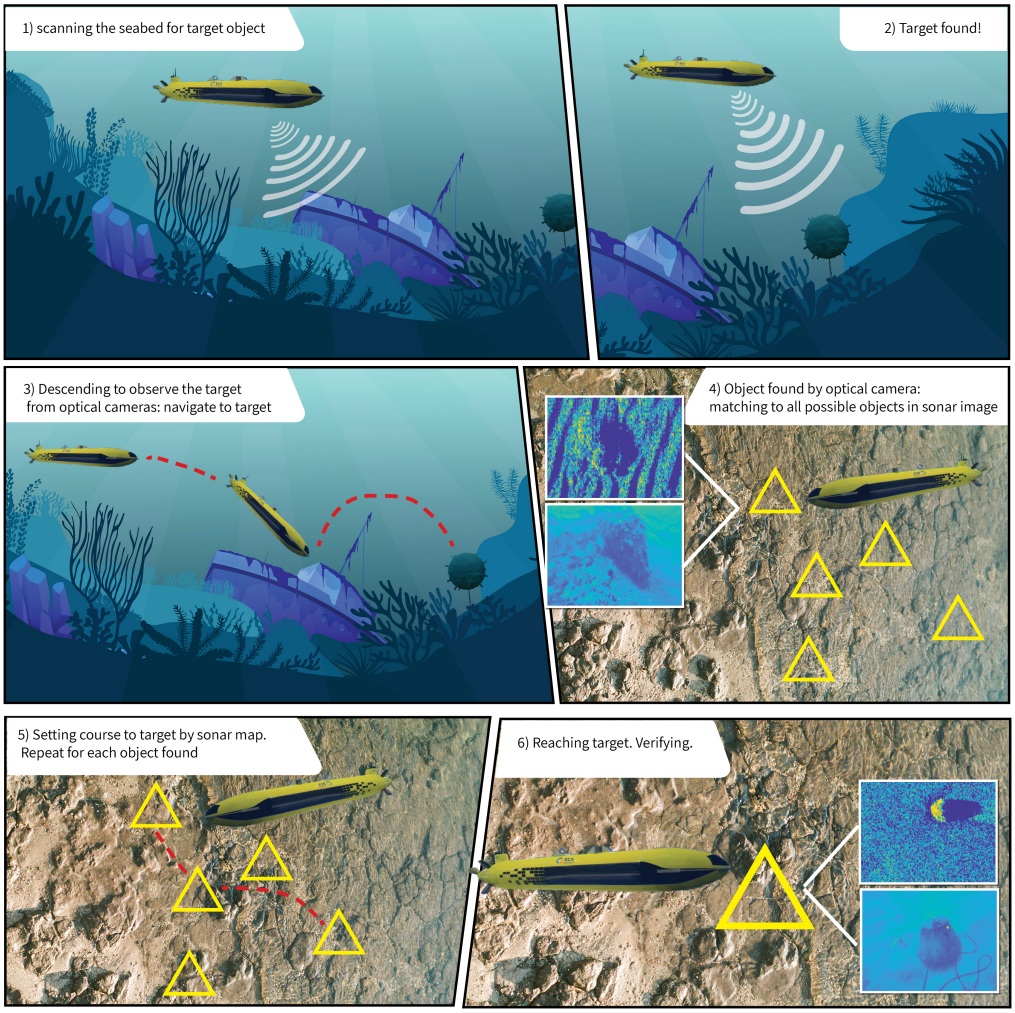
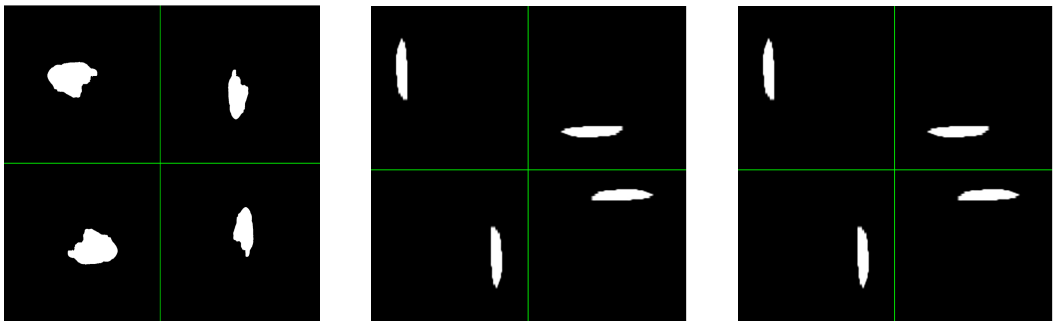


Figure 2. Demonstration of the considered scenario. Triangles represent location of targets.

2.1. Main Assumptions

We assume that several objects are found in the SAS image and that the SAS image acquisition takes place at a different time and at a different distance than the optical image acquisition. This means that the process of matching between the objects in the SAS image and an object found in an optical image should be insensitive to the viewing angle, size and shape of the object. Examples for the contour of such rotated images are given in Figure 3. Instead, such a matching procedure requires a statistical relationship between the objects in the images generated by the two modalities. Here, we assume that the relationship between SAS and optical images can be explored through the statistical distribution of the features of the objects. The preprocessing steps required for our solution are listed below.

We assume an underlying process that identifies a single object in both the SAS and optical images. This detection leads to regions of interest (ROIs), which are assumed to contain a signal object. These can be obtained by segmentation, which separates potential targets from their surrounding background. While several detectors are offered for SAS and for optical images, we choose the segmentation scheme presented in [4] for SAS and the algorithms in [18] for optical images. The former performs fuzzy clustering, the latter involves the use of Markov Random Fields (MRFs). We believe that these methods are well suited for our multimodal matching problem, as they provide robust results at a relatively low computational cost. See Section 3 below for more details. Another set of tools we use are feature descriptors and entropy matrices. Some preliminary remarks on these tools follow.



(**a**) (**b**) (**c**)

Figure 3. Augmentations performed in optical, SAS highlight and SAS shadow contour, in order to obtain the feature distributions. The contour in each figure belongs to the same object, a cylinder. Each figure above is a square grid with the original contour, and three rotations: 90, 180 and 270 degrees. (**a**) Contour of rotated optical images. (**b**) Contour of SAS-highlight images. (**c**) Contour of SAS-shadow images.

2.2. Preliminaries

2.2.1. Feature Descriptors

Our SAS–optical matching scheme relies on comparing features of the two objects. Given the two very different modalities physics, we avoid shape-based feature descriptors such as SIFT. Instead, we rely on geometry-based features of local curves to describe the objects within the images from the two modalities. A local curve refers to adjacent points on the contour, and is represented by a polynomial function of the range and angle of the curve, relative to the centroid of the contour. However, considering that the SAS- and optical-based objects are acquired at different times, distances and orientations, we use only angle-independent features; ones that avoid matching the two objects by the angle-of-view. In addition, region- and contour-based features are extracted, including the perimeter of the contour (P_c), its solidity, eccentricity, compactness and circularity, the FFT-based low-frequency density and the skewness of the discrete Fourier transform (DFT), as well as the central moments and the newly proposed entropy-angle feature [2]. Table 1 summarizes the definitions of the individual extracted features. The relationship between the features can be quantified using entropy metrics.

