remote sensing

MDPI

Article

Horizon Line Detection in Historical Terrestrial Images in Mountainous Terrain Based on the Region Covariance

Sebastian Mikolka-Flöry * and Norbert Pfeifer

Department of Geodesy and Geoinformation, TU Wien, 1040 Vienna, Austria; norbert.pfeifer@geo.tuwien.ac.at
* Correspondence: sebastian.floery@geo.tuwien.ac.at; Tel.: +43-1-58801-12253

Abstract: Horizon line detection is an important prerequisite for numerous tasks including the automatic estimation of the unknown camera parameters for images taken in mountainous terrain. In contrast to modern images, historical photographs contain no color information and have reduced image quality. In particular, missing color information in combination with high alpine terrain, partly covered with snow or glaciers, poses a challenge for automatic horizon detection. Therefore, a robust and accurate approach for horizon line detection in historical monochrome images in mountainous terrain was developed. For the detection of potential horizon pixels, an edge detector is learned based on the region covariance as texture descriptor. In combination with shortest path search the horizon in monochrome images is accurately detected. We evaluated our approach on 250 selected historical monochrome images in average dating back to 1950. In 85% of the images the horizon was detected with an error less than 10 pixels. In order to further evaluate the performance, an additional dataset consisting of modern color images was used. Our method, using only grayscale information, achieves comparable results with methods based on color information. In comparison with other methods using only grayscale information, accuracy of the detected horizons is significantly improved. Furthermore, the influence of color, choice of neighborhood for the shortest path calculation, and patch size for the calculation of the region covariance were investigated. The results show that both the availability of color information and increasing the patch size for the calculation of the region covariance improve the accuracy of the detected horizons.

Keywords: horizon detection; region covariance; historical images; edge detection

check for updates

Citation: Mikolka-Flöry, S.; Pfeifer, N. Horizon line Detection in Historical Terrestrial Images in Mountainous Terrain Based on the Region Covariance. *Remote Sens.* **2021**, *13*, 1705. https://doi.org/10.3390/rs13091705

Academic Editor: Bryan Gardiner

Received: 24 March 2021
Accepted: 26 April 2021
Published: 28 April 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/).

1. Introduction

Image archives hold impressive collections of historical terrestrial images from the Alpine regions. These images are an important, yet seldom used, source for identifying and documenting changes in the alpine landscape. Taken by early mountaineers without any auxiliary devices (e.g., global navigation satellite systems), the accurate position and orientation of these images are unknown. With georeferencing, the quantification of visible changes by monoplotting [1] becomes possible. Additionally, instead of relying on missing or wrong metadata, georeferenced images can be queried by their estimated location. Especially for geoscience displines (e.g., Hydrology, Glaciology), where historical archival images predate other sources by nearly 50 years (e.g., aerial imagery), the availability of geroreferenced archival images will promote their usage.

In order to estimate the unknown camera parameters by photogrammetric resection, ground control points are necessary. Without the availability of human-made structures (e.g., buildings, streets), and with vast changes in the topography (e.g., glaciers) and varying illumination conditions, their identification in historical terrestrial images is time consuming and requires experience. In combination with the limited field of view of terrestrial images, analysis is often limited to a few hand-selected images. To overcome this limitation and allow the quantification of changes based on archival terrestrial images at larger scales, automatic methods for the estimation of the unknown camera parameters are

necessary. Automatic image orientation is generally based on the extraction and matching of feature points (e.g., SIFT [2], SURF [3], ORB [4]). Especially in changing environments (e.g., glaciated mountains) in combination with collections of images captured over a time period of 100 years, image matching based on feature points is challenging ([5,6]). Therefore, in natural environments, like mountainous terrain, automatic methods are based on the visible horizon (also referred to as apparent horizon, horizon line, or skyline). Assuming that the visible horizon in an image is unique, it can can be used for matching with rendered horizons from digital terrain models [7–11]. While these methods vary within the matching stage and the required a priori information (e.g., approximate location, orientation, or focal length), they all are based on the visible horizon. Therefore, the accurate and robust detection of the horizon is a crucial prerequisite for the automatic estimation of the unknown camera parameters.

The apparent horizon is the border separating terrain from sky, which is characterized in images by a change in color and texture. These differences are exploited by methods approaching horizon detection as a segmentation task, where each pixel in the image is either assigned to ground or sky. In [7,12], the problem is formulated as a foreground–background segmentation, using multiple dense feature descriptors within a bag-of-words framework. On the other hand, the horizon line detection can be interpreted as an edge detection problem. Ref. [13] proposed a method based on edge detection and a subsequent extraction step using dynamic programming. In [14], the problem of non-continuous edges, often a result of traditional edge detection, is overcome by training a support-vector machine (SVM) on the normalized gray-scale values extracted from patches around each pixel. Based on the derived probability map, the horizon line is detected by searching for the minimum cost path. More recently, [15] used a convolutional neural network to classify patches extracted for each pixel. Instead of classifying individual patches, [16] use a fully convolutional network. A different approach was suggested by [17] based on the observation, that sky regions generally are more homogeneous compared to ground. The potential horizon line is found by maximizing the difference in variance above and below the horizon.

All of the proposed methods have only been applied to modern images. Except for the approach proposed by [14], color information is used. In contrast, the detection of the visible horizon in historical terrestrial images has not been addressed yet. Besides the unavailability of color information, these images pose additional challenges (e.g., Figure 1):

- Taken up to 100 years ago with the earliest available compact cameras, historical images have a poorer image quality resulting in blurred images, low contrasts, and overexposed regions.
- Most images found in archives were acquired from private collections. As those images have not been stored professionally, they show varying signs of usage (e.g., dirt, scratches, watermarks).
- Alpine environments, even with modern cameras, pose a challenging environment for photography. Especially snow and glacial areas are difficult to photograph due to strongly varying illumination. In combination with rapidly changing weather (e.g., fog, clouds) conditions, visual appearance is heavily affected.

Especially the lack of color information in combination with high alpine terrain, partly covered with snow or glaciers, poses a particular challenge for the detection of the horizon line. Therefore, a robust and accurate method for detecting the horizon in these challenging images is needed. Based on the work of [14,18], a new method is suggested. The region covariance is used in classification to describe texture within the neighborhood around pixels. Introducing the probabilities obtained from a random forest classifier as weights into a shortest path search the horizon is obtained. In order to evaluate the performance of the approach, two datasets representing modern and historical images were used.

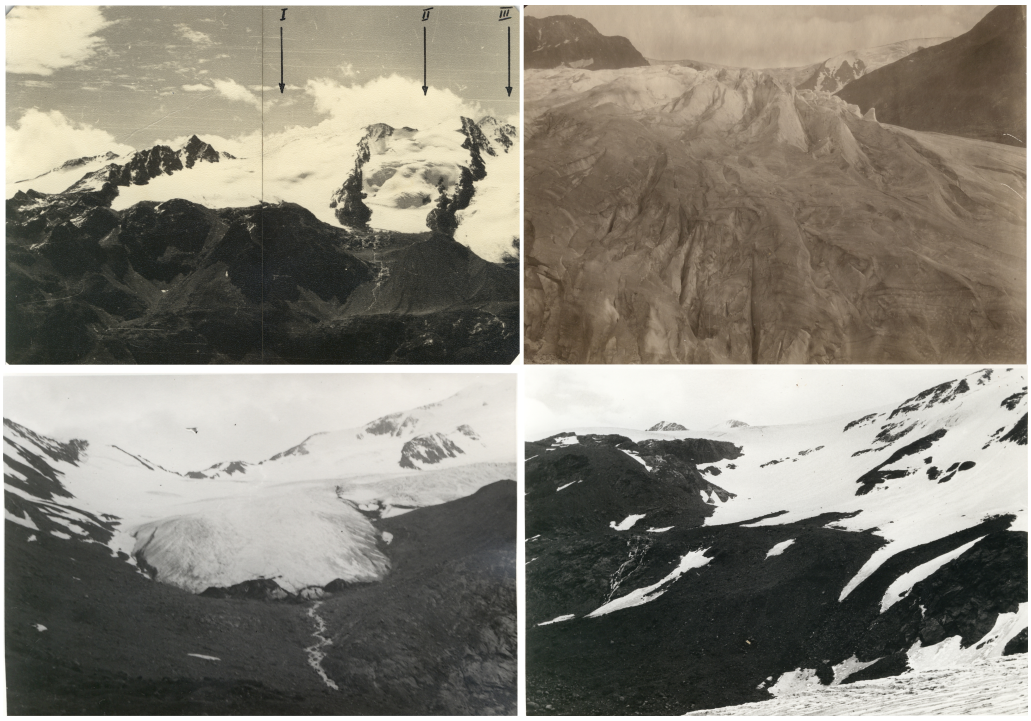


Figure 1. Selected images from the HIST dataset demonstrating the challenges posed by historical terrestrial images in mountainous terrain. (**Top left**): As images have been acquired from private collections, these partly contain handwritten marks. Additionally, the snow covered regions are barely distinguishable from the clouds in the background. (**Top Right**): Close up of a glacier. (**Bottom Row**): Without color information, the parts of the horizon which are covered with snow can only be vaguely guessed based on the slightly visible surrounding rock structures.

