

Article

## Optimizing SIFT for Matching of Short Wave Infrared and Visible Wavelength Images

Aleksandra A. Sima <sup>1,2,\*</sup> and Simon J. Buckley <sup>1</sup>

<sup>1</sup> Centre for Integrated Petroleum Research (Uni CIPR), P.O. Box 7810, N-5020 Bergen, Norway; E-Mail: simon.buckley@uni.no

<sup>2</sup> Department of Informatics, University of Bergen, P.O. Box 7803, N-5020 Bergen, Norway

\* Author to whom correspondence should be addressed; E-Mail: aleksandra.sima@uni.no; Tel.: +47-5558-3673; Fax: +47-5558-8265.

Received: 13 February 2013; in revised form: 12 April 2013 / Accepted: 15 April 2013 /

Published: 24 April 2013

---

**Abstract:** The scale invariant feature transform (SIFT) is a widely used interest operator for supporting tasks such as 3D matching, 3D scene reconstruction, panorama stitching, image registration and motion tracking. Although SIFT is reported to be robust to disparate radiometric and geometric conditions in visible light imagery, using the default input parameters does not yield satisfactory results when matching imagery acquired at non-overlapping wavelengths. In this paper, optimization of the SIFT parameters for matching multi-wavelength image sets is documented. In order to integrate hyperspectral panoramic images with reference imagery and 3D data, corresponding points were required between visible light and short wave infrared images, each acquired from a slightly different position and with different resolutions and geometric projections. The default SIFT parameters resulted in too few points being found, requiring the influence of five key parameters on the number of matched points to be explored using statistical techniques. Results are discussed for two geological datasets. Using the SIFT operator with optimized parameters and an additional outlier elimination method, allowed between four and 22 times more homologous points to be found with improved image point distributions, than using the default parameter values recommended in the literature.

**Keywords:** SIFT; optimization; SWIR; VIS; multi-wavelength; hyperspectral

---

## 1. Introduction

The broad area of automatic feature extraction methods in photogrammetry and remote sensing is constantly expanding in line with developments of new sensors and data acquisition techniques, and has driven research into fast, robust and reliable interest operators for aiding tasks such as image orientation, generation of digital surface models, 3D reconstruction and motion tracking. Although the principle of feature extraction is similar for all interest operators—*i.e.*, extraction and description of salient image features in such a way that homologous points identified in several images can be matched—robustness of interest operators to different radiometric and geometric image transformation varies. The success of automatic feature extraction relies on choosing an interest operator for a specific task and the characteristics of the input data [1]. For example, traditional corner detectors such as the Harris operator [2] do not handle different image scales well [3], FAST [4] and BRIEF [5] features are not rotationally invariant [6], and the Maximally Stable Extremal Regions (MSER) algorithm [7] performs better in high contrast image regions [8].

Although several newer matching approaches have been recently developed, e.g., ORB [6] or FREAK [9], the scale invariant feature transform (SIFT) [10] is one of the most established interest operators in current use. The robustness of SIFT to difficult geometric and radiometric conditions [11–13] makes it one of the most frequently employed interest operators, and therefore forms the focus of this paper. In photogrammetry it has been mostly used for image registration [3,13–16] and in computer vision for such tasks as 3D matching [17], 3D scene reconstruction [18], panorama stitching [19] and motion tracking [20].

The algorithm is frequently applied with the input parameters suggested by Lowe [21], without specifically tuning them for the task and data characteristics. Although in many cases using these default parameters brings satisfactory matching results, there are situations where it does not necessarily provide enough candidate points. In fact, for matching datasets acquired in different ranges of the electromagnetic spectrum, extensions to SIFT have been proposed such as scale restriction criteria [22] or descriptor modification [23] to increase the number of matched points, though the authors do not report attempts to evaluate or optimize the SIFT parameters. Modification of the SIFT parameter values has allowed for successful extraction and matching of homologous points in multi-wavelength datasets, such as intensity images from a near infrared (NIR) terrestrial laser scanner and the visible-light images (VIS) [3], or hyperspectral short wave infrared (SWIR) and VIS images [15]. With an expanding range of imaging sensors available, and an increasing focus on sensor integration and data fusion [24], the need for registration and integration with other 2D and 3D datasets provides new opportunities for many interdisciplinary studies, making the applicability of SIFT to multi-wavelength data especially relevant.

Hyperspectral imaging is an established technique that allows for mapping and quantification of materials indistinguishable to the naked eye [25], and is now being applied using terrestrial or close range sensors with encouraging results [14,26,27]. The high spectral resolution of hyperspectral imagery allows for detection of subtle geochemical differences [28] and quantitative analysis of pixel composition [29], especially attractive in geoscience, particularly geological applications. Although results of hyperspectral classifications are valuable as a standalone product, when integrated with high spatial resolution 3D surface models and image data, in the form of photorealistic models, they can

provide additional information for analyzing and interpreting mineral distribution [27]. Photorealistic 3D models are triangle meshes texture mapped with (normally VIS) digital images of the measured object. Registration of a hyperspectral image into the 3D model coordinate system enables the possibility to texture the terrain model with the hyperspectral products, assuming that the hyperspectral camera model is known. The integration between the hyperspectral imagery and the dataset used to create the photorealistic model can be solved using manual control point measurement [30], or semi-automated methods, as in [10,11]. The latter paper forms the motivation for extension in the current work. The authors presented an integration method relying on using SIFT to find corresponding points between the hyperspectral SWIR data and VIS images pre-registered in the 3D model coordinate system. Because the success of the registration depends on a suitable number and distribution of homologous points between the two image types, the influence of SIFT parameters was found to be important [11].

