Article

A Stereo Matching Method for 3D Image Measurement of Long-Distance Sea Surface

Ying Yang ^{1,2,*} and Cunwei Lu ¹¹ Information and Systems Engineering, Fukuoka Institute of Technology, Fukuoka 8110295, Japan; lu@fit.ac.jp² Department of Optic Engineering, School of Science, Nanjing University of Science and Technology (NJUST), Nanjing 210092, China

* Correspondence: bd19101@bene.fit.ac.jp

Abstract: Tsunamis are some of the most destructive natural disasters. Some proposed tsunami measurement and arrival prediction systems use a limited number of instruments, then judge the occurrence of the tsunami, forecast its arrival time, location and scale. Since there are a limited number of measurement instruments, there is a possibility that large prediction errors will occur. In order to solve this problem, a long-distance tsunami measurement system based on the binocular stereo vision principle is proposed in this paper. The measuring range is 4–20 km away from the system deployment site. In this paper, we will focus on describing the stereo matching method for the proposed system. This paper proposes a two-step matching method. It first performs fast sparse matching, and then complete high precision dense matching based on the results of the sparse matching. A matching descriptor based on the physical features of sea waves is proposed to solve the matching difficulty caused by the similarity of sea surface image textures. The relationship between disparity and the y coordinate is built to reduce the matching search range. Experiments were conducted on sea surface images with different shooting times and distances; the results verify the effectiveness of the presented method.

Citation: Yang, Y.; Lu, C. A Stereo Matching Method for 3D Image Measurement of Long-Distance Sea Surface. *J. Mar. Sci. Eng.* **2021**, *9*, 1281. <https://doi.org/10.3390/jmse9111281>

Academic Editor: Efim Pelinovsky

Received: 12 October 2021

Accepted: 9 November 2021

Published: 17 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: tsunami measurement; stereo matching; feature vector; leaning cost volume

1. Introduction

The tsunami warning systems operated by the Tsunami Warning Center in the United States rely on teleseismic measurements [1,2]. Limited by the number of measurement instruments, the monitoring scope is limited to the coast adjacent to the earthquake. The submarine cable system in Japan inputs the earthquake magnitude and hypocenter into pre-built simulation models to execute a tsunami warning [3,4]. A lack of parts will result in poor warning accuracy. The Great March 11th Earthquake in Japan provides a grim example for this. Three minutes after the earthquake happened, the earliest tsunami warning predicted that in the Iwate Prefecture the sea level height on shore would reach 3 m, but post-tsunami surveys showed that the average wave height was up to 11.8 m [5]. The German–Indonesian Tsunami Early Warning System takes an “end-to-end” approach to cover the complete tsunami warning chain; its working mechanism is similar to the submarine cable system and the Tsunami Warning systems in India and Australia [6–8]. Other methods include rapid determination of sea level variations caused by tsunamis and tsunami parameter estimation from Global Navigation Satellite Systems, tsunami detection and forecasting by radar on unconventional airborne crafts [9–11]. The sparseness and cost limit their capabilities. In this paper, a much more small-scale and flexible tsunami stereo measurement system is proposed; it scans the sea surface to increase measurement coverage and aims to measure sea level height in real time within its coverage area.

With the advancement in the field of computer vision, some scholars have turned towards building 3D geometry of sea waves [12–15]. Wave research based on the stereo system started to become more common after a partially supervised 3D stereo system

called Wave Acquisition Stereo System (WASS) was proposed [16]. A novel video observational system relying on variational stereo techniques to reconstruct the 3D wave surface was developed [17,18]. Other seminal reconstruction methods adopted local methods to compute the disparity map [13,16,19]. In 2017, Bergamasco et al. proposed an open-source pipeline for the 3D stereo reconstruction of ocean waves [20]. They specifically described all the steps required to estimate dense point clouds from stereo images. The system is mounted 12 m above the mean sea level, covering an area of $85 \times 65 \text{ m}^2$. Currently, mostly stereo systems for wave acquisition are used to compensate for the lost details of sea waves on a small scale, and the 3D reconstruction scope is limited to the sea area near the system. However, the excellent advances in this field of research have given us the inspiration and confidence to build a novel long distance stereo measurement system for tsunami measurement. With the use of telephoto lens, the measurement range of our proposed system is extended to 4–20 km away from the system deployment site, which can meet the requirement of tsunami warning.

Figure 1a shows the proposed system configuration. Figure 1b shows the data processing flowchart of the system. Firstly, we take sea surface images with the proposed stereo system which has one camera on the left and one on the right, then conduct stereo matching, calculate sea level height according to matching results and, lastly, judge if there is a tsunami happening. Stereo matching is one of the key steps, and in this paper we will focus on the stereo matching method of the proposed system. To realize accurate tsunami measurement, we must reduce stereo matching errors to smaller than eight pixels (we give the reasons for this accuracy requirement in the Discussion Section) as well as processing time to less than 24^{-1} s (real time measurement). Sea surface images lack texture and sea water is non-still. Additionally, long distance measurement suffers from large disparities in search range. For these reasons, stereo matching in this system is difficult. The existing stereo matching methods for wave reconstruction stereo systems can be classified into three categories: (1) local methods that suffer the delicate trade-off between the disparity window size (which influences the match localization accuracy) and the required surface smoothness [13,16,19]; (2) global methods that are so computationally intensive that they are unlikely to be used in practice [17,18]; and (3) semi-global methods that are based on well-packaged OpenCV library functions [21] and lose efficacy for long distance sea surface image pairs [20,22].

Figure 1. Proposed tsunami measurement system: (a) system configuration and (b) data processing flow chart.

In this paper, stereo matching is divided into two steps: (1) fast sparse sea wave matching is done by feature vectors to reduce disparity searching range, and (2) pixel-wise dense matching is done by leaning cost volume to realize accurate matching. In our method, a decision tree is built to accomplish sparse matching based on the feature vector [23]. According to the sparse matching result, we can build the leaning cost volume and a penalty volume, and then utilize them to complete the dense matching based on a simplified semi-global matching (SGM) algorithm [24]. The SGM algorithm suggested 16 searching paths

for providing good coverage of the 2D image. To increase the matching speed, in this paper, the number of searching paths is decreased from 16 to 8 as we conduct sparse matching first and a penalty volume is built instead of a constant penalty. The rest of the paper is organized as follows. In Section 2, we will introduce the whole stereo matching method, and in Section 2.2 the first step fast sparse matching method is introduced. In Section 2.3 the second step high precision dense matching is demonstrated. We also formulate the relationship between disparity d and the y coordinate, and a leaning cost volume based on the relationship is built to reduce the time and memory consumption of dense matching. In Section 3 we will show the experimental results. Finally, in the Section 4, the applicability for tsunami measurement and the limitations of the proposed method are discussed.