While the appearance of terrain and sky varies strongly, the horizon generally follows a certain spatial layout: The region below the horizon has larger intensity variations, whereas the region above, sky, appears homogeneous. Following [14], describing the spatial distribution of intensities surrounding horizon pixels, an edge detector is learned (Figure 2).

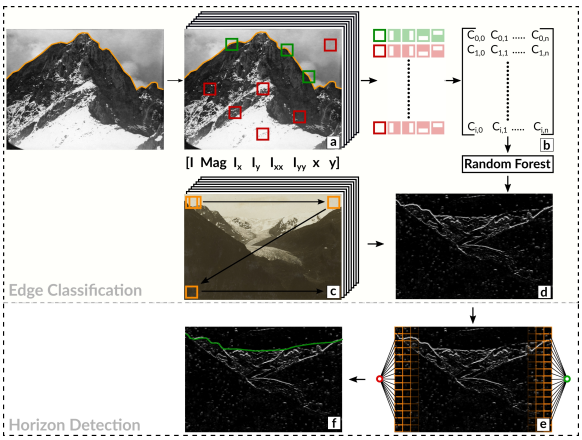


Figure 2. Extracting horizon (green) and not-horizon (red) patches from training images (**a**), the region covariance features are calculated (**b**), which are directly used as inputs to a random forest classifier. The probability map (**d**) is obtained by classification of densely extracted patches from a test image (**c**). Using shortest path search on a graph defined by the pixel grid (**e**), the horizon in the test image is detected (**f**).

2. Method

Various descriptors have been proposed for describing the appearance and layout of pixels in images (generally referred to as texture), e.g., LBP [19], GLCM [20], or filter banks [21]. Among those, the region covariance proposed in [18] shows interesting properties regarding our problem: Being calculated from various image statistics within a selected region, the covariance matrix is directly used as texture descriptor. Hence, no additional parameters must be defined except for the size of the region of interest. Depending on the used features, it is robust towards illumination, scale, and rotation changes. As originally proposed for object detection, the spatial appearance of objects can be easily incorporated. Furthermore, the region covariance is a compact descriptor which can be calculated efficiently using integral images [22].

2.1. Region Covariance

The covariance matrix C_R for a region R of a grayscale image is calculated using

$$C_R = \frac{1}{n-1}\sum_{k=1}^{n}(I_k - \mu)(I_k - \mu)^T \tag{1}$$

with I_k being the intensities of each pixel within the region R, and μ the mean intensity. For a monochrome image, C_R contains only the variance σ_I^2 of the intensity. Therefore, [18] extended the image by additional feature channels, which can be directly derived from the intensity values. With multiple dimensions, the main diagonal of C_R contains the variances of the respective features, whereas the other values represent the pairwise covariance. Among the proposed features are the intensity I, magnitude of the gradient Mag, the first and second order derivatives and x and y positions (Equation (2)).

$$F(x,y) = \begin{bmatrix} I & Mag & \frac{\partial I}{\partial x} & \frac{\partial I}{\partial y} & \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial y^2} & x & y \end{bmatrix} \tag{2}$$

In case color images are available, the feature channels can be easily changed to include color information instead of the intensity

$$F(x,y) = \begin{bmatrix} R & G & B & Mag & \frac{\partial I}{\partial x} & \frac{\partial I}{\partial y} & \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial y^2} & x & y \end{bmatrix} \tag{3}$$

The gradient magnitude Mag is calculated using the root squared sum of the first order image derivatives in x and y. The usage of the x and y image coordinates seems contradicting. While their variance is always the same, the covariances with the other features represent important information regarding their distribution within the selected patch.

2.1.1. Patch Representation

While the appearance of horizons vary, their spatial layout is consistent: A smoother sky region above, in contrast to the rougher terrain with stronger variations below. Therefore, not only the intensity values, but especially the image gradients and gradient magnitude, will have stronger variations for the terrain region in the bottom part of horizon patches. With the inclusion of the x and y coordinates into the calculation of the region covariance, some measure to describe the distribution of the features within each patch is already available.

To further emphasize the spatial structure of horizon patches, each patch is represented by five individual parts (Figure 3), which was already suggested by [18]. For each of the five individual regions of a patch, a covariance matrix describing the spatial–textural relationship is calculated. The concatenation of all covariance matrices of one patch then builds the basis within the classification.

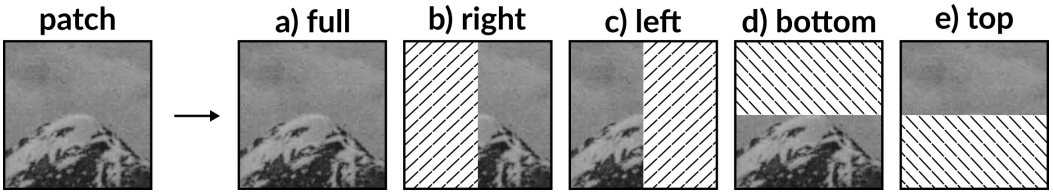


Figure 3. Five covariance matrices for each patch are calculated. Each one captures certain characteristics of the expected horizon regions.

2.1.2. Classification

Each covariance matrix is a symmetrical square matrix of dimension $d \times d$ with d being the number of features used. Due to symmetry, each matrix contains $(d^2 + d)/2$ unique values. As five regions for one patch are used, the length of the resulting feature vector is $5 \cdot ((d^2 + d)/2)$.

One limitation of covariance matrices is, that they are not embedded in an euclidean space [18]. Therefore, the individual elements of the covariance matrices cannot be directly interpreted as features in a subsequent classification task. In order to overcome this limitation, various approaches have been proposed. Ref. [18] used region covariances within a nearest neighbor classification making use of the distance metric proposed by [23] to calculate the pairwise distances. In case of support-vector machines, specifically designed kernels were suggested by [24,25]. Both also experimentally evaluated the influence of directly using covariance features in contrast to considering their geometry. For random forests, the log-euclidean transformed region covariances [26] have been used by various authors [27–29].

Following [26], region covariances C_R can be used within euclidean computations after transformation into their matrix logarithms. One prerequisite for this approach are symmetric positive–definite matrices. Per definition, C_R are symmetric positive semi-definite matrices, which can be transformed into positive definite matrices by adding a small constant factor $\epsilon = 1 \times 10^{-6}$:

$$C_R = C_R + \epsilon I_d \tag{4}$$

where I_d is the d-dimensional identity matrix. After transformation into a positive definite matrix, the logarithm $\log(C_R)$ can be calculated using the eigenvalue decomposition

$$C_R = UDU^T \tag{5}$$

with D being the diagonal matrix of the eigenvalues of C_R, and finally,

$$\log(C_R) = U \log(D) U^T \tag{6}$$

where $\log(D)$ is the diagonal matrix of the eigenvalue logarithms. The elements of $\log(C_R)$ now can be directly used within the random forest classification.

For training, positive training patches are sampled along the ground truth horizon. To account for the imbalance of horizon and not-horizon pixels in images, negative training patches are randomly sampled with a ratio of 4:1 (not-horizon:horizon).

2.2. Horizon Line Detection

Following [13,14] the horizon line detection is solved by searching for the shortest path from the left to right image border. A directed weighted graph $G(N, E)$ is defined with each node N corresponding to a pixel of the input image and edges E connecting two pixels of the image in the graph. Each edge has a weight $w(E)$. Additionally, two special nodes called "source" and "sink" are added to the graph G. The source node is connected to all pixels of the first image column, whereas the sink node to all pixels of the last image column. For all edges connected either to source or sink, the weight $w(E)$ is set to zero. As the edges have zero weight, they do not have any influence on the shortest path itself.