The main contribution of this paper is documentation of SIFT parameter optimization to maximize the number of correct matches between multi-wavelength imagery, *i.e.*, SWIR data acquired with a hyperspectral HySpex SWIR-320m camera (Norsk Elektro Optikk AS, Oslo, Norway) and conventional digital imagery (VIS) collected using a Nikon D200 camera. The two image types were acquired not only in different spectral ranges, but also with different spatial resolutions, geometric projections (cylindrical panoramic *vs.* central perspective) and locations of image capture. In contrast to [15] where one SIFT parameter was tuned and a single hyperspectral band was used in matching, in this paper five operator parameters are adjusted and multiple spectral bands are used in order to increase the number of matched points. The influence of the analyzed parameters on the matching results is presented and discussed in detail on imagery collected for geological purposes in two different locations. Although inspired by a practical need, the tuning reflects a lack in the literature and is also relevant to other applications.

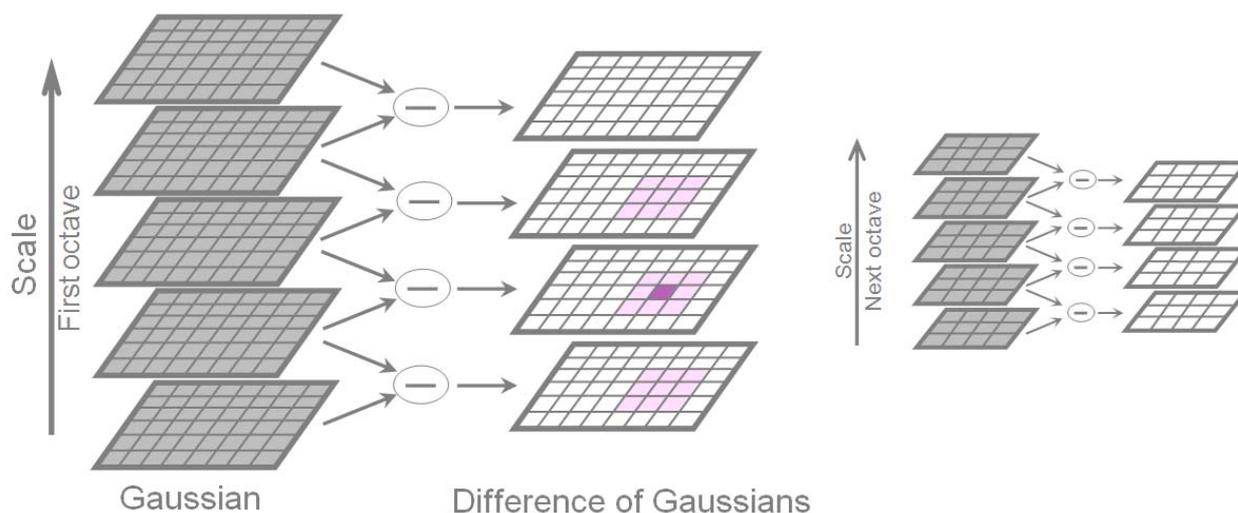
The paper is organized as follows: Section 2 presents the background on how SIFT interest points are extracted, described and matched, with focus on the specific parameters controlling the algorithm. An overview of different SIFT parameter adjustment approaches is given in Section 3, followed by the characteristics of the study datasets in Section 4. Section 5 defines the range of values for parameters to be considered, and the analysis approach. Final results, including the set of SIFT parameters optimized for the study datasets, are described and discussed in Section 6. Section 7 summarizes with conclusions and suggestions for future work.

## 2. Background

The SIFT interest point creation procedure is divided into four stages: extrema detection, keypoint localization, orientation assignment and keypoint descriptor extraction [21]. In the first step, a Difference of Gaussians (DoG) (Figure 1) is created and used to identify the potential interest points (keypoints). The input image is repeatedly Gaussian blurred to produce a set of images, termed an octave. The neighboring images in the set are used to compute the Difference of Gaussians (Figure 1, left). To create the next octave (Figure 1, right) each Gaussian blurred image is downsampled by a factor of two and the process is repeated [21]. The extrema in the DoG images are detected by comparing a pixel value to its 26 neighbors in  $3 \times 3$  regions at the current and adjacent scales (purple

in Figure 1). Three main parameters control this stage: the number of octaves, the initial Gaussian smoothing for the first level of each octave, defined by the *sigma* parameter, and the number of scales sampled in each octave (*nScales*).

**Figure 1.** Creation of octaves and Difference of Gaussians (adapted from [21]).



Next, the candidate interest points are located with sub-pixel accuracy by fitting a Taylor Expansion function to the local pixel values. At this stage potentially unstable points in low contrast areas are eliminated if the function value at the extreme values is lower than a defined minimum contrast difference, expressed by the value of *contrast threshold* (*CT*) parameter. Similarly, unstable keypoints located along edges are eliminated by defining an *edge threshold*.