2. Method

Our proposed stereo matching method is described in distinct processing steps. It is a combination of sparse matching and dense matching. There may be some terms used that can be unclear for a common reader. Thus, we define them in Table 1.

Table 1. Definition of technical terms.

Term	Definition
disparity	the x coordinates difference between the same object on stereo images
descriptor	a method used to describe a feature point/region
cost volume	a 3D matrix to store the cost of each point on the image at different disparities
penalty volume	a 3D matrix to store the discontinuity penalty value of each point in different directions
stereo matching	match the same object on stereo images
cost computation	compute the difference (see Equation (7) for detailed definition) between the two matched pixel points from left and right images of one object
cost aggregation	aggregate the computed cost volume and filter out obvious false cost

2.1. Motivations

Recently, stereo matching utilizing deep neural networks has achieved significant advances [25–27]. We are now also working on stereo matching of sea surface images using a deep learning network. However, there are still some difficulties that remain unsolved, such as the lack of ground truth for supervised networks, low accuracy of unsupervised network results, time and memory consumption, etc. Therefore, this paper focuses on stereo matching using traditional methods. A Siamese network was built to complete sparse stereo matching of sea surface images [28]. Figure 2a shows the matching result. However, it cannot be applied to tsunami measurements due to the high time consumption; the running time of one pair of stereo images lasts more than one minute.

Figure 2. Examples of sparse matching: (a) result of Siamese network and (b) result of RANSAC+SURF.

The key of sparse matching is the selection and description of feature points. Long distance sea surface images are low-texture images. Common feature point detectors, such as Harris, FAST, LoG and DOG can only detect feature points and are not sensitive enough for low-texture images [29–32]. Thus, for sea surface images, we detect feature regions for sparse matching. An adaptive dynamic threshold method is used to detect sea waves as feature regions [33]. Common descriptors, such as SIFT, SURF, GLOH, DAISY and LIOP cannot be adaptive to the feature region's size, which makes them difficult to describe the detected sea waves, because sea waves change randomly in shape, size and location [20,32,34–36]. Figure 2b shows one of the result of these methods (RANSAC+SURF [20,37]), where only well-characterized large waves are being correctly matched. Therefore, we propose a new descriptor for sea waves to perform sparse matching. We will describe this in detail in the next subsection.

The measurement of tsunami by the proposed stereo system requires a smaller than eight pixels matching error. It is difficult to ensure by sparse matching since it is a region-to-region matching method. Thus, the second step (dense matching) needs to be conducted. Dense matching contains: (1) cost computation, such as SAD, MI and NCC, etc. [38] and (2) cost aggregation, such as SGM, graph cuts and BF [24,38,39]. Little research on dense matching has studied the establishment of cost volume. The previous algorithms assume that the disparity changes within a constant small range $\mathbb{D}_s$ so that the algorithm can find the best matching within limited memory space and computing time. However, for long distance sea surface images, the disparity varies in a range of over 600 pixels. The traditional cost volume will lead to greater than 4 GB consumption of memory space. To solve this problem, this paper proposes a leaning cost volume based on the first step sparse matching result. We will give a detailed description in Section 2.3.

2.2. Sparse Matching by Feature Vector

2.2.1. Feature Vector Definition

Wave matching can usually be done using the shape and grayscale (or color) of the waves. However, the shape and grayscale can be easily influenced by some uncontrollable factors such as different extraction thresholds, different shooting angles, etc. Furthermore, using shape matching is computationally complex and time consuming, which makes it difficult to guarantee a processing time of 24^{-1} s or less. Thus, we define the wave feature vector by combining its barycenter, size, circularity, width, height, brightness and diagonal lengths to conduct sparse matching. The following content is the explanation of each element in the feature vector.

Epipolar constraint [40] is usually used to reduce the searching region of matching. The epipolar line can be calculated by RANSC [37]. If the same sea wave is extracted by different thresholds, it will lead to differences in extracted sizes, shapes and edges. Thus, compared with other points within the sea wave, the barycenter [41] is much more stable and it is chosen as the representative of the sea wave location.

Next, diagonal lengths (45° and 135°), height and width can be used to reflect the sea wave shape. Circularity can be used to measure the similarity of the sea wave to a circle. Brightness distribution can also be used as a discriminative feature, but it is not a decisive discriminant feature as shooting from different angles will change the brightness distribution. Figure 3 shows the features.

Figure 3. Illustration of each feature.

The feature vector of sea wave i is defined as follows:

$$F_i = [f_{i1} \ f_{i2} \ f_{i3} \ f_{i4} \ f_{i5} \ f_{i6} \ f_{i7} \ f_{i8}] \quad (1)$$

where:

f_{i1} the location of barycenter, displayed in x and y coordinates of the pixel;

f_{i2} the size of the sea wave, displayed in total number of pixels;

f_{i3}, f_{i4} the width and height of the sea wave, displayed in largest numbers of wave pixel in vertical and horizontal orientations;

f_{i5} the circularity of the sea wave, displayed in a number between 0 and 1. The larger it is the more similar it is to a circle;

f_{i6}, f_{i7} the lengths of the sea wave on the two diagonals' orientations of 45° and 135° ;

f_{i8} the brightness of the sea wave, displayed in the sum of grayscale of all pixel points of sea wave i .

In the following content, the k -th feature of the calculated sea wave is easily expressed in f_k .

From one image pair, we extract two sea wave feature vector sets. As a result, for the proposed stereo system, the problem of sea wave matching can be converted to the problem of matching between two feature vector sets F^L, F^R .

$$F^L = \{F_1^L, F_2^L, \dots, F_n^L\} \quad (2)$$

$$F^R = \{F_1^R, F_2^R, \dots, F_m^R\} \quad (3)$$

where F^L and F^R are the sea wave feature vector sets of the left and right images, n and m are the number of sea waves extracted on the left and right images and F_i^L is the feature vector of the i -th sea wave on the left image.

2.2.2. Fast Matching by Decision Tree

To make judgments based on a global analysis of each feature, the decision tree [23] classifier (DTC) is chosen as the matching strategy. The DTC can be selectively robust to some uncertainty features, so it reduces the influence of the uncontrollable factors while also reduces computational load.

The C4.5 [42] system takes the information gain ratio (4), to choose a split feature for each node. It can resist the uneven distribution of different categories of training samples. Pessimistic Error Pruning (PEP) is used to avoid overfitting.

$$InfoGainRation(P, f_k) = \frac{InfoGain(P, f_k)}{SplitInfo_{f_k}(P)} \quad (4)$$

where P represents the set of training samples and f_k is the k -th feature of calculated sea waves. Please refer to [42] for a detailed definition of $InfoGain(P, f_k)$ and $SplitInfo_{f_k}(P)$.