On the other hand, as both nodes are connected to all pixels of the first or last column, the shortest path can reach any pixel within the image. Therefore, no approximate location of the horizon must be known a priori.

2.2.1. Neighborhood Definition

Two different neighborhood definitions proposed in [14,16] are considered. The first one, N1, connects each pixel with the direct neighbors in the next image column (Figure 4-left). In contrast, for N3 all pixels in the next image column, which are separated by up to 3 rows below or above the current pixel (Figure 4-right), are regarded as neighbors and are connected with an edge in the graph used for the shortest path calculation. According to the authors, this should allow the detected horizon to follow along steeper mountain profiles.

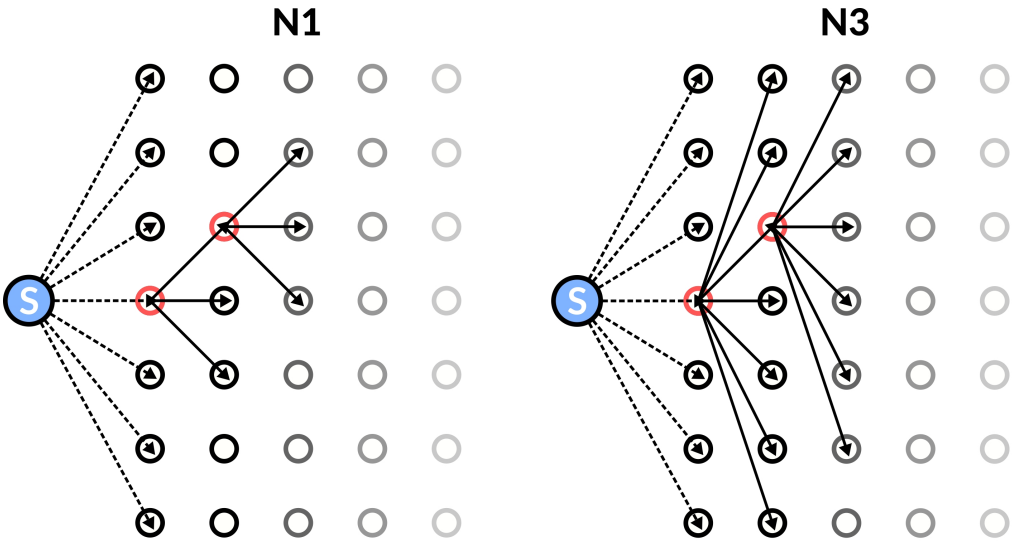


Figure 4. Different neighborhood definitions for the construction of the image grid graph used within the shortest path calculation. All pixels (indicated as circles) of the first column are connected to the source node (blue). For the N1 neighborhood, each pixel is only connected to direct neighbors in the next image column. For the N3 neighborhood, each pixel is connected to all pixels in the next column, which are separated by up to three rows. Connections are shown for two exemplary nodes highlighted in red.

2.2.2. Edge Weight

The horizon we want to detect corresponds to the shortest path minimizing the sum of the edge weights from the left to right image border. Therefore, the definition of the edge weights, except for those edges connecting nodes to either source or sink, is essential. The output of the random forest classifier is in the range $[0, 1]$ with 1 indicating a high probability that a pixel is part of the horizon. As we are searching for the path with the minimum sum of edge weights from left to right image border, the negative logarithm of the probability derived from the random forest classifier is used as the cost for the corresponding node N:

$$Cost_N = \log(P) * (-1) \tag{7}$$

where P is the probability of the pixel. Therefore, a pixel with a high horizon probability corresponds to a node in the graph with a low cost and vice versa. As $\log(0)$ is undefined, all pixels with a probability of zero are set to $1 * 10^{-6}$ prior to the logarithmic transformation. Each edge of the graph connects two pixels of the image. Hence, for the definition of the edge weights two candidates exist: the cost of the node where the edge originates from

(N_f) and the cost of the node where the edge points at (N_t). As defined in Equation (8), the cost of the target node is used:

$$w(E_{(N_f,N_t)}) = Cost_{N_t} \tag{8}$$

2.3. Evaluation

A commonly used metric [14,16,30], the average vertical deviation from the ground truth horizon, is used for accuracy evaluation:

$$D_{img} = \frac{1}{N_{col}} \sum_{j=1}^{N_{col}} |R_{gt}^j - R_{est}^j| \tag{9}$$

N_{col} is the number of columns in the image, R_{gt}^j is the row of the ground truth horizon in column j, and R_{est} is the row of the estimated horizon in column j. In order to derive the statistics for the whole dataset, D_{img} is calculated for each image, and the mean (μ) and standard deviation (σ) are calculated.

In both datasets, the ground truth horizons have no vertical or overhanging geometries, making this metric sufficient to determine accuracy. Only in cases of falsely detected horizons may a column contain multiple horizon pixels. In these rare cases, the position of the highest pixel is used.

3. Datasets

To evaluate the performance of our approach two different datasets, CH1 and HIST, are used. For the CH1 dataset results are reported for various approaches in [16], and is therefore suitable for comparison. The HIST dataset is a collection of historical terrestrial images used within our work.

3.1. CH1

This dataset (https://cvg.ethz.ch/research/mountain-localization/, accessed on 27 April 2021) was made available by [7], consisting of 203 color images, including ground truth sky masks, captured across Switzerland. All images have similar aspect ratios and dimensions (~1024 × 700 px). It contains modern photographs of mountainous terrain including challenging situations like cloud occlusions, low contrast horizons, and images taken from within cable cars (Figure 5). The dataset was used by a wide range of authors (e.g., [16,31]) for evaluation. While it consists of color images, for the developed method, the grayscale images are derived by averaging the color channels.



Figure 5. Challenging images from the CH1 dataset: Image taken from within a cable car (**left**); chairlift visible in the sky region (**middle**) and barely visible horizon due to strong haze (**right**).

3.2. HIST

The HIST dataset is a selection of 250 historical grayscale images with manually created ground truth horizons. All images were captured in high alpine terrain in the Alps. Compared to the CH1 dataset, many images show detailed views of glaciers and their

surroundings. Combined with the reduced image quality and lack of color information, these images present a particular challenge for horizon detection. Some examples are shown in Figure 1.

Collected from various archives, all images had already been digitized. For most, the original format (e.g., film, print) and information regarding the scanning process (e.g., scanner type and model, DPI) are unknown. Some of the images contain borders which were semi-automatically detected prior to the horizon line detection and excluded from all subsequent steps. In contrast to CH1, image dimensions range from approximately ~700 × 1000 px to ~2500 × 1800 px including both portrait and landscape images.

In total, the images span a time period of nearly 100 years. The earliest image dates back to 1895 with the most recent one taken in 1992. The majority of the images were captured between 1920 and 1950.

4. Results

4.1. CH1

For the CH1 dataset our approach is compared with the results of various methods reported in [14,16,32,33]:

DCNN-DS1: Fully convolutional network proposed in [16];
SVM: Patch classification using normalized gray-scale values [14];
CNN: Convolutional neural network trained on patches proposed in [14];
Structured Forest: Learned edge detection proposed in [32];
Random Ferns: Learned edge detection proposed in [33].

From the reported methods, SVM is most similar to our approach as it is based on normalized pixel intensities extracted from 16 × 16 patches. Therefore, we re-implemented SVM based on the details given in [14]. In contrast to the original approach, patches with an odd dimension of 17 pixels are extracted to symmetrically describe the regions surrounding each pixel. As the CH1 dataset contains color images, two different variants of our method were evaluated: The first one, "cov_gray", uses the feature channels as shown in Equation (2). Additionally, "cov_color" uses the R, G, and B channels instead of the intensity (Equation (3)).

The evaluation strategy from [16] is adopted by splitting the dataset into 60% training and 40% testing sets. As no details are given regarding which images were used for training and testing, random splitting of the dataset is repeated five times. μ and σ are calculated from the images in all testing sets.