In the third stage, a principal orientation is assigned to each keypoint location based on local image gradient directions. Transformation of the image data relative to the assigned orientation, scale, and location for each feature provides invariance to these geometric influences in all future operations.

Finally, for every keypoint, a set of eight-bin histograms are created for a  $4 \times 4$  pixel neighborhood. The resulting feature descriptor is a vector of 128 elements that is normalized to unit length to handle illumination differences. Details about the further nine parameters controlling the last two steps can be found in [10,31].

After the SIFT interest points are extracted from and described in two overlapping images, the keypoint descriptors need to be matched. The best candidate match for each keypoint on the first image is found by identifying its nearest neighbor in the database of keypoints on the second image based on Euclidean distance between the keypoint descriptors. In order to decrease the risk of incorrect keypoint matching, highly similar candidate matches are eliminated. This is realized by rejecting all matches where the ratio between the distances of the closest neighbor to that of the second-closest neighbor is larger than the value of the *nnRatio* parameter.

### 3. Related Work

Previous work to improve SIFT performance by parameter adjustment is limited. Numerous authors report changing the values of one or two parameters without reporting the means of reaching the used value [3,11,32,33]. In computer vision, several SIFT parameter optimizations have been reported, with

the work by May *et al.* [31,34] being the most extensive and covering the largest number of algorithm parameters. The efficiency of the SIFT algorithm was tested with different values of eight (out of the total of 17) parameters on several pairs of images, with results analyzed using information visualization techniques such as brushing, parallel coordinates, scatter plots and histograms. It was concluded that the number of resulting matches depends largely on the image content as well as on values of the initial Gaussian blur (sigma) and the contrast threshold while the other parameters are more robust to variations. Maestas *et al.* [35] proposed using a design of experiments to find the optimal values of six SIFT parameters. The authors focused on presentation of the method principle rather than discussion of results, with one image pair only and few parameters values assessed. In the biometrics field, several values of the contrast threshold were explored in the performance of SIFT for iris recognition purposes [36]. In photogrammetry, Lingua *et al.* [13] used SIFT for orientating images acquired with a low-cost unmanned aerial vehicle (UAV). Those authors proposed an auto-adaptive method to adjust the CT parameter in order to get more matches between images with homogenous texture (low-textured). In addition, the nearest neighbor ratio parameters were modified and better results were obtained than with the default SIFT values.

#### 4. Image Dataset Characteristics

The data used in this study were acquired with two systems, a tripod-mounted hyperspectral HySpex SWIR-320m camera, and a Nikon D200 camera mounted on the top of a terrestrial laser scanner acquiring the 3D terrain geometry. In order to reduce differences caused by the viewing angle of the two sensors, the two instruments were located next to each other in the field. Datasets were acquired in two locations: in the Pozalagua quarry (Cantabria, Spain) and at the TT Niche, a tunnel face in the Mont Terri underground laboratory, Jura Mountains, Switzerland. Both were collected for geological purposes, to map the distribution of different sedimentary materials (Spain) and to characterize variability in clay-rich rock (Switzerland).

##### 4.1. SWIR Hyperspectral Imagery

The hyperspectral HySpex SWIR-320m camera is a portable terrestrial line scanner, with a 14° field of view across track (320 pixels). The spectral range from 1.3  $\mu\text{m}$  to 2.5  $\mu\text{m}$  is covered by 241 bands with a sampling interval of 5 nm. The system is mounted on a tripod and uses a rotation stage to construct the image in the along-track direction. The resulting cylindrical imaging geometry can be successfully represented by a geometric model for panoramic cameras [37], allowing precise registration of the hyperspectral and conventional image data, as described in [30]. Detailed geometrical and spectral characteristics of the HySpex SWIR-320m are presented in [30].