The labels of the training data are determined by manual identification. We made 853 training samples with 336 correct correspondences and 517 incorrect correspondences. Figure 4 shows the built decision tree. We traverse from the root to the leaves to determine

if the two waves are a correct match. Figure 5 shows two of the sparse matching results on the sea surface images taken at 15:00 and 17:00, with the monitoring range being between 14 km and 20 km. For the sea surface image pair at 15:00 and 17:00: $n = 54$ and $m = 65$, 72 and 63 sea waves are extracted from the left and right images, respectively. A total of 25 and 26 pairs of sea waves are correctly matched by our method, the matching precisions are 100% and 92.3%. The matching accuracy is judged by visual inspection, if the same wave in the left and right images can be recognized by our method as one wave, we consider the matching result to be correct. The computational complexity and running time will be discussed in Section 3.2.

Figure 4. Final established decision tree.

Figure 5. Matching results of the decision tree method: (a,b) are the image pairs taken at 15:00 and 17:00, the top is taken by left camera and the bottom is taken by right camera.

2.3. Dense Matching by Building Leaning Cost Volume

2.3.1. Relationship between Disparity and y Coordinates of Sea Surface Images

In our research, the monitoring distance is 4–20 km, in this case the geodetic distance on the Earth's surface is approximately equal to straight-line distance. For simplicity, in this manuscript we use the straight-line distance in Euclidean space instead of using the distance in Riemannian space. To set up our user coordinate system (UCS), we set the camera location at coordinate origin $(0, 0, 0)$, with axis X , Y and Z as shown in Figure 6. We define the angle between the Z axis and the shooting camera's normal direction as θ and earth surface as E . We assume the earth is a sphere. Its center is $C = (0, H, 0)$, and its radius is R . The coordinate of target T in the UCS is (X, Y, Z) , and the coordinate of its projected point t in the image plane is (x, y) . According to Figure 6 and the pinhole camera

model, we know that the relationship between the target T and its image projection point t is as follows:

$$s \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = A \begin{bmatrix} R & T \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (5)$$

Figure 6. Set up user coordinate system (UCS).

Our UCS shares the same origin with the pinhole camera coordinate and rotates θ around the X axis. We assume this pinhole camera is the left camera. The rotation and translation of it is $\begin{bmatrix} R & T \end{bmatrix}$. To simplify we assume for two well calibrated and rectified cameras that the rotation and translation between left and right cameras are I and $\begin{bmatrix} b & 0 & 0 & 0 \end{bmatrix}^T$. Accordingly, we can formulate the relationship between disparity d and image coordinate y , as Equation (6) shows:

$$y = \frac{fy \times d}{fx \times b} (H \cos \theta - \sqrt{R^2 - X^2 - H^2 \sin^2 \theta - 2H \sin \theta \frac{fx \times b}{d} - (\frac{fy \times b}{d})^2}) + h \quad (6)$$

In (6), y is the y coordinate of the target in the image plane, h (pixel) is the length of half the image plane's height, fx and fy (pixel) are the focal lengths of the camera, d (pixel) represents the disparity between left and right images, b represents the baseline length between left and right cameras, R is the radius of earth, H is the length between the earth center and our UCS origin and X is the coordinate in the X axis of our UCS. From (6), we know the relationship between y and d . Figure 7a illustrates the relationship and the bold line represents the relationship between d and y within the image plane.

Given a sea surface image taken by our system, the disparity d increases when y increases. For target $T' : (X', Y', Z')$, the sea surface height is not 0, and the coordinate y of the projected target point will be smaller than the value calculated according to (6), which will cause a bright region in the disparity map, as Figure 7b illustrates (it is a schematic of a disparity simulation for the case where the sea surface height changes).

Figure 7. Illustration of the relationship between y and d : (a) the image of (6) and (b) illustrates the disparity map in the case where sea wave height is not 0 (like the target T' shown in Figure 6).

According to the relationship between d and y , we can drastically reduce the size of cost volume. A much smaller leaning cost volume is built to accomplish dense matching of long distance sea surface images, although the disparity d varies in a wide range for this class of images. We will introduce its construction in Section 2.3.2.

For simplicity, we can also simply assume that the relationship between y and d is proportional. It still yields a relatively accurate result, but the disparity range of the matching search will be extended and the accuracy will be slightly lower due to the low-texture nature of the sea surface images. In this paper, we still recommend one-time sparse matching for the same class of sea surface images to obtain an exact leaning cost volume first.

2.3.2. Build Leaning Cost Volume

Leaning cost volume can be built based on two facts: (1) the relationship between disparity d and y coordinate obeying (6), and (2) sparse matching has been accomplished, which can offer initial data for fitting the (6) function. This paper utilizes the least square method [43] to fit the function.

To resist the luminance change of left and right images caused by different shooting angles, the descriptor, which is used to compute the matching cost volume, is the combination of intensity, gradient and census [44] features. The matching cost of each pixel p (on the left image) is calculated from the difference between its feature descriptor and the feature descriptor of its suspected correspondence (on the right image), the suspected location being $q = f(p) + d$. The function $f(p)$ calculates the suspected location of the matching point on the right image according to (6), and d is the disparity between left and right pixel points caused by the unevenness of the sea surface. Equation (7) shows the calculation of cost $C(p, d)$ at point p :

$$C(p, d) = 2 - \exp\left(-\frac{(1 - \alpha) \times \min(\|I_p^L - I_q^R\|, \tau_1)}{\sigma_1}\right) \times \exp\left(\frac{\alpha \times \min(\|\nabla_x I_p^L - \nabla_x I_q^R\|, \tau_2)}{\sigma_1}\right) - \exp\left(-\frac{\text{Hamming}(\text{census}(p)^L, \text{census}(q)^R)}{\sigma_2}\right) \quad (7)$$

Here, I_p^L denotes the intensity of pixel point p . ∇_x is the grayscale gradient in the x axis direction. I_q^R is the intensity of the corresponding pixel of p on the right image. The disparity between them is $f(p) + d$. α balances the color and gradient terms' contributions, and τ_1, τ_2 are the truncation values. σ_1, σ_2 balance the census, intensity and gradient terms' contributions. $\text{Hamming}(\text{census}(p)^L, \text{census}(q)^R)$ represents calculating the hamming distance between vector $\text{census}(p)^L$ and $\text{census}(q)^R$. $\text{census}(p)^L$ denotes the census vector of pixel p on the left image, $\text{census}(q)^R$ is the census vector of the corresponding pixel of p on the right image. Census transformation is robust when the overall luminance of the image changes. We can choose the appropriate size of the neighborhood window to generate census vectors of different lengths.