Table 1. Summary of features extracted.

Name	Definition	Observations
Ecc	$\text{Eccentricity} = \frac{l_{\text{major}}}{l_{\text{minor}}}$	l_{major} and l_{minor} are equal to the two eigenvalues of the co-variance matrix [13]
Perimeter	Perimeter of the object's contour	
Comp	$\text{Compactness} = \frac{P_c}{A}$	A is the area of the contour and P_c is the perimeter
cir	$\text{Circularity} = \frac{A}{A_c}$	A_c is the area of the circle having the same length as the object's perimeter
Solidity	$\text{Solidity} = \frac{A}{A_{\text{convex hull}}}$	$A_{\text{convex hull}}$ is the area of the convex hull [13]

Table 1. *Cont.*

Name	Definition	Observations
roughness	$\text{Roughness} = \frac{P_c}{P_{\text{convex hull}}}$	$P_{\text{convex hull}}$ is the perimeter of the convex hull [13]
sigma$_{ij}$ and l_k	$\bar{\zeta}_{ij} = \frac{\zeta_{ij}}{\zeta_{00}^{l}}$, $l = 1 + \frac{i+j}{2}$, $\zeta_{ij} = \sum_u \sum_v (u-\bar{u})^i (v-\bar{v})^j I(u,v)$	Normalized central moment is $\bar{\zeta}_{ij}$, for every pixel (u,v), center mass $(\bar{u},\bar{v})$, $I(u,v) \in [0,1]$ is 1 if (u,v) is within the object's region, and each $sigma_{ij}$ refers to a central moment ζ_{ij} and l to the normalized version of lth calculated central moment [2]
Low Freq Den	$LowFrequencyDensity = \frac{1}{N_{\text{LF}}} \sum_{n_{\text{DFT}}=1}^{N_{\text{LF}}} D_{\text{cen}}(n_{\text{DFT}})$	D_{cen} is the magnitude of the Fourier coefficients of the centroid distance function and the DFT is implemented with N_{DFT} points
DFT-skewness	$\frac{\frac{1}{N_{\text{DFT}}} \sum_{n_{\text{DFT}}=1}^{N_{\text{DFT}}} (D_{\text{cen}}(n_{\text{DFT}}) - \mu_{\text{cen}})^3}{(\frac{1}{N_{\text{DFT}}} \sum_{n_{\text{DFT}}=1}^{N_{\text{DFT}}} (D_{\text{cen}}(n_{\text{DFT}}) - \mu_{\text{cen}})^2)^{1.5}}$	$\mu_{\text{cen}} = \frac{1}{N_{\text{DFT}}} \cdot \sum_{n_{\text{DFT}}=1}^{N_{\text{DFT}}} D_{\text{cen}}(n_{\text{DFT}})$
localCur$_i$	Local curve, polynomial coefficient i [2]	
Hist-BAS	Entropy-Angle Feature (EAF) [2]	

2.2.2. Entropy Measures

The Shannon entropy (H) for a dataset X is defined as:

$$H(X) = -\sum_{x \in X} p(x) \log p(x) \tag{1}$$

A direct calculation of this metric may be too computationally expensive and difficult to derive in its exact form. Instead, several alternatives are explored.

The Mutual Information provides knowledge about the reduction of uncertainty in one variable, given observations from another. For two datasets x, y, it is computed by:

$$I(X;Y) = -\sum_{x \in X, y \in Y} \log\left(\frac{p(x,y)}{p(y)\,p(x)}\right) p(x,y)\ . \tag{2}$$

The Differential Entropy provides insight on the expected uncertainty in one variable, given the values of another. It is calculated by:

$$h(X;Y) = -\sum_{x \in X, y \in Y} \log\left(\frac{p(x,y)}{p(y)}\right) p(x,y)\ . \tag{3}$$

The Transfer Entropy and the Normalized Transfer Entropy are, respectively, defined by:

$$T_{X \to Y} = H(Y_t \mid Y_{t-1:t-L}) - H(Y_t \mid Y_{t-1:t-L}, X_{t-1:t-L}) \tag{4}$$

$$\frac{T_{X \to Y}}{H(Y_t \mid Y_{t-1:t-L})}\ . \tag{5}$$

and are calculated by the compressed transfer entropy (cTE) [19]. The two measures are useful to determine dependencies between time series, as a form of mutual information

that also takes into account the transition probabilities between the ordered elements of the data sets. Vector entropy is used to describe the complexity of a signal and takes into account non-linear behavior. Formally define a constant γ and denote Γ as the Gamma function. The method for calculating the vector entropy is presented in Algorithm 1.

Algorithm 1 Vector Entropy KL

Input: feature vectors from the SAS and optical images, X **and** Y

$z \leftarrow X$ concatenated with Y
$n \leftarrow$number of rows in z ▷ the size of both X and Y
$p \leftarrow$number of columns in z
$D \leftarrow$distances between the $K = 4$ nearest neighbors between X and Y based on a KNN search
$n \leftarrow$ size of D
Output $\leftarrow \frac{p*\text{mean}(\log D)+\log(\frac{\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2}+1)})+\gamma+\log(n-1)}{\log 2}$

3. Methodology

3.1. Key Idea

The input to our scheme are two regions of interest (ROIs) containing possibly the same object; one acquired by SAS and the other by an optical camera. The scheme augment the two ROIs by rotation and scaling to calculate the statistics of extracted features of the two ROIs. Comparison of this statistics via entropy measures then determines if the two objects are indeed the same. A block diagram illustrating our process is shown in Figure 4. We distinguish between the path of detection (red) and the path of information exploration (blue). In the first step, we perform ROI detection and ROI segmentation. Next, in the detection route, each segmented image is expanded to l ROIs by rotating and rescaling the image. We then extract key features for each of the augmented ROIs to form a dataset with M descriptors for each extracted feature. This dataset is used to evaluate the distribution of the feature for each ROI. This step eliminates the need to learn the general statistics of each feature and contributes to the robustness of the method. The feature distributions of the SAS and optical ROIs are then compared using an entropy metric. The result is then thresholded to obtain a binary detection decision for a *positive* match for ROIs containing the same object and vice versa for *negative*. To determine which feature entropy metric is most effective for our similarity comparison, we quantify the overlap between the distributions of positive and negative matches using K-Means Clustering [20] accuracy.

Our goal is to divide the original ROIs from SAS and optical images into positive (matched) and negative (mismatched). Formally, given an ROI from an SAS image S, and an ROI from an optical image O, we determine whether $P(S|O) = P(S,O)/P(O) \neq P(S)$, or whether $P(O|S) = P(S,O)/P(S) \neq P(O)$. The former states that O provides valuable information about S, while the latter states that S is related to O. In both cases, the conclusion is that S and O contain the same object. This is explored via the information exploration route (blue lines), where we define feature vectors in terms of matching and non-matching optical and SAS data points, which we directly compare via an entropy measure and their conditional and marginal probabilities.