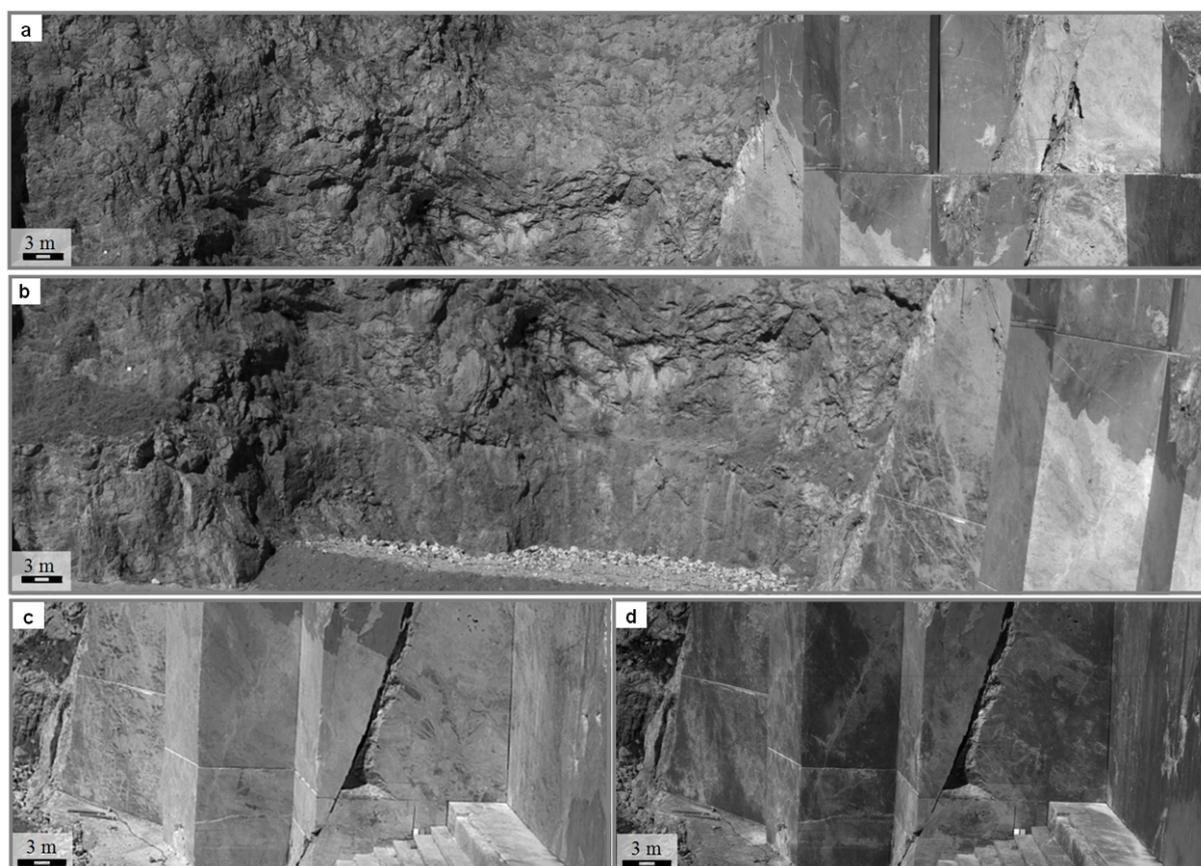
From each study area, three hyperspectral images were used and are characterized in Table 1. The spectral data recorded by the sensor were converted into at-sensor radiances, according to the spectral calibration report provided by the sensor manufacturer. Bands containing only random noise (wavelengths completely absorbed by the atmosphere in the outdoor Pozalagua dataset) were eliminated.

**Table 1.** Characteristics of SWIR and VIS images used for experimentation.

Location	HySpex SWIR-320m				Nikon D200	
	Image	Number of Columns	Object Sampling Resolution (m)	Number of Bands Used	Covering Photos	Object Sampling Resolution (m)
Pozalagua quarry, Cantabria, Spain	P-B1	1,200	0.075	205	5	0.007
	P-B2	1,500	0.068	211	6	0.007
	P-C3	700	0.049	204	3	0.005
TT Niche tunnel, Mont Terri, Switzerland	N-A1	510	0.010	241	2	0.002
	N-A2	510	0.010	241	2	0.002
	N-H1	1,600	0.004	241	1	0.002

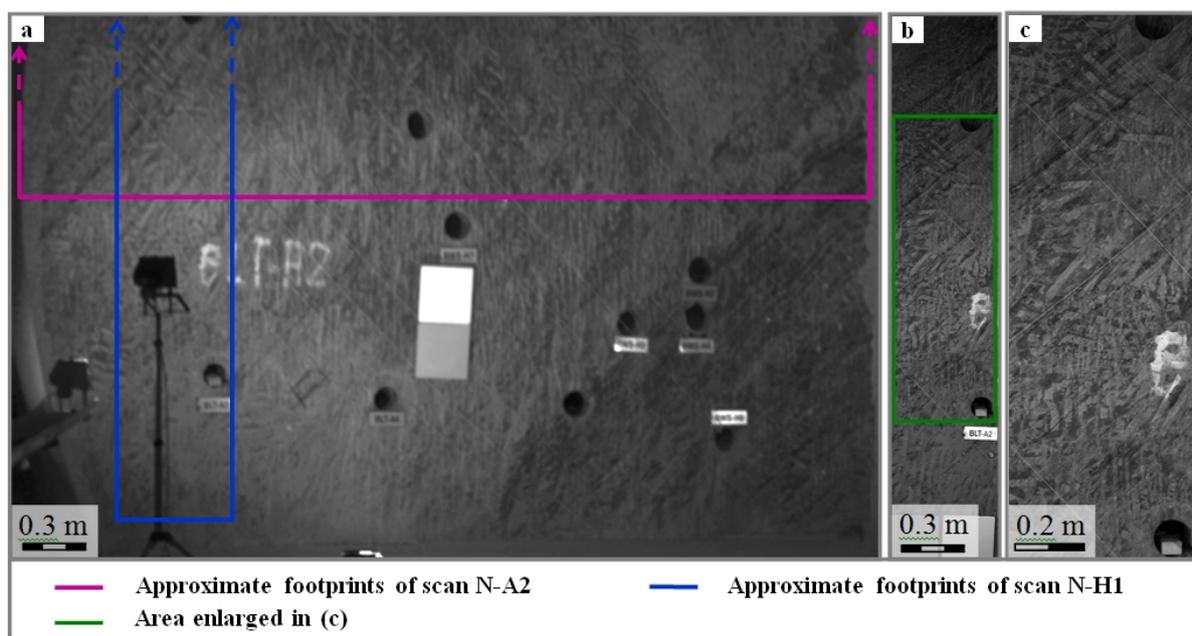
The selected hyperspectral imagery represented different objects, conditions and difficulties for matching. In the dataset from Spain, images P-B1 and P-B2 covered mostly (60%–80%) weathered rock surface areas, partially with vegetation (Figure 2(a,b)), while the remaining area represented low-textured fresh-cut rock surfaces. This proportion was different in image P-C3, where fresh-cut rock surfaces, very challenging for the SIFT interest operator, comprised around 90% of the image area (Figure 2(c,d)). The different spectral appearance of hyperspectral bands can be observed in Figure 2(c,d).