Figure 8 demonstrates the cost volume. (a) is the common cost volume. All the points in the $d = 1$ plane share the same disparity. It is usually adapted when disparity varies in a small range $\mathbb{D}_s$. (b) is the leaning cost volume proposed in this paper, the disparity plane is plane 2, and we can get it according to (6). Although the disparity varies in a large range $\mathbb{D}_l$, we can still search the best matching in the small range $\mathbb{D}_s$. By the leaning cost volume, large disparity change can be calculated in advance according to (6), and we just need to search the disparity $d \in \mathbb{D}_s (= [0, 20])$ interval for the best matching results. It greatly reduces the memory and time consumption of dense matching for the tsunami measurement system.

Figure 8. Illustration of cost volume: (a) common cost volume and (b) the leaning cost volume proposed in this paper for long distance matching.

2.3.3. Fast Dense Matching

Pixel-wise cost calculation is generally ambiguous and wrong matches can easily have a lower cost value than correct ones, due to the low-texture area, noise and so forth. Therefore, we need to add other constraints to remove wrong matches. Observing the sea surface, we find that sea level height change is continuous, thus, we add a constraint that adjacent pixels must have similar disparities, and define the following global energy in Equation [38]:

$$E(D) = E_{data}(D) + E_{smooth}(D) = \sum_p (C(p, d_p) + \sum_{p' \in N_p} u(p, p') T[|d_p - d_{p'}| \neq 0]) \quad (8)$$

$E_{data}(D)$ is the data term of disparity map D , representing the sum of all pixels' cost $\sum_p C(p, d_p)$. $E_{smooth}(D)$ is the smooth term, d_p is the disparity between pixel p on the left image and its suspected matching point on the right image, p' is the adjacent pixel point of p and $T[\bullet]$ equals 1 if the argument is true and 0 otherwise, N_p represents the set of neighboring points of p . The $u(p, p')$ multiplier can be interpreted as the penalty of a discontinuity between p and p' , and in this paper, it is the combination of the intensity difference and space distance, as Equation (9) shows:

$$u(p, p') = P_1 \exp\left(-\frac{\|p - p'\|}{\sigma_{sp}} - \frac{|I_p - I_{p'}|}{\sigma_I}\right) \quad (9)$$

It is inspired by the bilateral filter [39], where P_1 is the maximal penalty for a discontinuous pixel, and $u(p, p')$ decreases when the intensity difference or space distance of pixel p and p' increases, which can preserve the discontinuity of the edge. σ_{sp} and σ_I balance the contributions of intensity and space distance to the discontinuity penalty. The problem of dense matching now converts to find a disparity map D that minimizes the global energy $E(D)$. According to [24], it is an NP-hard problem. Graph cut and SGM are proposed in [45] and [24], respectively, and they can approximately minimize the energy function in polynomial time.

In this paper, we utilize the Semi-global matching algorithm to minimize the energy function; it is similar to [24]. However, different from [24], which adds a constant penalty P_1 for all pixels in the neighborhood of p when the disparity changes a little bit (that is, 1 pixel) and adds a larger constant penalty P_2 for all larger disparity changes, we calculate a dynamic penalty for each pixel in the neighborhood of p based on space and intensity differences, as (9) shows. Thus, we first need to calculate the penalty volume for the image.

The dimension of the penalty volume is $N \times H \times W$. N depends on the neighborhood window size, and W and H are the width and height of the image. Adjacent pixels share the same penalty, which can be used to reduce the size of the penalty volume in the N dimension, like Figure 9 shows. It is the 8-connection situation, where the window size is three.

Figure 9. Demonstration of an 8-connection situation.

$p' : (x + u, y + v)$ is one adjacent pixel of $p : (x, y)$ in the direction (u, v) , where u, v are the horizontal and vertical displacements of the neighboring point with respect to the point p . We know $u(p, p') = u(p', p)$, thus the penalty of p in the (u, v) direction equals the penalty of q in the $(-u, -v)$ direction.

For point p , we have marked all its neighborhood points in 8 directions (in gray). If we calculate the penalty volume from the top left point of image, the penalty of adjacent points in the top left corner of p (in dark gray) has already been calculated. In other words, the penalty in r_1, r_2, r_3, r_4 directions have already been calculated; therefore, we just need to calculate and store the penalty of the remaining r_5, r_6, r_7, r_8 directions. For point $p : (x, y)$, its penalty in the (u, v) direction equals the penalty of point $p' : (x + u, y + v)$ in the $(-u, -v)$ direction, which can be used to search the penalty in r_1, r_2, r_3, r_4 directions from the former pixels' penalty vectors. We know r_i and (u, v) are two different forms of adjacent directions. If $r_i = (u, v)$, we can formulate the relationship between i and u, v as the following Equation shows:

$$i = g(u, v) = \begin{cases} (v + \frac{wds}{2}) \times wds + u + \frac{wds}{2} + 1, v < 0 || (v = 0, u < 0) \\ (v + \frac{wds}{2}) \times wds + u + \frac{wds}{2}, v > 0 || (v = 0, u > 0) \end{cases} \quad (10)$$

i is the index of direction r_i , which can be calculated by function $g(u, v)$. wds is the window size of the neighborhood. For any pixel in the image (except the first row and first column), we only calculate its penalty in the last half of the directions. The penalty $P(x, y)$ in the first half of directions r_i can be searched from the former pixels' penalty vectors, for which we have:

$$\begin{cases} P(x, y)_{r_i} = P(x + u, y + v)_{r_j} \\ i = g(u, v), j = g(-u, -v) \end{cases} \quad (11)$$

After we have established the penalty volume, the minimization method is similar to [24]. Figure 10 shows the summary of all the processing steps of our proposed method, including sparse matching and dense matching.

Figure 10. Summary of processing steps in Section 3.

To complete dense matching, first, a leaning cost volume (size $W \times H \times \mathbb{D}_s$), and a penalty volume ($N \times H \times W$) need to be built in advance. Each element in the two volumes needs to be calculated, thus, the complexity is $O(WH\mathbb{D}_s + WHN)$. Then, we conduct cost aggregation similar to [24]. The calculation of smooth cost (cost aggregation) in one direction requires $O(\mathbb{D}_s)$ steps at each pixel. Each pixel is visited exactly K (aggregation path number) times, which results in a total complexity of $O(KWH\mathbb{D}_s)$.