Based on the mean error and standard deviation, cov_color and cov_gray rank second and third among all reported methods (Table 1). Only DCNN-DS1 based on a fully convolutional network achieves higher accuracy and lower standard deviation compared to cov_color. In contrast to the approaches achieving higher accuracy, cov_gray is solely based on the grayscale images. With regard to the HIST dataset, consisting only of historical monochrome images, comparison with SVM and CNN is of special interest. Using cov_gray, the average error was reduced from 22.57 px (SVM) and 5.78 px (CNN) to 3.51 px. As both SVM and cov_gray make use of exactly the same patches within training, the higher accuracy results exclusively from the improved horizon classification based on the region covariance.

Table 1. Difference between ground truth and estimated horizon line for the CH1 dataset. SVM (re-implemented), CNN [14], DCNN-DS1 [16], Structured Forest [32], Random Ferns [33], and cov_gray, cov_color (ours).

Color Space	Method	Error [px]	
		μ	σ
RGB	DCNN-DS1	1.53	2.38
RGB	**cov_color**	**2.78**	**7.29**
G	**cov_gray**	**3.51**	**11.82**
G	CNN	5.78	14.73
RGB	Structured Forest	8.07	18.92
RGB	Random Ferns	10.21	20.41
G	SVM	22.57	49.47

4.1.1. Influence of Color

As outlined in Section 2.1, the region covariance matrix can be easily adjusted to monochromatic or color images. As color images are available in the CH1 dataset, it is possible to evaluate the role of color for horizon detection. According to the results in Table 1, cov_color is more accurate than cov_gray. This indicates that color information is beneficial and contains additional information to distinguish the horizon from other apparent silhouettes. Analyzing the results of both variants in more detail revealed, that for most images, the influence of color is minor, as the region covariance based on the grayscale information is already a powerful and descriptive patch descriptor. However, in cases where additional structures having similar visual appearances in the grayscale domain exist, color information is important. Pixels along the true horizon are classified with higher probability when color information is available, whereas those along structures in the foreground are classified with lower probabilities leading to reduced errors. (Figure 6).

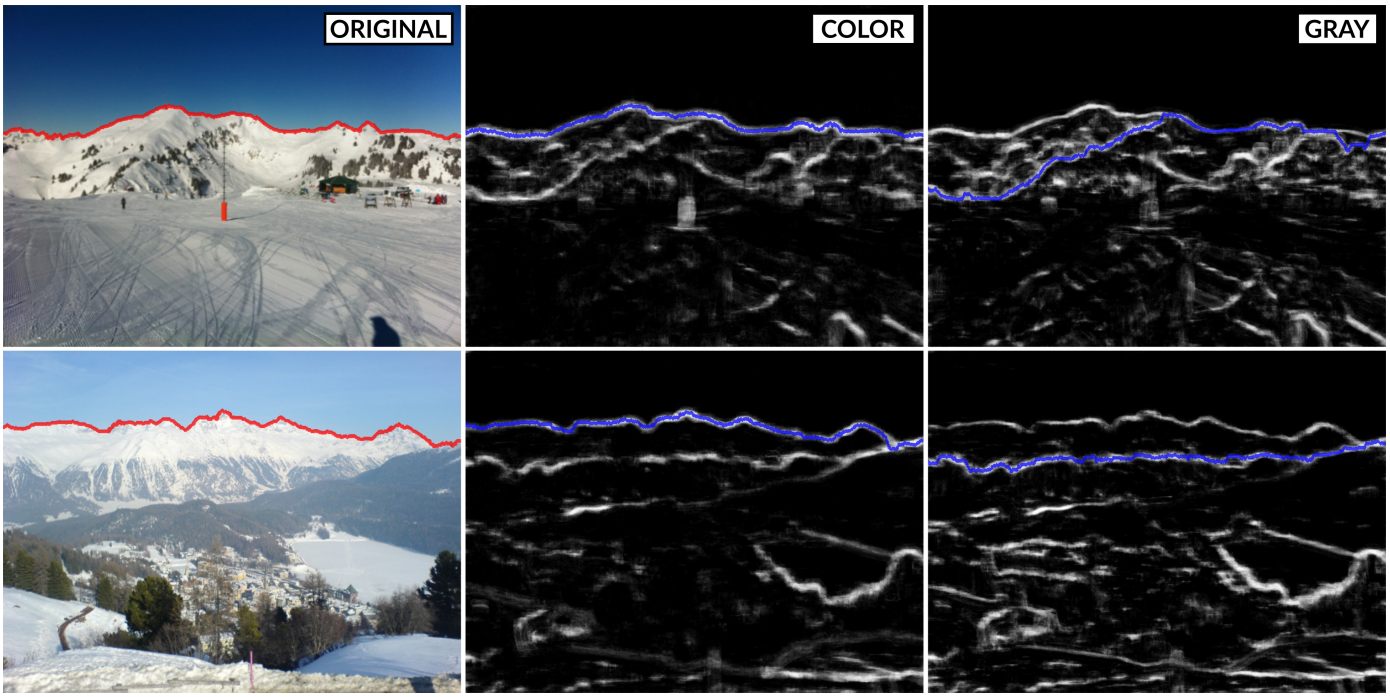


Figure 6. Detected horizon using color (**middle**) and grayscale (**right column**) images as the basis for the region covariance. (**Left column**) shows the original image and the ground truth horizon. In the middle and right column, the derived probability maps are shown where white indicates high horizon probability.

In the example shown in the top row of Figure 6, the horizon is correctly detected using cov_gray in the right part of the image. Rock faces, trees, and the skiing slope in the foreground visually form an apparent ridge line in the left part of the image. Without color information, the shortest path follows along these structures instead of the true horizon. In the second example (bottom row) the situation is similar. The tree line visible along the photographed mountain range is falsely detected as the horizon using cov_gray. As it can be seen in the probability map for cov_gray, multiple smaller parts of the true horizon are not correctly classified by the classifier. These gaps along the horizon result in graph nodes having high costs. Hence, the sum of edge weights along the true horizon is higher compared to the path along the tree line.

4.1.2. Influence of the Neighborhood DEFINITION

Refs. [14,16] employed two different neighborhood definitions within their shortest path calculations as graphically illustrated in Figure 4. While in [14] only pixels in consecutive columns and rows are connected, in [16] also pixels in consecutive columns which are separated by up to 3 rows are connected. According to [16] this should account for steeper mountain slopes encountered in the CH1 dataset.

For all approaches the results achieved based on the N3 neighborhood are better compared to N1 (Table 2). These results agree with the argumentation of [16] regarding the choice of N3.

Table 2. Influence of the used neighborhood.

Neighborhood	Method	Patch Size [px]	Error [px]	
			μ	σ
N1	cov_gray	17 × 17	4.26	13.08
N3	**cov_gray**	**17 × 17**	**3.51**	**11.82**
N1	cov_color	17 × 17	2.88	7.27
N3	**cov_color**	**17 × 17**	**2.78**	**7.29**
N1	SVM	17 × 17	26.99	58.27
N3	**SVM**	**17 × 17**	**22.57**	**49.47**

4.1.3. Influence of Step-Size

All images in the CH1 dataset have a resolution of ~1024 × 700 px. In general, it can be assumed that patches extracted from neighboring pixels will not differ greatly as they represent approximately the same image region. Accordingly, the region covariance of these patches will be very similar. With regard to the HIST dataset, where images with resolutions up to ~2500 × 1800 are available, the question arose, if it is really necessary to extract patches densely for each pixel or more sparsely in order to speed up computations and reduce memory requirements. Therefore, the accuracy of the detected horizon for the test images classifying each pixel, every 2nd, 4th, and 8th, were compared (Table 3).

Table 3. Influence of the step size on the detected horizon.

Step	Neighborhood	Method	Patch Size [px]	Error [px]	
				μ	σ
1	N3	cov_gray	17 × 17	3.51	11.82
2	N3	cov_gray	17 × 17	3.66	11.92
4	N3	cov_gray	17 × 17	4.04	11.79
8	N3	cov_gray	17 × 17	10.24	22.94

The accuracy of the detected horizon is only marginally influenced by extracting and classifying patches of every second pixel, whereas the number of patches is already reduced to 1/4. Further increasing the step size up to 8, which in combination with the patch size

used leads to half overlapping patches, the error further grows up to ~10 px, which is three times larger as for the densely extracted patches. While the similar standard deviations for the step sizes 1, 2, and 4 reveal that only minor differences occur, the standard deviation for the step size of 8 is twice as large indicating gross errors in the detected horizon.

Further results of the horizon detection for more images of the CH1 are given in the Appendix A.