**Figure 2.** (a) Hyperspectral image P-B2, band 1: 1.336  $\mu\text{m}$ ; (b) Image P-B1, band 1: 1.336  $\mu\text{m}$ ; (c) Band 62: 1.639  $\mu\text{m}$  and (d) band 225: 2.429  $\mu\text{m}$  of image P-C3.



The second dataset, acquired in the subsurface TT Niche tunnel, gave different challenges. The outcrop wall was smooth and covered with a pattern of surface gouging caused by the tunnel making process (Figure 3). The hyperspectral images were acquired at a different time than the digital photos, around six months apart, so that the illumination conditions were different (non-optimal artificial illumination that caused many specular reflections and shadows), and changes had occurred to the tunnel face itself (new identifying marks for geological study were added). Most changes appeared in the lower part of image N-A1 (Figure 3) and overlapping area of image N-A2. Hyperspectral image N-H1 was acquired from a much closer distance to the tunnel face and covered a smaller area with higher spatial resolution (Figure 3(b,c)).

**Figure 3.** (a) Hyperspectral image N-A1, band 62: 1.634  $\mu\text{m}$ ; (b) Band 62: 1.634  $\mu\text{m}$  of image N-H1; (c) Enlarged area marked in (b).



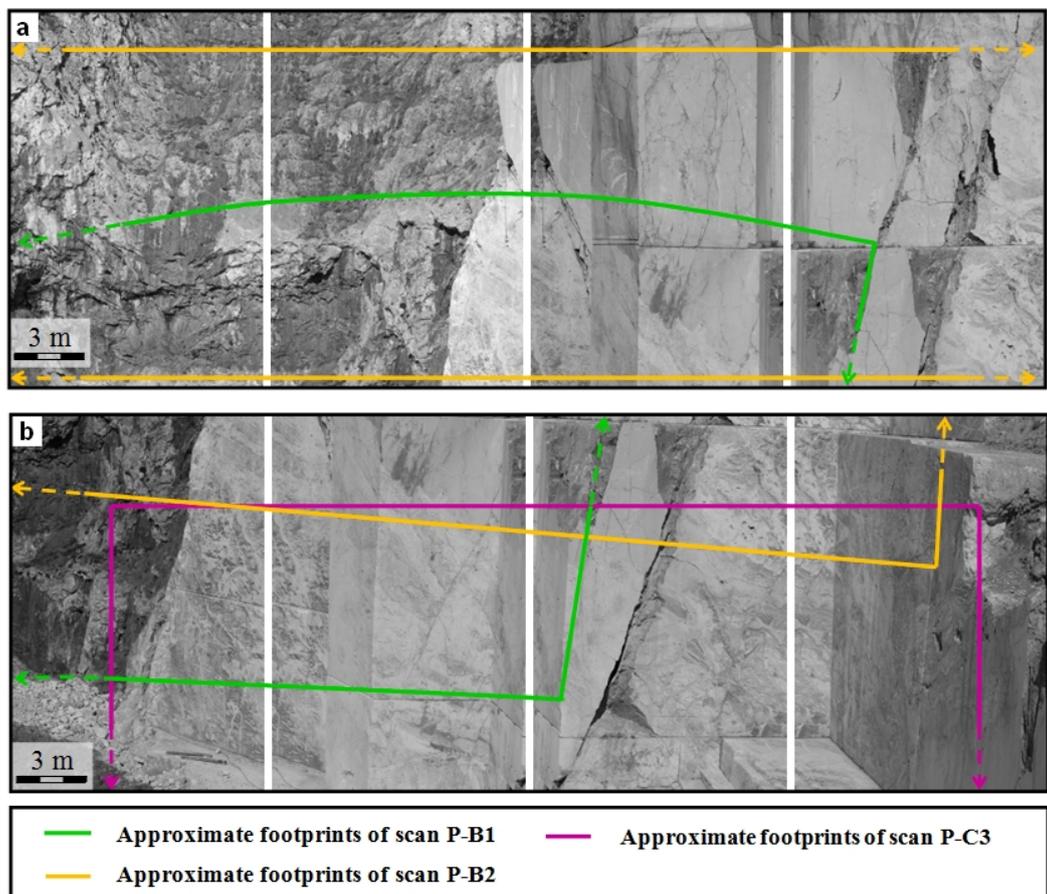
#### 4.2. VIS Imagery

The Nikon D200 is a 10.2 megapixel single-lens reflex (SLR) camera. Two calibrated lenses were used while acquiring data for this study—a Nikkor 85 mm lens (Polazagua), and a Nikkor 50 mm lens (TT Niche)—with focal lengths fixed for the duration of the data collection. Due to different sensor optics and lens field of views, several VIS images were required to cover the area covered by one hyperspectral image (Table 1, Figures 4 and 5). The much higher spatial resolution of the Nikon camera resulted in significantly different object sampling resolutions (Table 1).