With the leaning cost volume, the complexity of the building cost volume decreases from $O(WH\mathbb{D}_l)$ to $O(WH\mathbb{D}_s)$. For real long distance sea surface images, $\mathbb{D}_s$ is approximately 20, $\mathbb{D}_l$ is approximately 600, and nearly 97% of running time and memory consumption are saved. In the cost aggregation process, by using the penalty volume, we can reduce the aggregation path number by half (compared to what the author of [24] used), which can save 50% in running time.

3. Experiment Results

In this section, we apply our method to sea surface images taken at different times with different illuminance conditions and shooting locations. We show two kinds of experiment results: sparse matching results and dense matching results. The sea surface images are captured by our tsunami measurement system in three periods: 29 February–6 March 2016; 8–16 March 2017; 18–23 August 2018. There were two sites, Fukuoka Kenritsu Suisan High School and Fukuoka Institute of Technology, with 3 monitoring distances: 14–20 km, 4–10 km and 8–14 km. The experiments were performed on a desktop with a 3.4 GHz Intel core CPU and 6GB of memory with C++ code.

3.1. Configuration of the Proposed System

The stereo system consists of two telephoto cameras to take sea surface images, two Pan/Tilt/Zoom (PTZ) heads that can be rotated in the pan and tilt directions to adjust the sight of each camera, one console panel to manually input control signal to PTZ heads, and two client computers to control the photography by adjusting the parameters of the cameras and receive the images captured by the cameras. There is also one total server to communicate with client computers, and control and send the necessary commands for photography to the client computers. Figure 11 is one of the deployed experiments of the proposed system. The two cameras were deployed 27 m apart on the two ends of Fukuoka Institute of Technology's teaching building A. The measurement area was approximately 8 km (closest point) to 14 km (furthest point) away from the system at a height of approximately 30 m above the sea level. (a) is the system's monitoring area, the red point depicts the position of system, and the area within the yellow lines is the cameras' field of view. (b) is the specific configuration of the proposed system. The top figure of (b) shows the components of our system, the middle figure demonstrates the actual deployment location of two telephoto cameras (the red points), and the bottom figure is the captured images of our system.

Figure 11. Experiment deployment of the proposed system: (a) the monitoring area and (b) actual components and deployment.

3.2. Sparse Matching Results

In order to evaluate the performance of the sparse matching method, we compare the matching results of the proposed method with the RANSAC+SURF method and Euclidean distance method. Precision, recall and runtimes are the three main evaluation terms, with precision and recall defined as the following, respectively:

$$\text{precision} = \frac{nTP}{(nTP + nFP)} \times 100 \quad (12)$$

$$\text{recall} = \frac{nTP}{nTP + nFN} \times 100 \quad (13)$$

where nTP and nFP are the numbers of correctly and wrongly detected correspondences in the matching method, respectively. nFN is the number of correct correspondences that are not detected.

To illustrate the comparative result intuitively, we show six representative images' matching results taken at different times. In Figure 12, ①, ②, ③, ④, ⑤ and ⑥ are the six groups of comparison results of the representative image pairs. The (a) column is the RANSAC+SURF matching results, the (b) column is the Euclidean distance matching results and the (c) column is the results of our proposed method. ① and ② were taken 18–23 August 2018 from Fukuoka Institute of Technology, with a monitoring distance of 8–14 km. ③ and ④ were taken 8–15 March 2017 from Fukuoka Institute of Technology, with a monitoring distance of 4–10 km. ⑤ and ⑥ were taken 29 February–6 March 2016 from Fukuoka Kenritsu Suisan High School, with a monitoring distance of 14–20 km.

From the matching results, we find that the RANSAC+SURF algorithm can generate stable and correct matching results only within the region where features are obvious and distinctive, such as mountains or large size sea waves. Our proposed method can match more than 90% of the sea waves correctly and stably. Among the three methods, our proposed method correctly matches the largest number of sea waves.

We also conducted a quantitative comparison on different image pairs, the performance of each method is shown in Table 2. The average precision of the RANSAC+SURF algorithm was 88.4%, showing that sea waves can be correctly matched. However the average recall was 29.1%, meaning that many sea waves are missed by the algorithm, and it will influence the final precision. The average precision of our proposed method is 95.3%; it is sufficient for the second step dense matching. The ground-truth is established by manual checking. From the comparison results, we can conclude that more than 90% of sea waves can be matched stably and correctly by feature vectors, regardless of different illumination conditions. It can solve the problem of no obvious feature points on the sea surface image during the sparse matching process.

The precision of Euclidean distance matching is 5.3% greater than the RANSAC+SURF method's. It means that the feature vector defined in this paper is effective. Meanwhile, the recall of our method is 39.7% greater than Euclidean distance method's. Thus, we can conclude that decision tree we built is more suitable for sea surface image matching.

We also conducted a much more general comparison on sea surface image pairs with the RANSAC+SURF and Euclidean distance method. Figure 13 shows the comparison of the matching results of RANSAC+SURF, Euclidean distance and the proposed methods. The comparison is conducted on 367 pairs of sea surface images taken in 2016, 2017 and 2018.

Figure 12. Comparison of RANSAC+SURF, Euclidean distance and the proposed method: Rows ① and ② are the images taken at 18–23 August 2018 (8–14 km); Rows ③ and ④ are the images taken at 8–15 March 2017 (4–10 km); Rows ⑤ and ⑥ are the images taken at 29 February–6 March 2016 (14–20 km). Column (a) shows the results of RANSAC+SURF; Column (b) shows the results of Euclidean distance; and column (c) shows the results of the proposed method.

Table 2. Comparison of sparse matching results.

NO	RANSAC+SURF		Euclidean Method		Our Method	
	Precision	Recall	Precision	Recall	Precision	Recall
(a)	100.0	36.6	86.7	63.4	88.1	90.2
(b)	100.0	42.0	88.2	60.0	92.5	98.0
(c)	83.3	9.6	100.0	73.1	99.0	100.0
(d)	81.8	9.7	95.2	43.0	96.7	95.7
(e)	72.7	28.6	92.3	42.6	100.0	92.6
(f)	92.3	48.0	100.0	44.0	95.7	88.0
Average	88.4	29.1	93.7	54.4	95.3	94.1

(a)

(b)

(c)

(d)

(e)

(f)

Figure 13. Comparison of three matching methods on correctly matched numbers and precision: (a,b) 29 February–6 March 2016 (14–20 km), (c,d) 8–15 March 2017 (4–10 km) and (e,f) 18–23 August 2018 (8–14 km).