4.2. HIST

For the HIST dataset, the same evaluation strategy as for the CH1 dataset was used. The 250 manually labeled images were randomly split 5 times into 60% training and 40% testing data. Following the evaluation on the CH1 dataset, a step size of 2 is used to calculate the probability maps of the test images.

As for the CH1 dataset, our method achieves significantly better results compared to SVM (Table 4). In contrast to the results obtained for CH1, the N1 neighborhood results in higher accuracy for both methods. The idea behind using N3 as neighborhood definition was to allow horizon lines to follow steep mountain profiles. In case of the HIST dataset, this has the opposite effect as shown in (Figure 7). In situations where ridge lines close to the true horizon, formed, e.g., by rock faces, the increased neighborhood size allows the detected horizon to jump across smaller vertical gaps and follow along these apparent structures.

Table 4. Difference between ground truth and estimated horizon line for the HIST dataset.

Method	Neighborhood	Step	Patch Size [px]	Error [px]	
				μ	σ
cov_gray	**N1**	**2**	**17 × 17**	**26.96**	**116.78**
cov_gray	N3	2	17 × 17	41.48	146.11
SVM	N1	2	17 × 17	116.25	184.58
SVM	N3	2	17 × 17	122.88	183.85

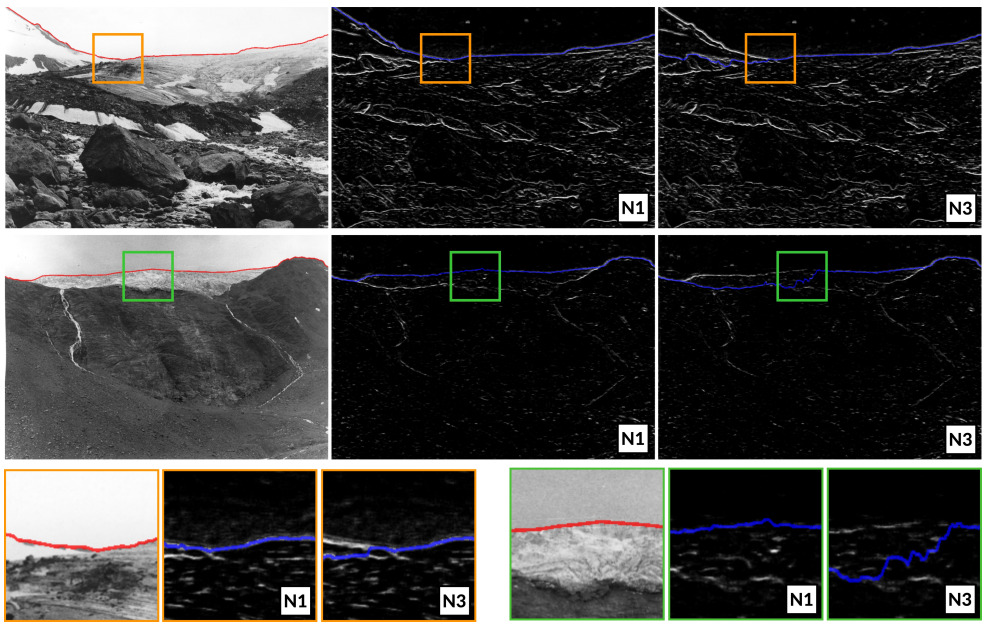


Figure 7. Connecting pixels based on the N3 neighborhood results in detected horizons which jump from the true horizon to ridges with strong visual edges below the true horizon. These ridge lines have high horizon probabilities, as they are often the result of a change from rock to snow or glacier. In contrast, the true horizon, also caused by the lack of color information, especially in case of snow coverage, has lower horizon probabilities or gaps.

Influence of the PATCH Size

The patch size controls the size of the region surrounding each pixel used within the calculation of the region covariance. While for the CH1 dataset, a patch size of 17 × 17 was used in accordance of the original implementation of SVM and for comparison with CNN, other dimensions might be more suitable. Therefore, additional patch sizes of 9 × 9 and 33 × 33 were evaluated (Table 5).

Table 5. Influence of patch size used for the region covariance calculation on the average error.

Method	Neighborhood	Step	Patch Size [px]	Error [px]	
				μ	σ
cov_gray	N1	2	9 × 9	43.91	132.36
cov_gray	N1	2	17 × 17	26.96	116.78
cov_gray	**N1**	**2**	**33 × 33**	**19.26**	**103.26**

The lowest average error and standard deviation is obtained for a patch dimension of 33 × 33 pixels. For smaller patch sizes, especially 9 × 9, smaller structures are classified with a high probability as potential horizon pixels leading to noisier probability maps. With increasing patch sizes and therefore consideration of larger regions surrounding each pixel, the influence of smaller structures is reduced. This leads to less possible paths within the probability maps for the shortest path to follow. In the example shown in Figure 8 the error of the detected horizon is reduced from ~156 px for the smallest patch size to ~9 px for the patch size of 33 × 33 pixel. The remaining error is caused due to the snow covered horizon in contrast to an apparent ridge line formed by rock faces below the true horizon (Figure 8-orange rectangle). Without any contrast between the true horizon and the cloudy sky in the background, cov_gray is not possible to detect this part of the horizon.

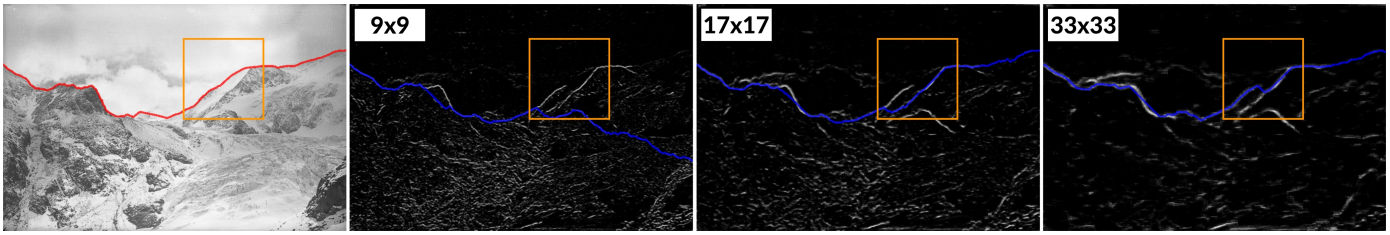


Figure 8. Detected horizon using color (**middle**) and grayscale (**right column**) images as the basis for the region covariance. (**Left column**) shows the original image and the ground truth horizon. In the middle and right column, the derived probability maps are shown, where white indicates high horizon probability.

Another benefit of the region covariance not addressed so far is directly related to the used patch size. While for SVM, the length of the feature vector used within classification is dependent on the patch size, the length of the feature vector for the region covariance, as shown in Section 2.1.2, is only dependent on the number of feature channels and number of regions used within each patch. Therefore, for all patch sizes used, the length of the feature vector is 135 for cov_gray, whereas for SVM, it ranges from 81 (9 × 9) to 1089 (33 × 33).

5. Discussion

For both datasets, cov_gray achieves the highest accuracy among the methods based on monochrome images. These results demonstrate that the region covariance is a powerful texture descriptor within the context of horizon detection. While for the CH1 dataset, an average error of 3.51 px was achieved, the error for the HIST dataset is roughly 6 times larger for the best performing variant of cov_gray. Analyzing the deviation from the ground truth horizon for each image (Table 6) for both datasets shows that the increased average error is mainly caused by a larger percentage (14.4%) of images having an error above 10 pixels, whereas for the CH1 dataset, there are only 6.6%. Additionally, the maximum

error for HIST (~1200 px) is approximately 10 times larger than for the CH1 (~180 px) dataset. This increased error is also related to the dimensions of the images contained in the respective dataset. While for CH1 all images have the same dimension (~1024 × 768 px), in the HIST dataset these range up to ~2500 × 1800 px.

Table 6. Percentage of images by difference from ground truth horizon in [px] for the best performing variants of cov_gray for both dataset.

Dataset	<1 [px]	1–2 [px]	2–3 [px]	3–4 [px]	4–5 [px]	5–10 [px]	>10 [px]
HIST	0.5	30.7	28.1	8.7	6.1	11.6	14.4
CH1	48.0	34.9	6.1	1.2	1.7	1.5	6.6

To further understand the remaining challenges, selected images from the HIST dataset with deviations larger than 10 pixels are shown in Figure 9.