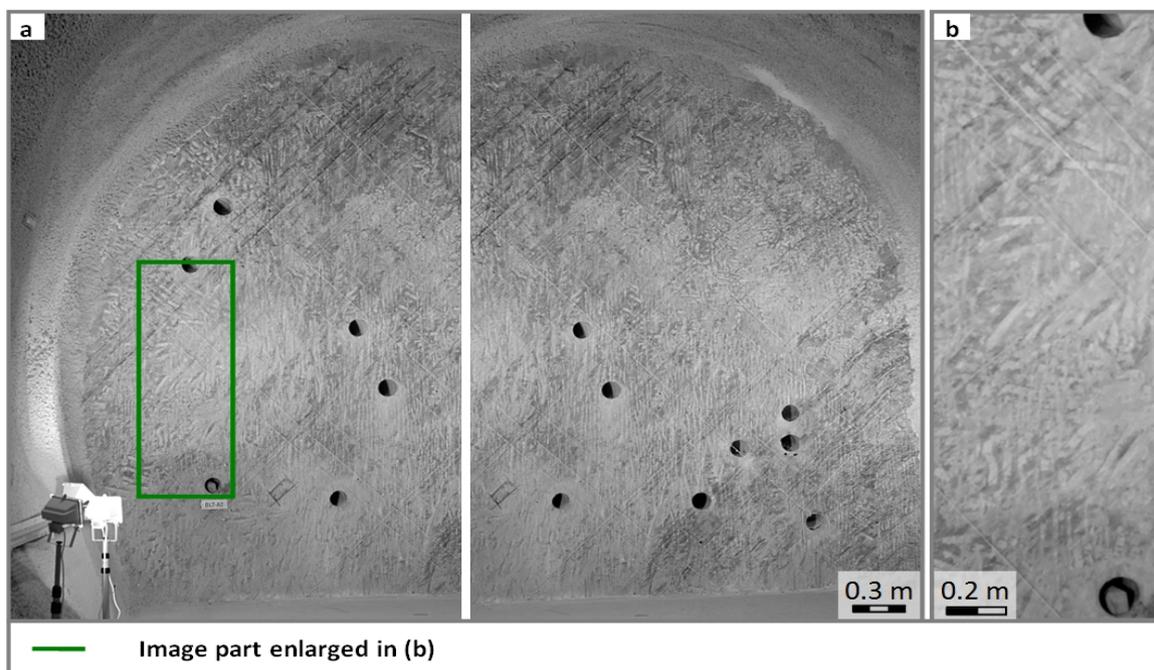
Because the SIFT detector is applied to a single image band, rather than using all VIS bands, for simplicity reasons, the color (red, green, blue; RGB) digital photos were transformed into a single-band grayscale image, as explained in [38]. Additionally, in order to optimize the matching conditions and shorten the processing time, the VIS images were downsampled by a scale factor approximately equal to the ratio of the differences in pixel dimensions between the two data types (see Table 1). This resulted in the hyperspectral and digital images having near-identical ground sampling distances.

The difference between the two image types is apparent in Figure 6 for the Pozalagua dataset.

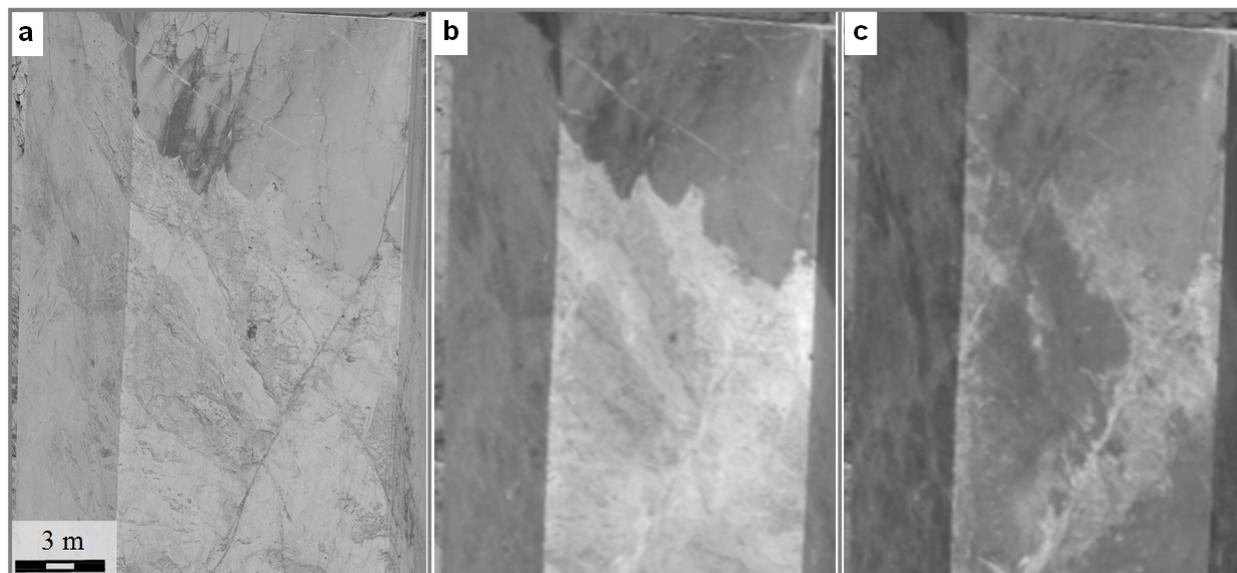
**Figure 4.** Nikon D200 images covering part of the Pozalagua quarry, converted to grayscale. Arrows indicate extents of SWIR image areas outside of the used VIS images. (a) upper and (b) lower image strip.