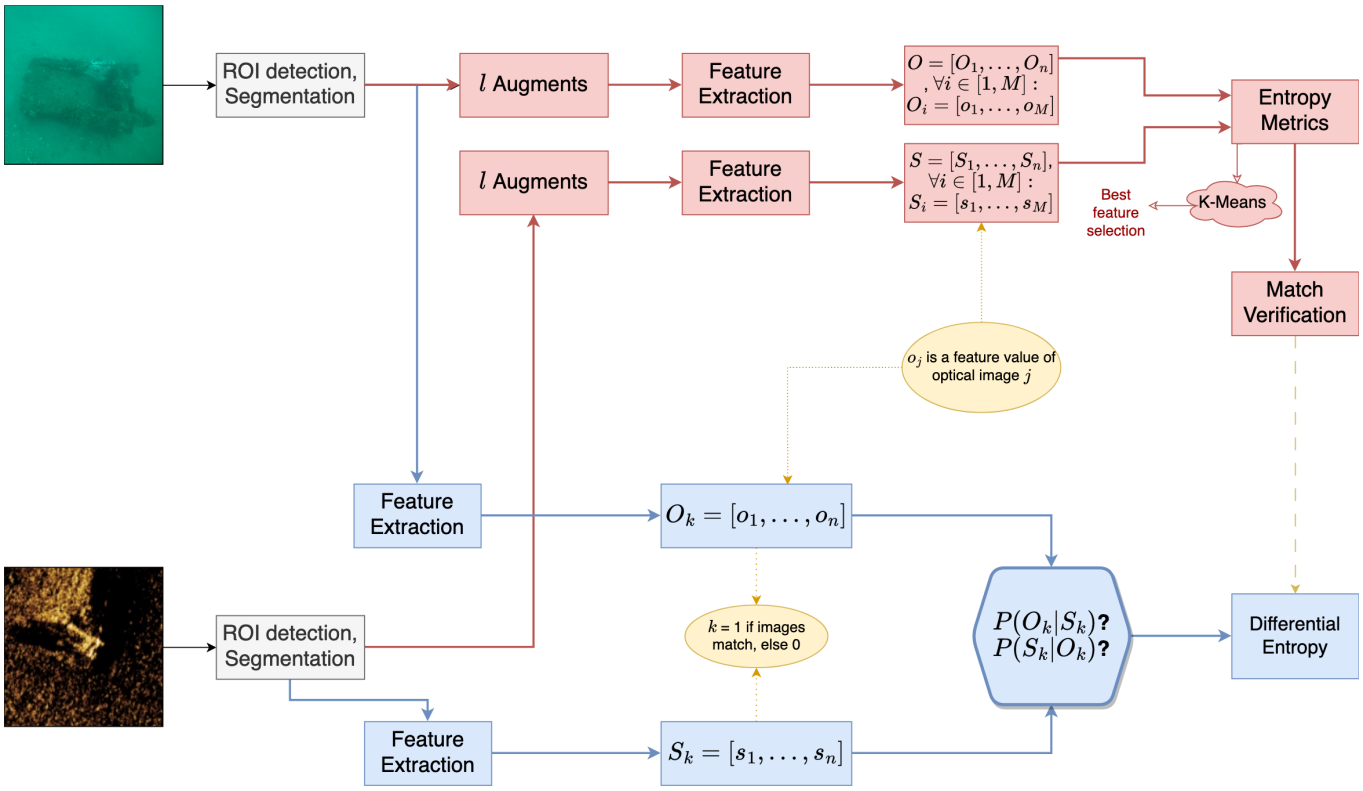


Figure 4. Diagram summarizing the steps of our proposed SAS–optical comparison approach. The detection process involves the steps on the top side of the diagram (red lines), logic steps (yellow lines) and the bottom side gives an overview of the steps taken to explore the statistical connection between the images (blue lines).

3.2. Entropy Matching Decision

Joint entropy measures can express the information exchange between two feature vectors. In this paper, we investigate the use of differential entropy, mutual information, transfer entropy and normalized transfer entropy, and vector entropy over the concatenation of two feature vectors. We chose these metrics because of their efficiency in identifying the exchange of information between statistical populations. The above entropy measures take as input pairs of features derived from the ROIs of the SAS and the optical images. The entropy is then calculated after the distribution of each feature set is computed. Note that since the entropy measures are not symmetric, we perform the analysis both from the SAS to the optical side and from the optical to the SAS direction. In addition, the analysis is performed separately for the shadow and highlight regions in the SAS ROI. We leave the investigation of a possible merging of the two regions to future work.

The entropy-based matching method is based on the distribution function of the features extracted from the SAS ROI and the optical ROI. Depending on the direction of exploration, from SAS to optical or vice versa, the two distributions are computed and labeled as $p(x)$ and $p(y)$. These are then fed into one of the entropy measures listed in Section 2.2.2, the result of which is called a soft decision. Crossing a threshold for this value would result in a positive or negative decision for each pair of SAS and optical ROIs. This process is applied separately for the shadow and highlights segments of the SAS ROIs and separately for the SAS to optical and optical to SAS directions. A positive match is determined if one of these tests leads to a positive decision. In the following sections, we provide details on how to obtain the distribution of the individual features.

3.3. Feature Extraction

The feature extraction process begins with a segmentation of the ROI to determine the position of the object's pixels within the ROI. For SAS segmentation, we choose the method in [4], which relies on kernel-based fuzzy clustering for segmentation. The process includes a denoising step that contributes to the segmentation process. The result is a three-grade segmentation, where the pixels of the ROI are divided into contiguous regions of background, highlights and shadows. For optical segmentation, we chose the method in [18], as it is a low complexity approach that incorporates spatial information to segment images via a Markov Random Fields (MRF)-based mixture model. From each segmented image, we extract a set of feature descriptors from the list described in Section 2.2 to express geometric information from both optical and SAS images.

The entropy detection process described above requires the calculation of the distribution of the feature. Since there is no probabilistic model for the feature descriptors, we turn to a numerical evaluation of the distribution. This is obtained by augmenting each ROI by rotation and scaling operations. Formally, for images defined as a two-dimensional matrix, these two operations are obtained by multiplying the image by an affine transformation matrix. For each pixel (x, y), we perform Rotation and Scaling.

3.3.1. Rotation

Considering that the SAS- and optical-based matched objects are most probably obtained at different (any) orientations, the angle-of-view does not contain valuable information that may affect performance. We thus use rotation to augment our dataset. This is performed by

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \tag{6}$$

where α as the specific angle of rotation.

3.3.2. Scaling

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \tag{7}$$

where s_x, s_y are the scaling factors in the x, y dimensions, respectively. Note that in this case, we use the same scale factor in both dimensions, that is $s_x = s_y$.