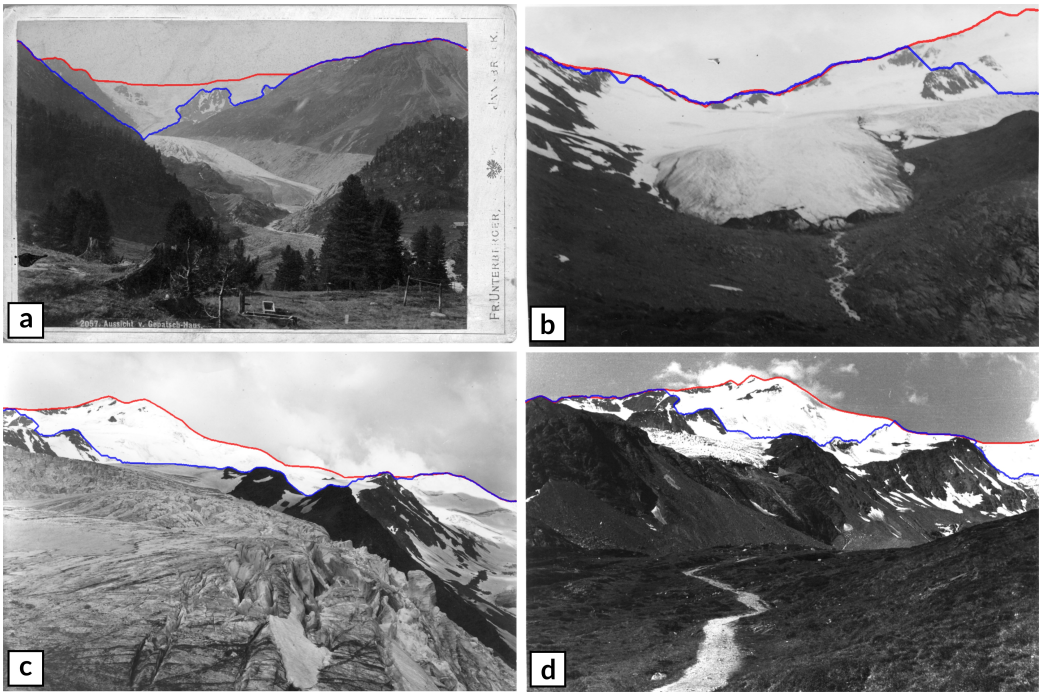


Figure 9. Selected images of the HIST dataset with errors greater than 10 pixels. Red indicates the ground truth horizon and blue the detected horizon. (**a**,**b**): Parts of the horizon are barely visible. (**c**,**d**): Paths along rock formations are interpreted as horizon.

Subfigures (a) and (b) show images where parts of the true horizon are barely distinguishable from the sky above due to the lack of color, low contrast, and reduced image quality. In those regions, where the horizon is visible, it is also correctly detected. Due to the formulation of the horizon detection as a shortest path problem from the left to the right image border, a continuous horizon for the whole image is enforced. Therefore, in those cases where parts are not visible, the next best path is chosen within the shortest path.

In (c) and (d), the situation is slightly different and similar to the example shown in Figure 6. While the horizon is visible, it is completely covered in snow. Due to the visually apparent ridge lines formed by rock faces below the horizon, the detected horizon partly follows along these structures. These situations might be solved if color was available, as was the case for the CH1 dataset.

Nevertheless, with the developed method, it is possible to overcome major problems due to challenging situations posed by historical terrestrial images captured in high alpine terrain (Figure 10).

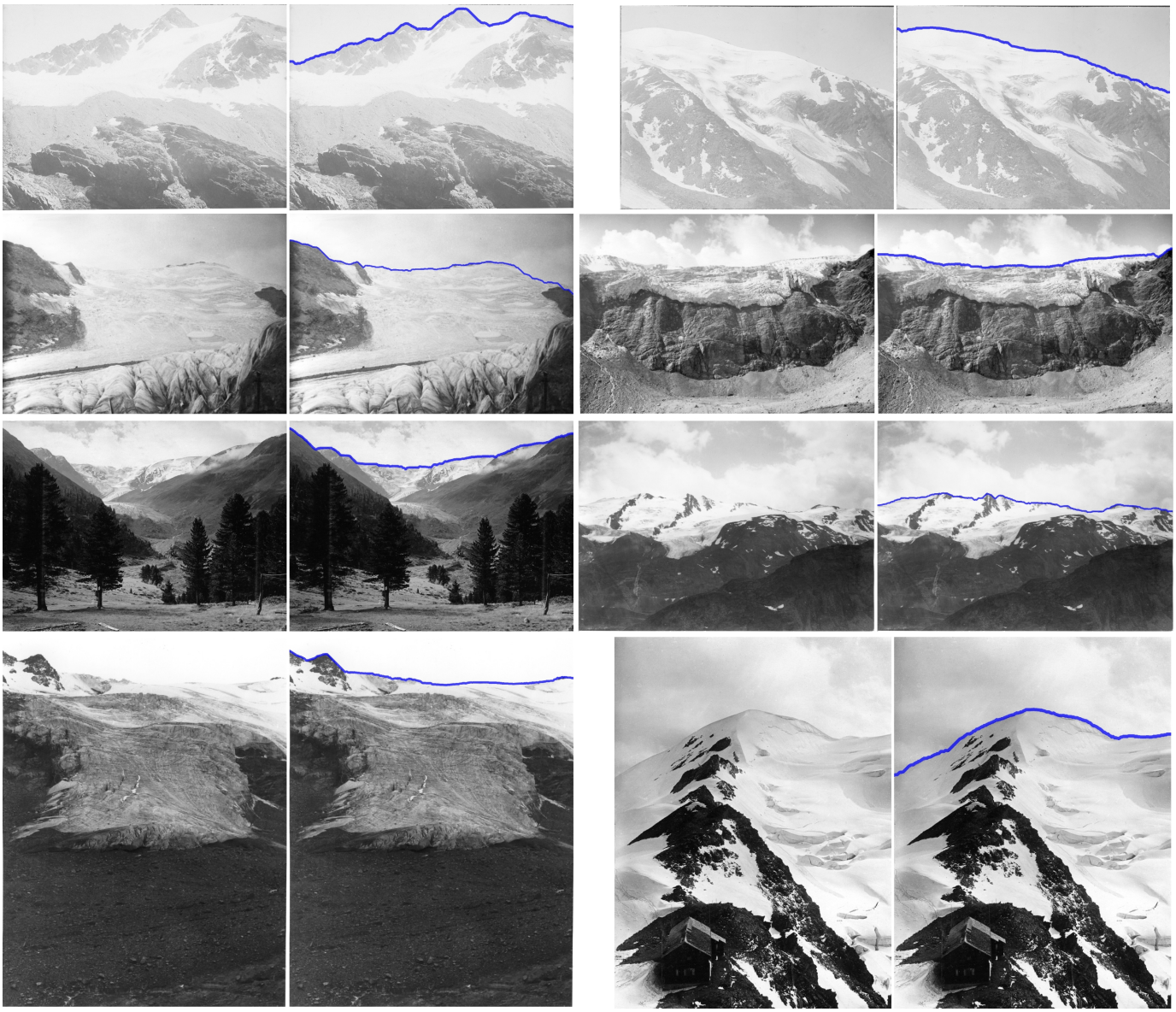


Figure 10. Successfully detected horizons using cov_gray despite challenging situations. (**Top row**): low contrast images. (**Second row**): Horizons covered with glacier. (**Third row**): Barely visible horizons due to snow and clouds. (**Fourth row**): Horizon barely visible due to glacier (left) and snow (right).

More results of the horizon detection for images of the HIST dataset, including both successful and unsuccessful examples, are given in the Appendix B.

6. Conclusions

Based on the region covariance as texture descriptor, an improved method for the detection of horizon lines in monochrome images was presented. Especially comparison with [14] on the CH1 dataset shows the discriminative power of the region covariance matrix. Only approaches using color information achieve higher accuracy, indicating that color is an important additional information regarding the horizon line detection in mountainous terrain. With the HIST dataset, the developed method was evaluated on historical images in average dating back to 1950. Particularly, the combination of high alpine terrain covered with snow and glaciers and the reduced image quality pose extremely challenging situations for horizon line detection. Despite those, in 85% of the images the horizon was detected with an average error smaller than 10 pixels.