**Figure 5.** (a) Nikon D200 images covering the TT Niche tunnel face, converted to grayscale; (b) Area corresponding to Figure 3(c).



**Figure 6.** (a) Sample of the greyscale Nikon D200 and the corresponding area in image P-B1 (b) band 1: 1.336  $\mu\text{m}$  and (c) band 158: 2.102  $\mu\text{m}$ .



## 5. Optimization of Selected SIFT Parameters

### 5.1. Experiment Description

The entire SIFT keypoint extraction and matching procedure is controlled by 17 different parameters (see details in [31]). Determining the optimum of all these parameters would require millions of image matching instances, especially if each hyperspectral image is covered by several digital images, and multiple hyperspectral image bands are used. The resulting analysis would be extremely complex to conduct. Therefore, five key parameters (Table 2) were chosen to restrict the scope of the computations, carefully selected following recommendations from related papers (*i.e.*, [31]) and empirical results.

**Table 2.** SIFT parameters considered in the optimization.

Parameter	Default Value [21]	Range in Pre-Analysis	Interval in Pre-Analysis	Range in Analysis	Interval in Analysis
Scales in octave ( $nScales$ )	3	3–6	1	3–6	1
Initial Gaussian blur ( $\sigma$ )	1.6	0.7–2.0	0.1	1.0–1.3	0.1
Contrast threshold ( $CT$ )	0.03	0.0–0.05	0.01	0.01–0.03	0.01
Edge threshold	10	5–15	5	10	-
Nearest neighbor ratio ( $nnRatio$ )	0.8	0.8–0.98	0.05	0.8–0.98	0.05

Modification of the SIFT parameters resulted in an increased number of matched points as well as a high number of false matches. Therefore, as in [15], the Random Sample Consensus method (RANSAC) [39] was used to eliminate the remaining outliers. Depending on the number of matches, a fundamental matrix or a homography model was fit to the data to check the consistency of the matched image points. Although these models are inappropriate for orientation purposes when applied to

images in central perspective and panoramic cylindrical projections, they can be employed in combination with RANSAC as an effective method for outlier detection.

In order to analyze the difference in number of points matched between the digital images and different bands of the hyperspectral image (recorded in different wavelengths) a subset of nine bands equally distributed along the spectral range of the HySpex camera was selected, starting at band 1. This subset was identical for all the images (Table 3) in both study sites, with bands strongly affected by atmospheric absorption excluded.

**Table 3.** HySpex SWIR-320m band subset used in analysis.

Band number	1	31	62	93	124	155	186	210	235
Wavelength ( $\mu\text{m}$ )	1.336	1.482	1.643	1.785	1.936	2.087	2.239	2.356	2.478

Before conducting the full analysis, the allowable range of all parameters values and increments was first defined. During this pre-analysis, three selected hyperspectral bands (bands 1, 124 and 235) were matched with all the covering digital images, using a combination of all possible parameter values listed in Table 2. In result the range of values for the Gaussian blur, contrast threshold and the edge threshold were restricted as shown in Table 2 and discussed in Section 6. In the detailed analysis, all nine bands of the six hyperspectral images were matched with the covering VIS photos using all possible (240) parameter value combinations, resulting in over 71,000 image matching instances (individual VIS photo matched with a single hyperspectral band using a single SIFT parameter configuration). For the sake of simplicity, in both image types the keypoints were extracted with identical parameter configurations. The total number of points matched between each hyperspectral band and all the covering photos was recorded for each parameter configuration (matching results per single hyperspectral band). Due to the fact that one unique point was often matched on several hyperspectral bands, additionally the total number of unique points matched on all nine considered hyperspectral bands was recorded (matching results per hyperspectral image).

## 5.2. Uncertainty

It is worth stressing that full evaluation of the matching performance (e.g., matching accuracy or recall) is outside the scope of this paper. The main goal of the research was to analyze general dependencies between the SIFT parameters and the number of matched points in order to find an optimized set of parameters providing the highest number of conjugate points, allowing the number of matches to be maximized. Consequently, the number of correct matches rejected as false (false negatives) has not been assessed. To evaluate the risk that the set of homologous points estimated as correct using RANSAC contained an unacceptable number of incorrect matches (false positives), a visual check of 25% of the image matching instances was carried out. Results showed single false matches among the RANSAC inliers in 0.45% of the cases (80 false matches per 17,750 matching instances). The result of this analysis was assessed as sufficient for the scope of this paper. Furthermore, any potentially remaining false matches are eliminated automatically during later stage of processing, using bundle block adjustment, which is employed as the final stage of hyperspectral image registration (see [15] for more details).