The left column is the correctly matched sea wave numbers of the three matching methods. The horizontal axis is the image number ordering by image shooting time, the vertical axis is the correctly matched sea wave number, and each point (t, n) on the line represents that there are n correctly matched sea waves on the t th sea surface image. The blue line represents the ground-truth sea wave number; it is counted by manual check. The purple line represents RANSAC+SURF matching results, the green line represents Euclidean distance matching results and the red line represents the number of correctly matched waves by our proposed method. We can find that most of the time, the red line is higher than the green and purple lines, meaning that our proposed method can match most sea waves at the most times. The right column is the matching precision of these three methods. The precision of our proposed method and the Euclidean distance matching is close in many cases, and in some cases, our proposed method is a little bit better than the Euclidean distance matching. Additionally, in many cases, our proposed method is better than RANSAC+SURF method.

3.3. Computational Complexity

To judge if two sea waves can be matched, we need to traverse from the root to the leaf of the decision tree. The maximum comparison time is the depth of the decision tree, six times, and the minimum comparison time is two times. There are nine comparison paths in the decision tree, the average comparison time is 4.25. It is nearly half the length of the feature vector. The Euclidean distance method and normalized cross correlation (NCC) used in SURF/SIFT need to traverse the whole feature vector. Thus, compared with these methods, the decision tree built in this paper can save half the time.

Figure 14 shows the comparison results. The red line represents our proposed method's runtime, the green line is the runtime of the Euclidean distance matching and the purple line represents the runtime of the RANSAC+SURF (NCC is used to calculate similarity) method. We find the runtime of our proposed method and the Euclidean distance matching are very close. Observing the partial original running time data (see the small table in the upper right corner), we find our proposed method is 1–2 ms faster than Euclidean distance matching method in some cases. Our method is surprisingly more effective than the RANSAC+SURF algorithm. As RANSAC+SURF algorithm is point to point matching, the number of extracted feature points is much greater than our method's sea wave number. At the same time, the feature vector length is 128, much longer than our proposed feature vector length.

3.4. Dense Matching Results

Figure 15 shows one of the dense matching results of sea surface images by leaning cost volume. The image was taken at 17:00, with a monitoring distance of 8–14 km. (a) (b) are the left and right images and (c) is the disparity map of (a) and (b). The intensity of each pixel represents the disparity of each pixel between the left and right images. Higher intensity represents larger disparity, and the stripe-like areas in the bottom left and top right corners are disparity missing areas caused by the occlusion between the left and right images. Comparing the original image (a) and (b) with the disparity map (c), we can find that, consistent with the conclusion drawn in Section 2.3, the intensity in the presence of a sea wave is larger than the background intensity. Since there is no planar structure on sea surface images, there is rarely area with the same disparity, it is consistent with figure (c).

To validate the correctness of dense matching results, we choose a line on the left image (a). For each pixel on the line, we compute its matching pixel on the right image according to the disparity map (c).

Figure 14. Comparison of three methods on runtime: (a,b) 29 February–6 March 2016 (14–20 km), (c,d) 8–15 March 2017 (4–10 km) and (e,f) 18–23 August 2018 (8–14 km).

Figure 15. Dense matching results: (a,b) are the left and right images and (c) is the disparity map of (a,b).

Figure 16 shows the matching results, and the two matched pixels are connected by a line. The matching accuracy is checked by manual operation. The black line represents correct matching. The red line represents incorrect matching, when the two linked points break the disparity continuity criterion. The matching accuracy of Figure 16 is 87.0%. In Section 2.3.1, the relationship between y and d is shown in Figure 7a, and it is consistent to Figure 16 result, in which the change tendency of the lines' slopes are consistent with the change of y .

Figure 16. Demonstration of the matching result of pixel points on an oblique line. The top is the image taken by the left camera, and the bottom is the image taken by the right camera.

We also conduct dense matching experiments on sea surface images taken at other times and locations. Figure 17 is the dense matching results of sea surface images, The (a) column is taken by the left camera, the (b) column is taken by the right camera and the (c) column is the disparity map. The irregular striped areas in the disparity map are non-overlapping areas in the left and right camera fields of view. ① and ② were taken 18–23 August 2018 from Fukuoka Institute of Technology, with a monitoring distance of 8–14 km. ③,④ were taken 8–15 March, 2017 from the same site, with a monitoring distance of 4–10 km. ⑤ and ⑥ were taken 29 February–6 March 2016 from Fukuoka Kenritsu Suisan High School, with a monitoring distance of 14–20 km. Due to the lack of ground truth, it is difficult for us to evaluate the dense matching accuracy and manual point by point inspection would be costly, thus, we can only perform a general accuracy check such as Figure 16. Limited by this paper's length, in this paragraph, we only make a general evaluation of the final results. According to the conclusion of Section 2.3.1, the disparity will increase in the area where there is a sea wave, coast or mountains, causing a whitish area on the disparity map. This phenomenon is consistent with our experimentally derived disparity maps. In this case, we tentatively conclude that the our dense matching method is valid, and in the future, we will focus on obtaining ground truth to accurately evaluate the dense matching results.

Figure 17. Disparity maps generated by dense matching: ① and ② are images taken at 18–23 August 2018 (8–14 km); Rows ③ and ④ are images taken at 8–15 March 2017 (4–10 km); and rows ⑤ and ⑥ are images taken at 29 February–6 March 2016 (14–20 km). Column (a) shows the images taken by the left camera; Column (b) shows the images taken by the right camera; and column (c) shows the disparity maps.

4. Discussion and Conclusions

We know that the wavelength of a tsunami wave is very long and usually the increase of sea surface height is small in the place where the tsunami occurs. According to its speed calculation equation $c = \sqrt{gh}$, where g is the acceleration of gravity and h is the water depth, the tsunami moves fast in deep sea areas and slows down near the coast. Thus, even a small increase in sea surface height of only 20 cm could be caused by a tsunami [46] and that has the potential to cause a large sea surface height increase near the coast. As the tsunami waves slow down near the coast, the wavelength becomes shorter while the wave energy is the same, resulting in a significant increase in sea surface height near the coast.

For different subsea topography, a tsunami 20 km away takes about 15–30 min to reach the coast. If our proposed stereo system can detect an abnormal sea surface rise of 20 cm and higher, we will have the opportunity to be able to provide coastal people with 15–20 min of escape time before the tsunami comes ashore. The proposed system has two long focal length cameras of 1140 mm, an acquisition rate of 30 fps, resolution of 1920×1080 pixels and two cameras are deployed 27 m apart from each other. The uncertainties in the calibration processing are controlled [47]. Thus, to realize real-time measurement for tsunami warning, the stereo matching error must be smaller than eight pixels and the running time of stereo matching must be shorter than 24^{-1} s. Furthermore,

the stereo matching method must be capable of processing stereo images under different lighting conditions.