The augmentation process leads to a series of 89 augmentations per ROI. A single descriptor is drawn from each augmented ROI to obtain a set of 89 features for each ROI. This process is performed separately for both the SAS and optical ROIs. To demonstrate the effectiveness of this augmentation step, we show in Figure 5 an example feature set drawn from the ROI of an SAS and an optical image of a cylinder, as shown in Figures 6 and 7. We observe a variability in the feature values that reflects a change in the ROI due to the augmentation procedure. However, as the differences between the values show, a direct comparison between the feature distribution of the SAS and the optical ROIs is not possible for the task of matching the two, and a deeper statistical investigation through the concept of information exchange using joint entropy is required.

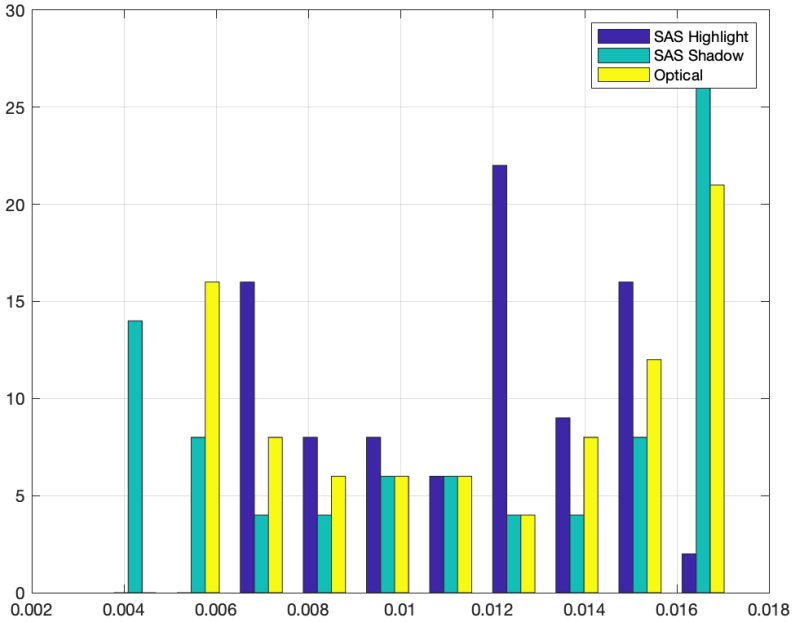


Figure 5. Distribution (histogram) of central moment with i = 0, j = 2 of SAS image of cylinder vs. optical image of cylinder.

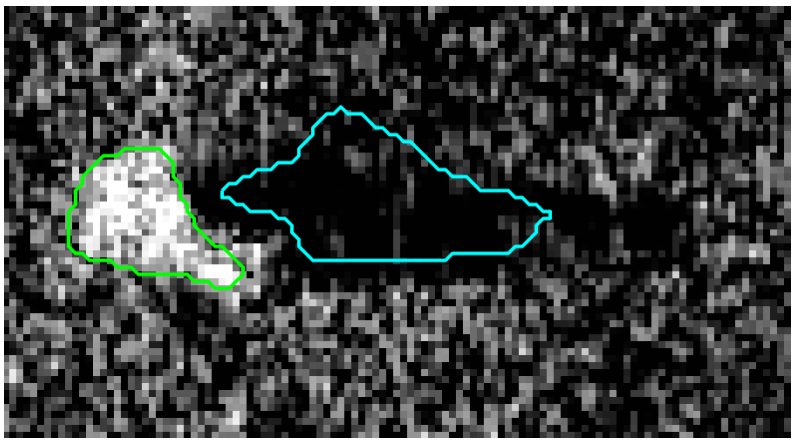


Figure 6. Original ROI of the SAS image of the cylinder segmented into highlight in green and shadow in blue.

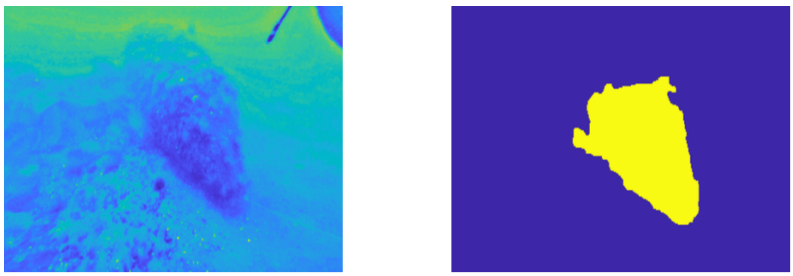


Figure 7. Original ROI of the optical image of the cylinder and its segmentation.

4. Exploration of Statistical Relations between SAS and Optical Images

We now turn to investigate whether there is indeed a relationship between optical and SAS images that can be used for matching, comparing or, even better, classifying a detected object. To this end, we are looking for a statistical relationship in the sense of an exchange of information between the optical and sonar images. The exploration of statistical relationships between data sets can be performed by calculating correlation and covariance measures, such as the Pearson correlation coefficient, under certain conditions,

such as linearity [21]. In this paper, we turn again to testing the relationship between SAS and optical images by using joint entropy measures to infer the statistical relationship between a set of features from an SAS ROI and a set of features from an optical ROI. To complete this analysis, we explore the direct calculation of the conditional probabilities $P(S|O)$ and $P(O|S)$.

Exploring the statistical relationships involves extracting the same feature from a set of ROIs derived from SAS and optical images. Some of these image pairs contain the same object and some different objects. We denote the positive (i.e., matched objects) dataset for each feature as $S_1 = [s_1, ..., s_m]$, $H_1 = [h_1, ..., h_m]$ and $O_1 = [o_1, ..., o_m]$, and the negative dataset as $S_0 = [s_1, ..., s_n]$, $H_0 = [h_1, ..., h_n]$ and $O_0 = [o_1, ..., o_n]$, where S, H and O represent the SAS-based ROI with the segmented shadow, the SAS-based ROI with the segmented highlight and the segmented ROI of the optical images, respectively. Each member in the S, H and O groups reflects the result of the selected feature over a particular ROI.

5. Results

Our goal is to compare a pair of optical and SAS images to determine whether they depict the same object. For this reason, we compare the performance of our approach with that of recent work [17], we refer to as *Abu-2023*. This benchmark scheme quantifies the likelihood that an SAS and an optical image contain the same object using a generic Gaussian radial basis function, which is known to improve robustness to image noise [22].

5.1. Dataset

Our dataset contains sonar images collected with our ECA robotics A18D AUV. The AUV is equipped with a Kraken-manufactured SAS sensor and a pair of optical cameras. The data were collected during four sea trials off the coasts of Caesarea and Haifa, Israel. The optical images of the objects found were obtained either from the AUV's cameras or from handheld optical cameras operated by divers. The ROIs were determined according to the scheme in [2] and checked manually.