Author Contributions: Conceptualization, S.M.-F. and N.P.; methodology, S.M.-F. and N.P.; software, S.M.-F.; validation, S.M.-F. and N.P.; writing—original draft preparation, S.M.-F. and N.P; supervision, N.P.; funding acquisition, N.P.; All authors contributed to the discussion and the editing of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work is part of the SEHAG project (project number I 4062) funded by the Austrian Science Fund (FWF). Open Access Funding by the Austrian Science Fund (FWF).

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

Appendix A. Detected Horizon Lines on the CH1 Dataset

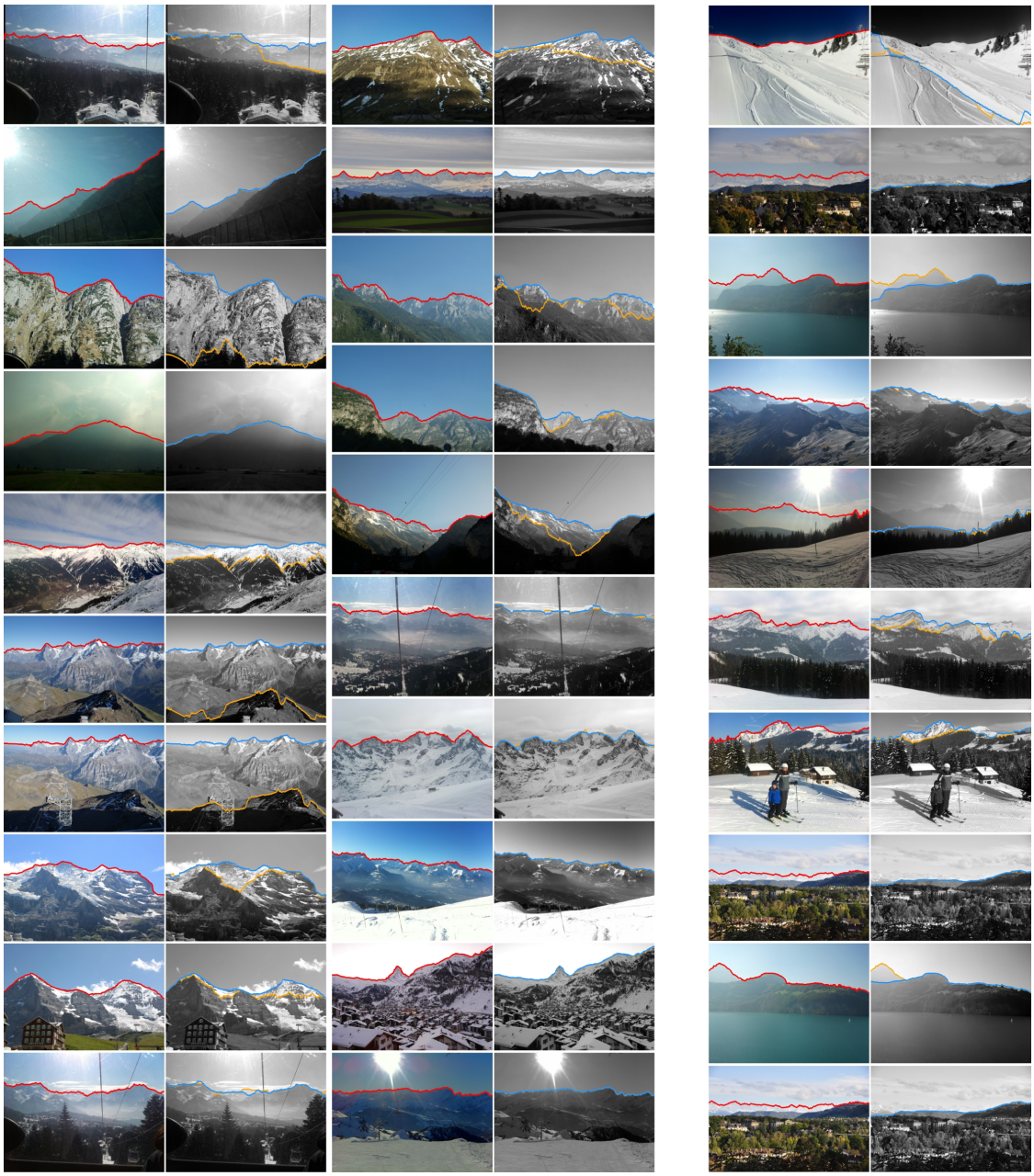


Figure A1. Selected images from the CH1 dataset with ground truth horizons (red). The results using SVM (orange) and cov_gray (blue) are shown in the grayscale images. Left and center column show correctly detected horizons using cov_gray, right column horizons which are erroneous.

Appendix B. Detected Horizon Lines on the HIST Dataset

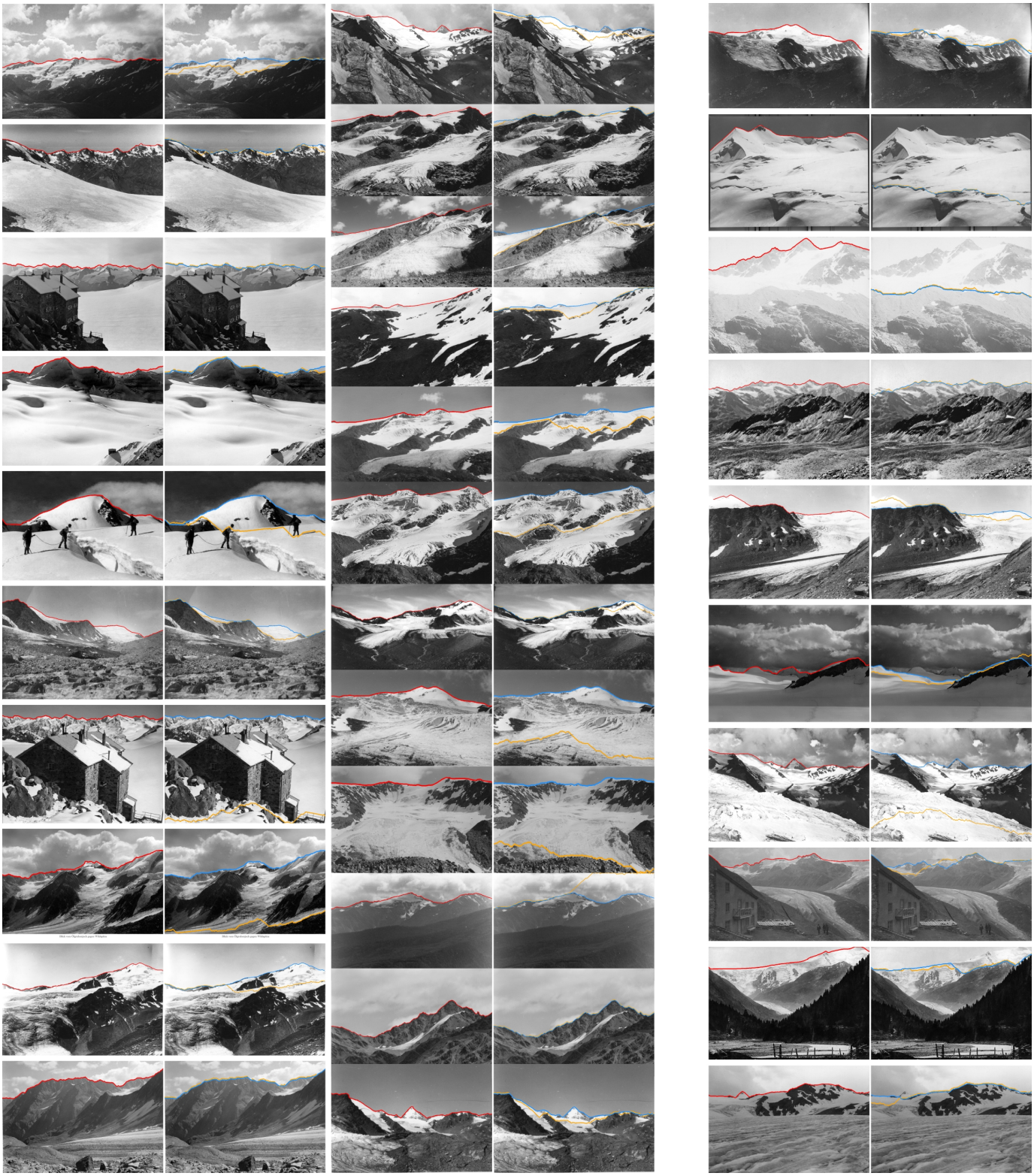


Figure A2. Selected images from the HIST dataset with ground truth horizons (red) and detected horizons using SVM (orange) and cov_gray (blue). Left and center column show correctly detected horizons using cov_gray, right column horizons which are erroneous.