To verify the effectiveness of the proposed method for sea surface images under different lighting conditions, we conducted three groups of experiments, acquiring sea surface images from three different locations over three weeks, and performed sparse matching and dense matching on them. The experiment results show that our proposed method can correctly match greater than 90% of sea waves on the sea surface images, see Figure 13, regardless of the distance from the surface, shooting angles and sea state conditions. The running time of sparse matching ranges from 0–400 ms when the correctly matched number ranges from 0 to 140 for stereo images under different conditions, see Figures 13 and 14, for most stereo images, the running time is less than 40 ms. Without considering the time consumption of dense matching, it can meet the requirement for real-time monitoring. We did not record the running time of dense matching, because it is much longer than sparse matching as we need to traverse the leaning cost volume for K (the searching path number) times (time complexity of $O(KWHD_s)$). It is the major time consumption process, and causes our method to not be able to output matching results in real time. In the future, we will focus on increasing the speed of the dense matching by adding parallel operations and matrix operations. By manually checking one of the dense matching results, the dense matching accuracy is 87.0%, thus making it able to meet the accuracy requirement of a tsunami warning.

Indeed, the tsunami problem is very serious and many elements can affect the system. For example, there are many elements affecting the sea surface elevations such as tidal waves, typhoon waves, tsunami waves, etc. Sea surface changes alone are not sufficient to indicate a tsunami. In our project, we believe that the wavelength of a tsunami wave is much longer than the wavelength of typhoon-induced waves, so when the sea surface height changes in a small area, we consider it to be caused by typhoons or sea breezes, etc. The change in sea surface height caused by tides has a certain time pattern, so when the sea surface height is abnormal in a wide area and the tidal factor is excluded, we will consider the occurrence of a tsunami. Furthermore, in our project, there are also many other research subjects such as 24 h image capture for long-distance sea surface of 4–20 km, image measurement in bad weather such as rain and snow, stereo mapping for binocular stereo vision, calculation of sea level height, method of determining the presence or absence of tsunami and how to estimate the arrival time. As we are limited by the paper length, in this paper, we only introduce the stereo matching algorithm. To achieve a high precision and fast stereo matching of proposed tsunami measurement system, we proposed a two step sea surface image matching method based on feature vectors and leaning cost volume.

To resist the computation load caused by a large disparity range, sparse matching is proposed for sea surface image stereo matching. Experiments on multiple groups of sea surface images show that the sparse matching method based on feature vectors can achieve an average precision of 95.3%, and recall of 94.1%, which is better than the RANSAC+SURF method. Furthermore, the accuracy is large enough for second step dense matching.

We formulated the relationship between the disparity d and y coordinate. To reduce the computation load of dense matching, we constructed leaning cost volume based on sparse matching results. During this process, a dynamic penalty method is taken and penalty volume is calculated before we minimize the energy function. The final dense matching results validate our conclusions. In addition to time consumption, another limitation is the high requirement on image quality, the sparse matching can be conducted only in the case, where there are sea waves on the sea surface and the sea waves are captured by the stereo system. To address this insufficiency, we are considering fusing point-to-point sparse matching.

Differently from traditional sea surface stereo matching methods [13,16,19,20], we firstly perform the stereo matching on long distance sea surface images (4–20 km). Although the method here has many shortcomings, our expectation is that this attempt can open up new and exciting possibilities in terms of wave measurements for tsunami warnings.

Author Contributions: Conceptualization, methodology and validation, Y.Y., methodology implementation and supervision, C.L., writing—original draft preparation, Y.Y., writing—review and editing, C.L., sea surface image resource, C.L. and Y.Y. and other members in Lulab. All authors have read and agreed to the published version of the manuscript

Funding: This research was funded by JSPS KAKENHI Grant Numbers JP17K01331, and the MEXT-Supported Program for the Strategic Research Foundation at Private Universities S1311050.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors want to give their sincere gratitude to Nose Toshihiro for his kind help on derivation the equations in this paper. We would also like to thank the other members of lulab for going with me to complete the surface photography experiment. Also sincere thanks to the Samantha Hawkins for the grammar and spelling check of the manuscript

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hirshorn, B.; Weinstein, S.; Tsuboi, S. On the application of Mwp in the near field and the March 11, 2011 Tohoku earthquake. *Pure Appl. Geophys.* **2013**, *170*, 975–991. [CrossRef]
2. Meinig, C.; Stalin, S.E.; Nakamura, A.I.; González, F.; Milburn, H.B. Technology developments in real-time tsunami measuring, monitoring and forecasting. In Proceedings of the OCEANS 2005 MTS/IEEE, Washington, DC, USA, 17–23 September 2005; pp. 1673–1679.
3. NRI for Earth Science; Resilience, D. Seafloor Observation Network for Earthquakes and Tsunamis along the Japan Trench. Available online: <http://www.bosai.go.jp/inline/seibi/seibi01.html/> (accessed on 2 March 2021).
4. Tatehata, H. The new tsunami warning system of the Japan Meteorological Agency. In *Perspectives on Tsunami Hazard Reduction*; Springer: Berlin/Heidelberg, Germany, 1997; pp. 175–188.
5. Mori, N.; Takahashi, T.; The 2011 Tohoku Earthquake Tsunami Joint Survey Group. Nationwide Post Event Survey and Analysis of the 2011 Tohoku Earthquake Tsunami. *Coast. Eng. J.* **2012**, *54*, 1250001-1–1250001-27. [CrossRef]
6. Lauterjung, J.; Letz, H. *10 Years Indonesian Tsunami Early Warning System: Experiences, Lessons Learned and Outlook*; GFZ German Research Centre for Geosciences: Potsdam, Germany, 2017.
7. Nayak, S.; Kumar, T.S. Indian tsunami warning system. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. Beijing* **2008**, *37*, 1501–1506.
8. Allen, S.; Greenslade, D. Developing tsunami warnings from numerical model output. *Nat. Hazards* **2008**, *46*, 35–52. [CrossRef]
9. Larson, K.M.; Lay, T.; Yamazaki, Y.; Cheung, K.F.; Ye, L.; Williams, S.D.; Davis, J.L. Dynamic sea level variation from GNSS: 2020 Shumagin earthquake tsunami resonance and Hurricane Laura. *Geophys. Res. Lett.* **2021**, *48*, e2020GL091378. [CrossRef]
10. Yu, K. Tsunami-wave parameter estimation using GNSS-based sea surface height measurement. *IEEE Trans. Geosci. Remote Sens.* **2014**, *53*, 2603–2611. [CrossRef]
11. Mulia, I.E.; Hirobe, T.; Inazu, D.; Endoh, T.; Niwa, Y.; Gusman, A.R.; Tatehata, H.; Waseda, T.; Hibiya, T. Advanced tsunami detection and forecasting by radar on unconventional airborne observing platforms. *Sci. Rep.* **2020**, *10*, 1–10.
12. Shemdin, O.H.; Tran, H.M.; Wu, S. Directional measurement of short ocean waves with stereophotography. *J. Geophys. Res. Ocean.* **1988**, *93*, 13891–13901. [CrossRef]
13. Wanek, J.M.; Wu, C.H. Automated trinocular stereo imaging system for three-dimensional surface wave measurements. *Ocean. Eng.* **2006**, *33*, 723–747. [CrossRef]
14. Bechle, A.J.; Wu, C.H. Virtual wave gauges based upon stereo imaging for measuring surface wave characteristics. *Coast. Eng.* **2011**, *58*, 305–316. [CrossRef]
15. Kosnik, M.V.; Dulov, V.A. Extraction of short wind wave spectra from stereo images of the sea surface. *Meas. Sci. Technol.* **2010**, *22*, 015504. [CrossRef]
16. Benetazzo, A. Measurements of short water waves using stereo matched image sequences. *Coast. Eng.* **2006**, *53*, 1013–1032. [CrossRef]
17. Gallego, G.; Yezzi, A.; Fedele, F.; Benetazzo, A. A variational stereo method for the three-dimensional reconstruction of ocean waves. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4445–4457. [CrossRef]
18. Gallego, G.; Yezzi, A.; Fedele, F.; Benetazzo, A. Variational stereo imaging of oceanic waves with statistical constraints. *IEEE Trans. Image Process.* **2013**, *22*, 4211–4223. [CrossRef]
19. Brandt, A.; Mann, J.; Rennie, S.; Herzog, A.; Criss, T. Three-dimensional imaging of the high sea-state wave field encompassing ship slamming events. *J. Atmos. Ocean. Technol.* **2010**, *27*, 737–752. [CrossRef]
20. Bergamasco, F.; Torsello, A.; Sclavo, M.; Barbariol, F.; Benetazzo, A. WASS: An open-source pipeline for 3D stereo reconstruction of ocean waves. *Comput. Geosci.* **2017**, *107*, 28–36. [CrossRef]