The dataset contains images of a cylinder, a manta mine, a box and a variety of unclassified natural objects (rocks, corals, etc.). The natural objects are not labeled in either the SAS images or the optical images, and are therefore not matched. In addition, there were only optical images of the box. Therefore, the positive (matched) sonar-optical pairs were only formed with combinations of images of the cylinder and the manta objects, and the rest were only used to generate more negative (unmatched) samples. Figure 8 shows examples of optical and SAS images of the objects. With a total of 1217 samples—185 positive samples formed by pairing multiple images of the same objects (cylinder and manta mine) and 1032 negative samples—Table 2 quantifies the set of images in each modality per object type.

Table 2. Overview of images in the dataset.

Object Type	SAS	Optical
Cylinder	9	9
Manta Mine	8	13
Box	0	6
Natural	25	14

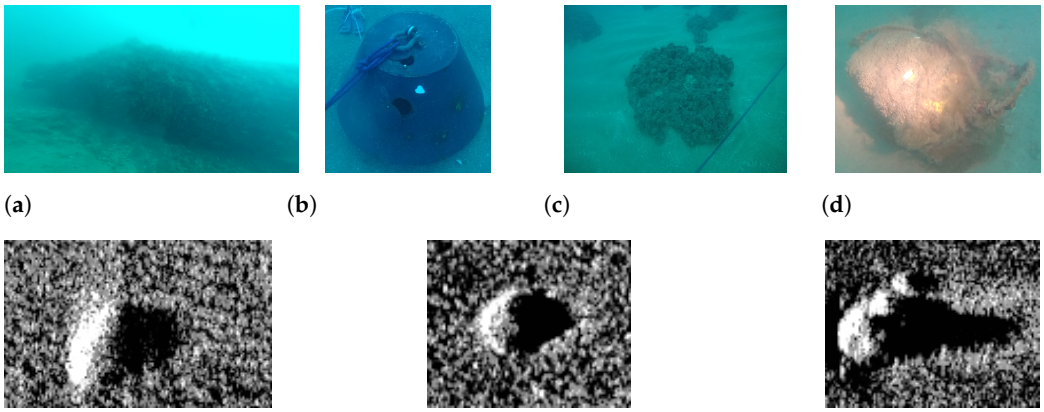


Figure 8. Examples of images from each type of object in the dataset. (**a**) Optical image of cylinder. (**b**) Optical image of manta mine. (**c**) Optical image of natural object. (**d**) Optical image of box. (**e**) SAS image of cylinder. (**f**) SAS image of manta mine. (**g**) SAS image of natural object.

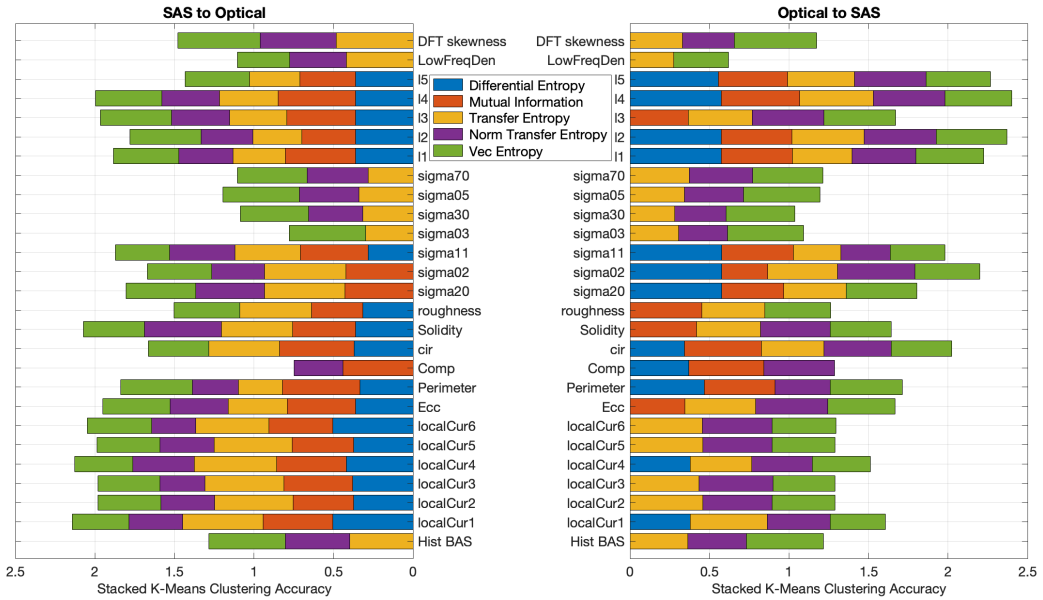
5.2. Feature Analysis

To quantify how well each feature's entropy measure separates between a positive and a negative match, we use the K-Means Clustering algorithm. It aims to divide a number of observations into K clusters. In our case, $K = 2$, and we assume that the feature whose entropy measures are most correctly clustered should be used for identifying matches. This choice of $K = 2$ is due to the fact that here we cluster the entropy values comparing optical–SAS image pairs into either "matching" or "non-matching", which is a binary task. For the SAS optical matching case, the accuracy of the K-Means clustering is used to indicate for each direction (from SAS to optical and from optical to SAS), for each entropy measure, and for each selected feature descriptor whether the positive and negative cases are well observed and distinguished from each other.

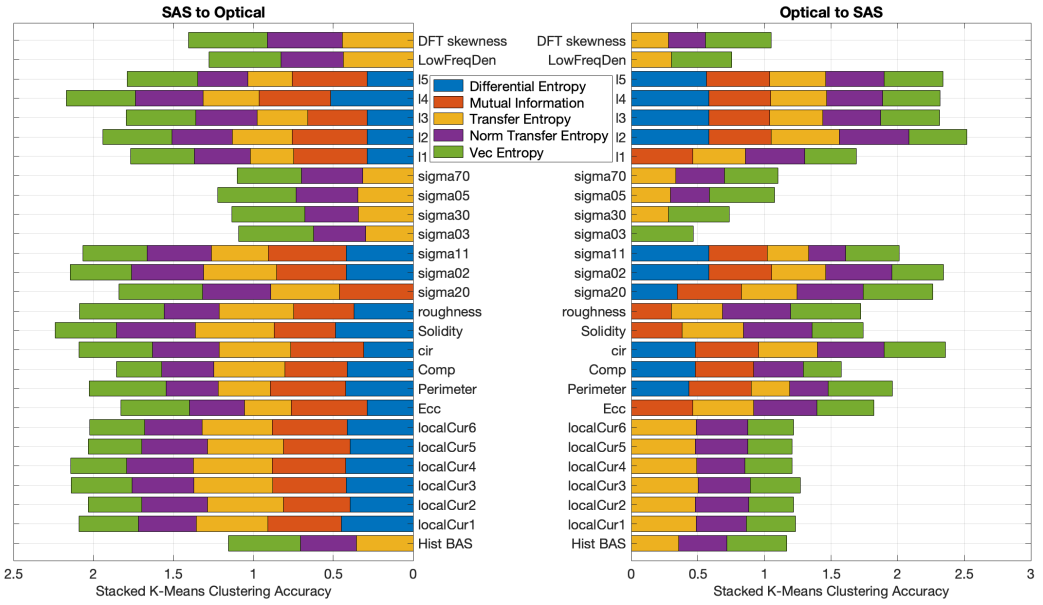
To determine the best feature for the similarity test, we show the accuracies of the K-Means clusterings over all tested features and entropy measures for both positive and negative matches in Figure 9. To account for the known distribution of positive and negative samples in the dataset, we set a threshold, λ, on the ratio between the sizes of the two clusters. The threshold takes into account the extreme case where, for example, all observations would be clustered in the negative cluster and we would obtain a high clustering accuracy. Nevertheless, the detection rate of matches would be low due to the imbalance between the number of positive and negative samples. This threshold was empirically set to 0.2, and if the criterion was not met, an accuracy of 0 was assigned for better visual discrimination. Table 3 summarizes which feature had the best accuracy, and was therefore selected for the matching procedure, for each entropy measure and direction, i.e., optical to SAS and SAS to optical, for SAS highlights (HL) and shadows (SH).