References

1. Bozzini, C.; Conedera, M.; Krebs, P. A New Monoplotting Tool to Extract Georeferenced Vector Data and Orthorectified Raster Data from Oblique Non-Metric Photographs. *Int. J. Herit. Digit. Era* **2012**, *1*, 499–518. [CrossRef]
2. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
3. Bay, H.; Tuytelaars, T.; Van Gool, L. SURF: Speeded Up Robust Features. In Proceedings of the Computer Vision—ECCV 2006, Graz, Austria, 7–13 May 2006; pp. 404–417. [CrossRef]
4. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2564–2571.
5. Produit, T.; Tuia, D.; Golay, F.; Strecha, C. Pose estimation of landscape images using DEM and orthophotos. In Proceedings of the 2012 International Conference on Computer Vision in Remote Sensing, Xiamen, China, 16–18 December 2012; pp. 209–214. [CrossRef]
6. Gat, C.; Albu, A.B.; German, D.; Higgs, E. A Comparative Evaluation of Feature Detectors on Historic Repeat Photography. In *Advances in Visual Computing*; Bebis, G., Boyle, R., Parvin, B., Koracin, D., Wang, S., Kyungnam, K., Benes, B., Moreland, K., Borst, C., DiVerdi, S., et al., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2011; pp. 701–714. [CrossRef]
7. Baatz, G.; Saurer, O.; Köser, K.; Pollefeys, M. Large Scale Visual Geo-Localization of Images in Mountainous Terrain. In Proceedings of the Computer Vision—ECCV 2012, Florence, Italy, 7–13 October 2012; pp. 517–530. [CrossRef]
8. Baboud, L.; Čadík, M.; Eisemann, E.; Seidel, H. Automatic photo-to-terrain alignment for the annotation of mountain pictures. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 41–48.
9. Cozman, F.; Krotkov, E. Position estimation from outdoor visual landmarks for teleoperation of lunar rovers. In Proceedings of the Third IEEE Workshop on Applications of Computer Vision, WACV'96, Sarasota, FL, USA, 2–4 December 1996; pp. 156–161. [CrossRef]
10. Naval, P.; Mukunoki, M.; Minoh, M.; Ikeda, K. Estimating Camera Position and Orientation from Geographical Map and Mountain Image. In *38th Research Meeting of the Pattern Sensing Group, Society of Instrument and Control Engineers*; Citeseer: Tokyo, Japan, 1997.
11. Stein, F.; Medioni, G. Map-based localization using the panoramic horizon. *IEEE Trans. Robot. Autom.* **1995**, *11*, 892–896. [CrossRef]
12. Saurer, O.; Baatz, G.; Köser, K.; Ladický, L.; Pollefeys, M. Image Based Geo-localization in the Alps. *Int. J. Comput. Vis.* **2016**, *116*, 213–225. [CrossRef]
13. Lie, W.N.; Lin, T.C.I.; Lin, T.C.; Hung, K.S. A robust dynamic programming algorithm to extract skyline in images for navigation. *Pattern Recognit. Lett.* **2005**, *26*, 221–230. [CrossRef]
14. Ahmad, T.; Bebis, G.; Nicolescu, M.; Nefian, A.; Fong, T. An Edge-Less Approach to Horizon Line Detection. In Proceedings of the 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 9–11 December 2015; pp. 1095–1102. [CrossRef]
15. Frajberg, D.; Fraternali, P.; Torres, R.N. Convolutional Neural Network for Pixel-Wise Skyline Detection. In Proceedings of the Artificial Neural Networks and Machine Learning—ICANN, Alghero, Italy, 11–14 September 2017; pp. 12–20. [CrossRef]
16. Porzi, L.; Rota Bulò, S.; Ricci, E. A Deeply-Supervised Deconvolutional Network for Horizon Line Detection. In Proceedings of the 24th ACM International Conference on Multimedia, Amsterdam, The Netherlands, 15–19 October 2016; [CrossRef]
17. Shen, Y.; Wang, Q. Sky Region Detection in a Single Image for Autonomous Ground Robot Navigation. *Int. J. Adv. Robot. Syst.* **2013**, *10*, 362. [CrossRef]
18. Tuzel, O.; Porikli, F.; Meer, P. Region Covariance: A Fast Descriptor for Detection and Classification. In Proceedings of the Computer Vision—ECCV 2006, Graz, Austria, 7–13 May 2006; pp. 589–600. [CrossRef]
19. Ojala, T.; Pietikäinen, M.; Mäenpää, T. Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *Lect. Notes Comput. Sci.* **2000**, *1842*, 404–420. [CrossRef]
20. Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural Features for Image Classification. *IEEE Trans. Syst. Man, Cybern.* **1973**, *SMC-3*, 610–621. [CrossRef]
21. Leung, T.; Malik, J. Representing and Recognizing the Visual Appearance of Materials using Three-dimensional Textons. *Int. J. Comput. Vis.* **2001**, *43*, 29–44. [CrossRef]
22. Porikli, F.; Tuzel, O. Fast Construction of Covariance Matrices for Arbitrary Size Image Windows. In Proceedings of the 2006 International Conference on Image Processing, Atlanta, GA, USA, 8–11 October 2006; pp. 1581–1584.
23. Förstner, W.; Moonen, B. A Metric for Covariance Matrices. In *Geodesy-The Challenge of the 3rd Millennium*; Grafarend, E.W., Krumm, F.W., Schwarze, V.S., Eds.; Springer: Berlin/Heidelberg, Germany, 2003; pp. 299–309. [CrossRef]
24. Barachant, A.; Bonnet, S.; Congedo, M.; Jutten, C. Classification of covariance matrices using a Riemannian-based kernel for BCI applications. *Neurocomputing* **2013**, *112*, 172–178. [CrossRef]
25. Guan, D.; Xiang, D.; Tang, X.; Wang, L.; Kuang, G. Covariance of Textural Features: A New Feature Descriptor for SAR Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3932–3942. [CrossRef]
26. Arsigny, V.; Fillard, P.; Pennec, X.; Ayache, N. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magn. Reson. Med.* **2006**, *56*, 411–421.

27. Li, X.; Hu, W.; Zhang, Z.; Zhang, X.; Zhu, M.; Cheng, J. Visual tracking via incremental Log-Euclidean Riemannian subspace learning. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska, 23–28 June 2008; pp. 1–8.
28. Kluckner, S.; Mauthner, T.; Roth, P.; Bischof, H. Semantic Image Classification using Consistent Regions and Individual Context. In Proceedings of the British Machine Vision Conference 2009 (BMVC), London, UK, 7–10 September 2009.
29. Li, P.; Wang, Q. Local Log-Euclidean Covariance Matrix (L2ECM) for Image Representation and Its Applications. In Proceedings of the Computer Vision—ECCV 2012, Florence, Italy, 7–13 October 2012; pp. 469–482. [CrossRef]
30. Ahmad, T.; Campr, P.; Čadik, M.; Bebis, G. Comparison of semantic segmentation approaches for horizon/sky line detection. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 4436–4443.
31. Ahmad, T.; Bebis, G.; Nicolescu, M.; Nefian, A.; Fong, T. Fusion of edge-less and edge-based approaches for horizon line detection. In Proceedings of the 2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA), Corfu, Greece, 6–8 July 2015; pp. 1–6. [CrossRef]
32. Dollár, P.; Zitnick, C.L. Fast Edge Detection Using Structured Forests. *arXiv* **2014**, arXiv:1406.5549.
33. Porzi, L.; Buló, S.R.; Valigi, P.; Lanz, O.; Ricci, E. Learning Contours for Automatic Annotations of Mountains Pictures on a Smartphone. In Proceedings of the International Conference on Distributed Smart Cameras, ICDSC '14, Venezia Mestre, Italy, 4–7 November 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 1–6. [CrossRef]