21. Marengoni, M.; Stringhini, D. High level computer vision using opencv. In Proceedings of the 2011 24th SIBGRAPI Conference on Graphics, Patterns, and Images Tutorials, Alagoas, Brazil, 28–30 August 2011; pp. 11–24.
22. Vieira, M.; Guimarães, P.V.; Violante-Carvalho, N.; Benetazzo, A.; Bergamasco, F.; Pereira, H. A Low-Cost Stereo Video System for Measuring Directional Wind Waves. *J. Mar. Sci. Eng.* **2020**, *8*, 831. [CrossRef]
23. Safavian, S.R.; Landgrebe, D. A survey of decision tree classifier methodology. *IEEE Trans. Syst. Man Cybern.* **1991**, *21*, 660–674. [CrossRef]
24. Hirschmuller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *30*, 328–341. [CrossRef]
25. Cheng, X.; Wang, P.; Yang, R. Depth estimation via affinity learned with convolutional spatial propagation network. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 103–119.
26. Zhong, Y.; Dai, Y.; Li, H. Self-supervised learning for stereo matching with self-improving ability. *arXiv* **2017**, arXiv:1709.00930.
27. Ren, H.; Raj, A.; El-Khamy, M.; Lee, J. SUW-Learn: Joint Supervised, Unsupervised, Weakly Supervised Deep Learning for Monocular Depth Estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 750–751.
28. Chen, C.H.; Lu, C.W.; Ying, Y. Method of Sea Wave Extraction and Matching from Images Based on Convolutional Neural Network. In Proceedings of the 5th International Conference on Engineering, Applied Sciences and Technology, Luang Prabang, Laos, 2–5 July 2019; pp. 1–4.
29. Harris, C.G.; Stephens, M. A combined corner and edge detector. In Proceedings of the Alvey Vision Conference, Manchester, UK, 31 August–2 September 1988; Volume 15, pp. 10–5244.
30. Rosten, E.; Drummond, T. Machine learning for high-speed corner detection. In Proceedings of the European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 430–443.
31. Marr, D.; Hildreth, E. Theory of edge detection. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **1980**, *207*, 187–217.
32. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
33. Yang, Y.; Lu, C. Long-distance sea wave extraction method based on improved Otsu algorithm. *Artif. Life Robot.* **2019**, *24*, 304–311. [CrossRef]
34. Mikolajczyk, K.; Schmid, C. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1615–1630. [CrossRef] [PubMed]
35. Tola, E.; Lepetit, V.; Fua, P. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *32*, 815–830. [CrossRef] [PubMed]
36. Wang, Z.; Fan, B.; Wu, F. Local intensity order pattern for feature description. In Proceedings of the 2011 International Conference on Computer Vision, Colorado Springs, CO, USA, 20–25 June 2011; pp. 603–610.
37. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [CrossRef]
38. Kim, J. Visual correspondence using energy minimization and mutual information. In Proceedings of the Ninth IEEE International Conference on Computer Vision, Nice, France, 13–16 October 2003; pp. 1033–1040.
39. Tomasi, C.; Manduchi, R. Bilateral filtering for gray and color images. In Proceedings of the Sixth international conference on computer vision (IEEE Cat. No. 98CH36271), Bombay, India, 7 January 1998; pp. 839–846.
40. Zhang, Z.; Deriche, R.; Faugeras, O.; Luong, Q.T. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artif. Intell.* **1995**, *78*, 87–119. [CrossRef]
41. Miclo, L. Isoperimetric stability of boundary barycenters in the plane. *Ann. Math. Blaise Pascal* **2019**, *26*, 67–80. [CrossRef]
42. Quinlan, J.R. *C4. 5: Programs for Machine Learning*; Elsevier: Amsterdam, The Netherlands, 2014.
43. Ledvij, M. Curve fitting made easy. *Ind. Phys.* **2003**, *9*, 24–27.
44. Zabih, R.; Woodfill, J. Non-parametric local transforms for computing visual correspondence. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 2–4 March 1994; Springer: Berlin/Heidelberg, Germany, 1994; pp. 151–158.
45. Boykov, Y.; Veksler, O.; Zabih, R. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 1222–1239. [CrossRef]
46. Office, C. Information about Tsunami. Available online: http://www.bousai.go.jp/kohou/kouhoubousai/h22/05/special_01.html/ (accessed on 16 September 2021).
47. Yi, H.; Tsujino, K.; Lu, C. 3-D image measurement of the sea for disaster prevention. *Artif. Life Robot.* **2018**, *23*, 304–310. [CrossRef]