Table 3. Highest K-Means accuracy feature per entropy and direction of comparison. See the feature preliminaries in Section 2.2 for each feature's definition.

Direction / Entropy	Optical to SAS HL	SAS HL to Optical	Optical to SAS SH	SAS SH to Optical
Differential Entropy	sigma20	localCur6	l2	l4
Mutual Information	l4	l4	sigma20	sigma11
Transfer Entropy	localCur1	localCur4	l2	localCur4
Normalized Transfer Entropy	sigma02	Solidity	l2	Solidity
Vector Entropy	DFT skewness	DFT skewness	roughness	roughness



(**a**)



(**b**)

Figure 9. Stacked K-Means clustering accuracies of each entropy type, in different colors (X-axis), as tested with each each of the 27 features (Y-axis). The accuracy here refers to the accuracy of the resulting clusters at splitting the positive and negative image pairs (their entropy comparison values). (**a**) K-Means accuracies for entropy measures by feature with SAS highlight only. (**b**) K-Means accuracies for entropy measures by feature with SAS shadow only.

5.3. Entropy Matching

For each selected feature, we now examine the results of the entropy matching procedure. The results are measured using the receiver operating characteristics (ROC) curve to observe the trade-off between the detection rate and the false alarm rate. Detection is defined here as the correct assignment of a positive match, while a false alarm refers to the case of a positive decision on a non-similar object pair. The results are shown in Figure 10, and are compared to the work of [17] as a benchmark. We report that the highest clustering accuracy was achieved with differential entropy. We compute the ROC for this entropy

measure in all possible directions, i.e., from optical to SAS HL and to SAS SH and from SAS HL and SAS SH to optical.

As can be seen from Figure 10, the detection was not successful when the entropy measures were examined from SAS to optical, while in the other direction, the differential entropy provided a match far above the chance level. This was the case for both the highlights and the shadow regions, indicating that both the highlights and the shadows in SAS may provide relevant shape information that is statistically related to the shape information extracted from the optical images.

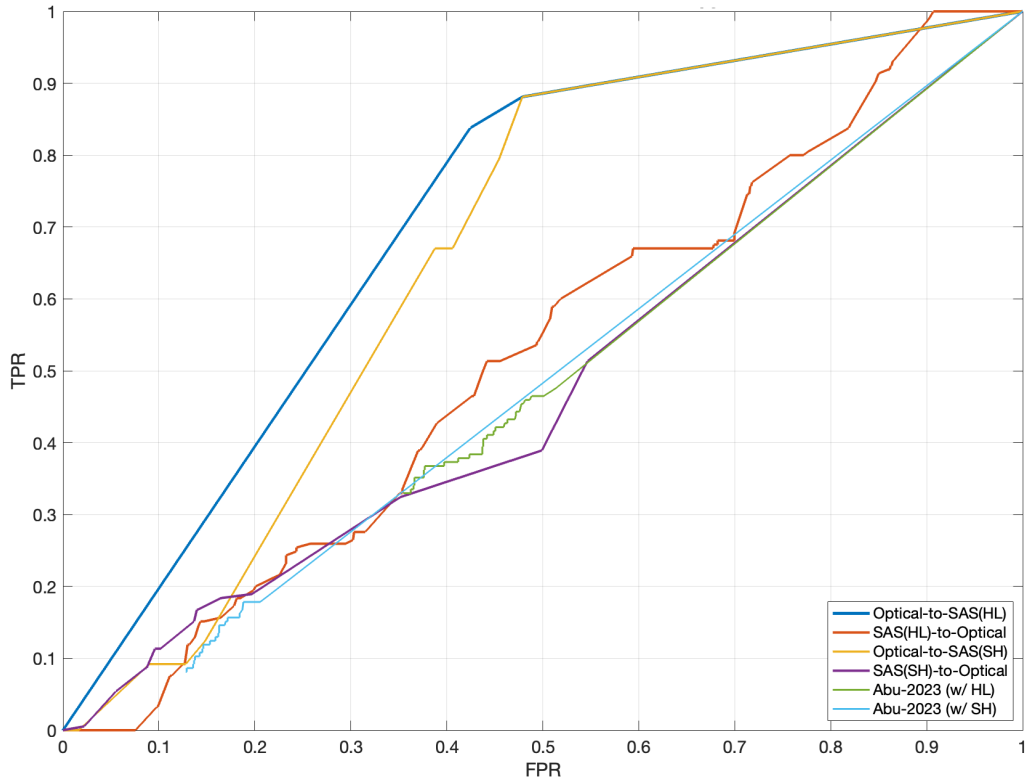


Figure 10. ROC for match detection between all directions with Differential Entropy, as well as benchmark, with SAS highlight (HL) and with SAS shadow (SH) segmentation. We note that there was little variation across over 10,000 threshold values tested, which may lead the resulting curves to contain points concentrated in specific parts of the curve only.

5.4. Information Exchange

Our final analysis examines the statistical relationship between the SAS and the optical images. In Table 4, we report the differential entropy values for the set of positive and negative matches. We conclude that the information from optical and SAS images is indeed statistically related, with values of approximately 3 and 5 for positive and negative series, respectively, compared to roughly 3 and 4 from the SAS to optical direction. As for the highlight and shadow effects, respectively, we find that there is little difference when comparing the two with the optical features.

Table 4. Values of differential entropy between optical–SAS (HL: highlight only, SH: shadow only) pairs, for the most appropriate features (see Table 3).

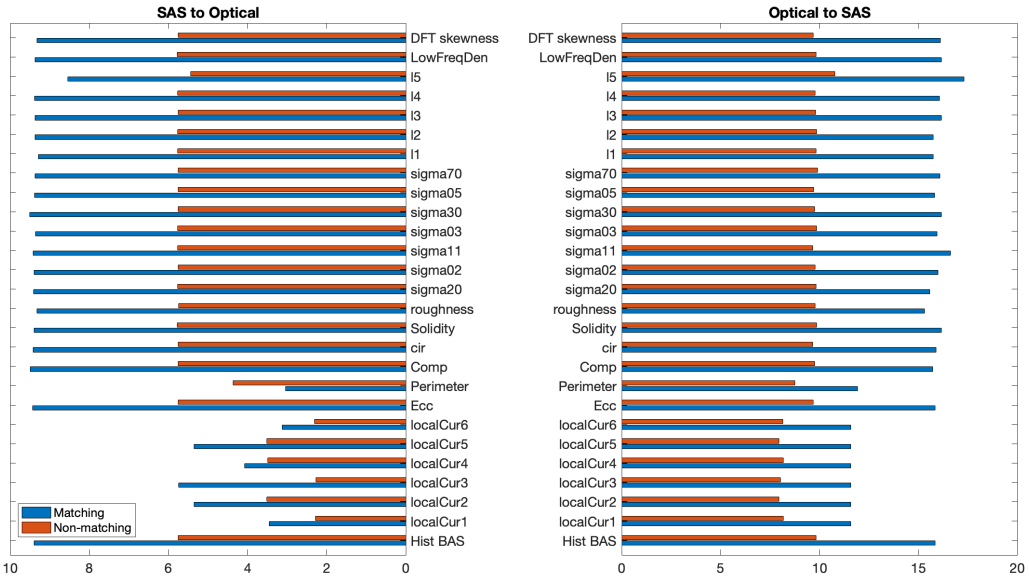
Optical to SAS			
Positive (HL)	**Negative (HL)**	**Positive (SH)**	**Negative (SH)**
3.0744	5.1850	3.3481	5.2553
SAS to Optical			
Positive (HL)	**Negative (HL)**	**Positive (SH)**	**Negative (SH)**
3.4682	4.4457	3.4682	4.4457

To calculate the joint entropy and conditional probabilities, we estimate the probability density function of each data set using a numerical histogram. We find that the entropy between matched optical and SAS datasets is consistently lower than the entropy between unmatched optical and SAS datasets, for all features. This means that there is indeed more information flowing between optical and SAS images of the same object than in images of different objects.

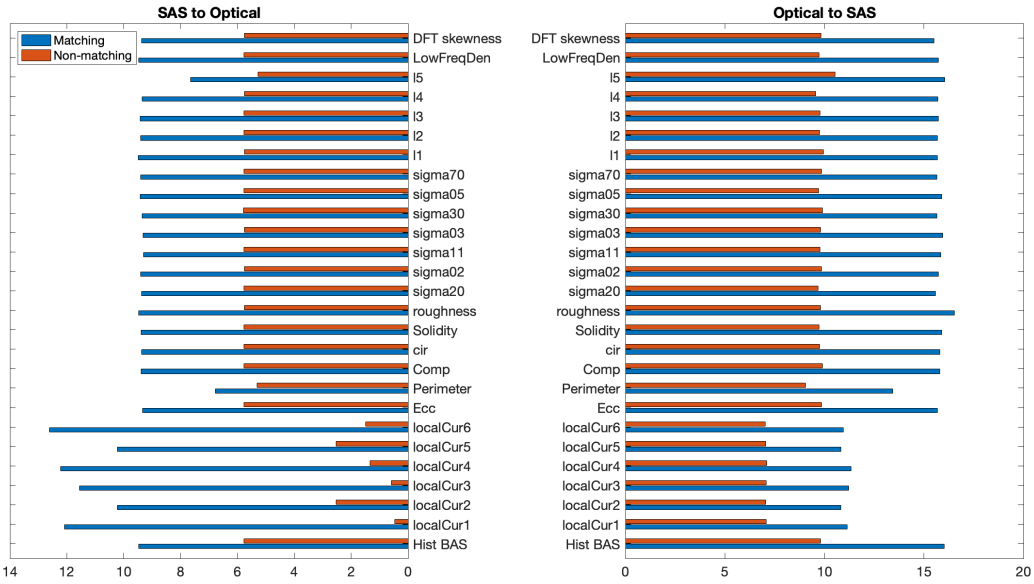
Finally, we examine the conditional probability of the feature vectors to determine whether $P(O|S) = P(S,O)/P(S) \neq P(O)$. To do this, we calculate the conditional probabilities $P(O_1|S_1), P(O_0|S_0)$ and the marginal probabilities $P(O_1), P(O_0)$, and determine their distance using the Kullback–Leibler divergence (KLD) [23]. Formally, the KLD for two distributions P and Q over one variable X, the KLD is derived as follows:

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in X} P(x) \log\left(\frac{P(x)}{Q(x)}\right). \tag{8}$$

In Figure 11, we show the resulting KLDs, both when the SAS feature vector comes from the highlight and when it comes from the shadow segmentation. We find that, for all features, the divergence between the conditional and marginal probabilities of the matched sonar-optical feature vectors in the optical-to-SAS direction is significantly higher than for unmatched feature vectors. While only small differences can be seen between SAS highlight and shadow, the differences between the later features, including those selected for the similarity test, sigma20 and l2, are larger. Regarding the SAS to the optical direction, smaller deviations and a smaller difference between matched and mismatched deviations are observed. This is to be expected, as was also seen in the matching test, where this direction did not provide much relevant information flow. An exception is the localCur features from the SAS shadow, which showed a much higher divergence than all other features in the SAS-to-optical direction. This seems consistent with the fact that these features also had the highest K-Means accuracy for some entropy measures in this direction, such as differential entropy and transfer entropy, although no entropy measure was successful in detecting positives above chance level, likely due to the lack of information that can flow from SAS to optical compared to the reverse direction.



(**a**)



(**b**)

Figure 11. KLD for each feature, between optical and SAS feature vectors, matching (blue) and non-matching (red). (**a**) KLD between conditional and marginal with SAS highlight only. (**b**) KLD between conditional and marginal with SAS shadow only.

5.5. Discussion

An interesting result is that the detection performance depends on the direction of the comparison, i.e., from optical to SAS or from SAS to optical. In particular, better performance is obtained in the optical to SAS direction. This is to be expected since the optical images contain more information due to their much higher resolution and color coding. The results of the statistical relation exploration support this observation by showing a higher differential entropy for the optical to SAS direction than vice versa. Another observation in this context concerns the performance when considering the highlight vs. the shadow ROI. The results suggest that, at least for our dataset, there is generally little difference when only one or the other is considered. This means that both can correctly express the geometric information from the SAS image relative to the geometric information from the optical image, at least when considered separately. Further work can be derived from

our results to explore the significance of this information gain. In other words: which areas of the object contribute most to the similarities between optical and SAS images?

Another consideration for our system is the resolution of the SAS image. The higher the resolution of the SAS image, the better the object matching performance is likely to be. This is because a higher resolution is directly reflected in a better representation of the object features. Possible methods for high-resolution SAS can be found in [24,25]. In our system, we used a two-sided SAS system manufactured by Kraken on an AUV with a resolution of 2 cm per pixel and a one-sided aperture of 100 m.

We recognize that the statistical nature of our approach is compromised by the few pairs of optical and SAS images. However, since, to our knowledge, there is no analytical model for directly converting or comparing these two modalities, this statistical approach provides a simple and intuitive way to detect matched multimodal optical and SAS data. Furthermore, this method does not require the images of both modalities to be transformed or filtered, which could lead to a loss of relevant information, but instead examines the information from the image of each modality as it is. It should also be noted that the acoustic and optical images in our database were recorded from different angles. This might affect the results when comparing them directly, but to obtain a robust method, we indeed wish to have different representations of the same object. We apply a number of enhancements (rotations and scaling) to the images to obtain different representations and assume that these cover as many different angles as possible in order to be able to compare the images as statistical distributions.

6. Conclusions

In this work, we have used statistical measures and tests to evaluate and identify the relationship between optical and SAS images of different objects. In particular, we examined the performance of an entropy-based matching test and investigated the differential entropy and the divergence between the conditional and marginal probabilities of optical and SAS feature vectors. Despite the limited data available for this study, the results suggest a statistical relationship between matching optical and SAS image pairs. These results could serve as a basis and stimulus for further research on combining these two modalities.

Author Contributions: Conceptualization, R.D.; methodology, R.C. and R.D.; software, R.C.; validation, R.C.; formal analysis, R.C. and R.D.; investigation, R.C.; resources, R.D.; data curation, R.C. and R.D.; writing—review and editing, R.C. and R.D.; visualization, R.C. and R.D.; supervision, R.D.; project administration, R.D.; funding acquisition, R.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Horizon Europe programme of the European Union under the UWIN-LABUST project (project #101086340).

Data Availability Statement: Dataset is available on https://drive.google.com/file/d/1ggz8BA0W6CtRXH-48XTfy9eO23-Ag1Qt/view?usp=sharing, accessed on 14 February 2024.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Chantler, M.; Lindsay, D.; Reid, C.; Wright, V. Optical and acoustic range sensing for underwater robotics. In Proceedings of the OCEANS'94, Brest, France, 13–16 September 1994; Volume 1, pp. 205–2010.
2. Abu, A.; Diamant, R. Feature set for classification of man-made underwater objects in optical and SAS data. *IEEE Sensors J.* **2022**, *22*, 6027–6041. [CrossRef]
3. Fumagalli, E.; Bibuli, M.; Caccia, M.; Zereik, E.; Del Bianco, F.; Gasperini, L.; Stanghellini, G.; Bruzzone, G. Combined acoustic and video characterization of coastal environment by means of unmanned surface vehicles. *IFAC Proc. Vol.* **2014**, *47*, 4240–4245. [CrossRef]
4. Abu, A.; Diamant, R. Enhanced fuzzy-based local information algorithm for sonar image segmentation. *IEEE Trans. Image Process.* **2019**, *29*, 445–460. [CrossRef] [PubMed]

5. Gubnitsky, G.; Diamant, R. A multispectral target detection in sonar imagery. In Proceedings of the OCEANS 2021: San Diego–Porto, San Diego, CA, USA, 20–23 September 2021; pp. 1–5.
6. Schettini, R.; Corchs, S. Underwater image processing: State of the art of restoration and image enhancement methods. *EURASIP J. Adv. Signal Process.* **2010**, *2010*, 1–14. [CrossRef]
7. Ghani, A.S.A.; Isa, N.A.M. Underwater image quality enhancement through integrated color model with Rayleigh distribution. *Appl. Soft Comput.* **2015**, *27*, 219–230. [CrossRef]
8. Wang, F.; Vemuri, B.C. Non-rigid multi-modal image registration using cross-cumulative residual entropy. *Int. J. Comput. Vis.* **2007**, *74*, 201–215. [CrossRef] [PubMed]
9. Lowe, G. Sift-the scale invariant feature transform. *Int. J.* **2004**, *2*, 2.
10. Ye, Y.; Shen, L.; Hao, M.; Wang, J.; Xu, Z. Robust optical-to-SAR image matching based on shape properties. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 564–568. [CrossRef]
11. Ye, Y.; Bruzzone, L.; Shan, J.; Bovolo, F.; Zhu, Q. Fast and robust matching for multimodal remote sensing image registration. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9059–9070. [CrossRef]
12. Jang, H.; Kim, G.; Lee, Y.; Kim, A. CNN-based approach for opti-acoustic reciprocal feature matching. In Proceedings of the ICRA Workshop on Underwater Robotics Perception, Montreal, QC, Canada, 24 May 2019; Volume 1.
13. Fei, T. Advances in Detection and Classification of Underwater Targets Using Synthetic Aperture Sonar Imagery. Ph.D. Thesis, Technische Universität, Berlin, Germany, 2015.
14. Xu, S.; Luczynski, T.; Willners, J.S.; Hong, Z.; Zhang, K.; Petillot, Y.R.; Wang, S. Underwater visual acoustic SLAM with extrinsic calibration. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September–1 October 2021; pp. 7647–7652.
15. Xu, Z.; Haroutunian, M.; Murphy, A.J.; Neasham, J.; Norman, R. An Integrated Visual Odometry System With Stereo Camera for Unmanned Underwater Vehicles. *IEEE Access* **2022**, *10*, 71329–71343. [CrossRef]
16. Cardaillac, A.; Ludvigsen, M. Camera-Sonar Combination for Improved Underwater Localization and Mapping. *IEEE Access* **2023**, *11*, 123070–123079. [CrossRef]
17. Abu, A.; Diamant, R. A SLAM Approach to Combine Optical and Sonar Information from an AUV. *IEEE Trans. Mob. Comput.* **2023**. https://doi.org.10.1109/TMC.2023.3336697. [CrossRef]
18. Nguyen, T.M.; Wu, Q.J. Fast and robust spatially constrained Gaussian mixture model for image segmentation. *IEEE Trans. Circuits Syst. Video Technol.* **2012**, *23*, 621–635. [CrossRef]
19. Faes, L. cTE—Matlab Tool for Computing the Corrected Transfer Entropy. Available online: http://www.lucafaes.net/cTE.html (accessed on 28 July 2023).
20. MacQueen, J. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*; University of California Press: Oakland, CA, USA, 1967; Number 14, pp. 281–297.
21. Wackerly, D.; Mendenhall, W.; Scheaffer, R.L. *Mathematical Statistics with Applications*; Cengage Learning: Singapore, 2014.
22. Shang, R.; Tian, P.; Jiao, L.; Stolkin, R.; Feng, J.; Hou, B.; Zhang, X. A spatial fuzzy clustering algorithm with kernel metric based on immune clone for SAR image segmentation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 1640–1652. [CrossRef]
23. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]
24. Zhang, X.; Wu, H.; Sun, H.; Ying, W. Multireceiver SAS imagery based on monostatic conversion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 10835–10853. [CrossRef]
25. Yang, P. An imaging algorithm for high-resolution imaging sonar system. *Multimed. Tools Appl.* **2023**, 1–17. [CrossRef]