### 5.3. Analysis of Results

Analysis of Variance (ANOVA) [40] with a linear least squares model was used to assess the influence of the individual SIFT parameters on the matching results. ANOVA is a statistical technique used for revealing the influence (level of significance) of factors (or interactions) on a particular response (here: number of matched points). This method separates the total variability of the response into contributions of every factor and the error term (unexplained variance). Influence (significance) of the parameters is assessed on the basis of the F-ratio, a ratio of variance due to the effect of a factor and variance due to the error term. Significance of a factor indicates if the factor has an impact on the result, but does not indicate how important that impact is. In high power experiments with a large number of observations it is possible to measure significant but minor factor effects. The effect size can be assessed as a proportion of total variation accounted for by a factor [41].

The JMP Statistical Discovery (version 10.0, September 2012) software was used to perform the ANOVA and to analyze the trends and interactions in the dataset using variability graphs. The significance level was set to  $\alpha = 0.05$  level, indicating that 5% of the time, a significant difference is incorrectly declared.

## 6. Results and Discussion

### 6.1. Pre-Analysis

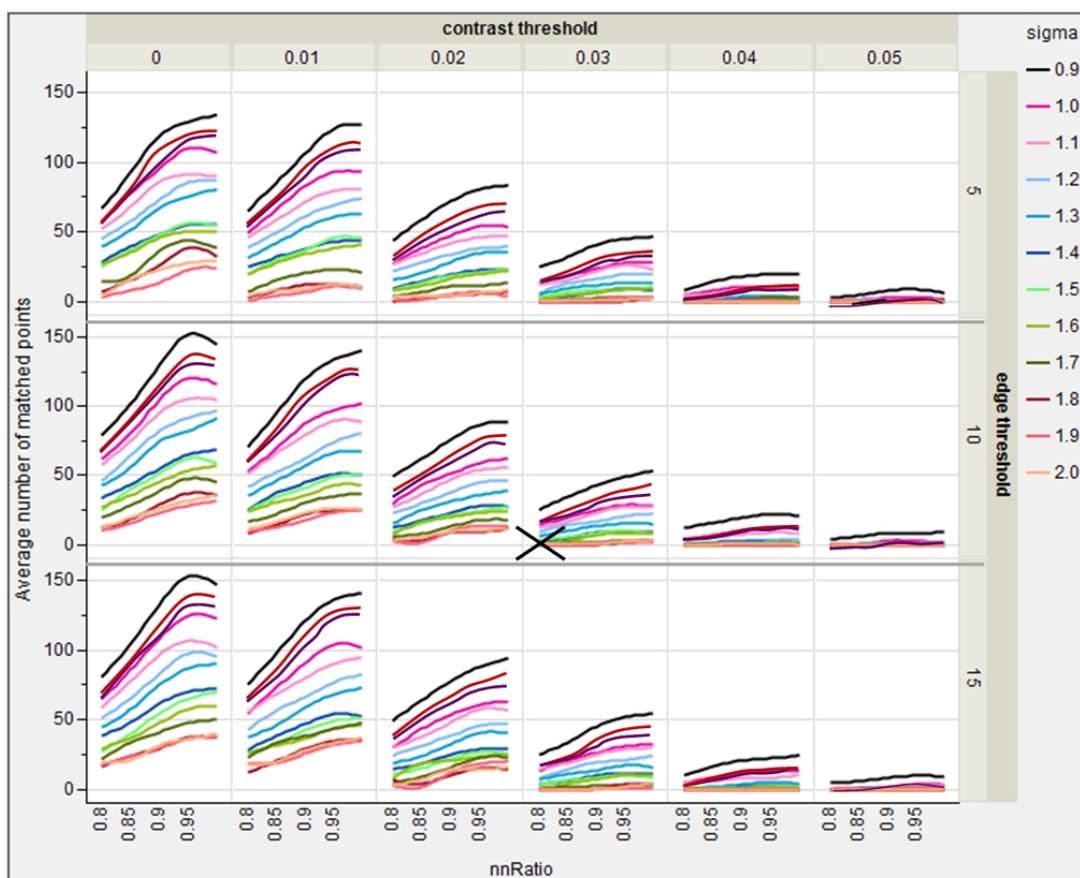
In order to assess the results of the pre-analysis, the variance in the number of points matched per single hyperspectral band was estimated for all the parameters. The trends revealed that even if all the evaluated factors were statistically significant, the edge threshold parameter had a very small impact on the average number of points matched per hyperspectral band (Figure 7). With increasing edge threshold, the average number of points matched per single hyperspectral band also increased, though the change was minimal. Consequently, the edge threshold was fixed at 10 in the detailed analysis, as suggested by Lowe [21].

Another outcome of the pre-analysis was to limit the range of values for the parameter controlling initial image smoothing ( $\sigma$ ), as shown in Table 2. Fewer points were found when high  $\sigma$  values were used, where the contribution of the neighborhood to the pixel value in the smoothed image is larger. This can be observed in Figure 7, by comparing the number of matched points for different  $\sigma$  values. On average, for  $\sigma$  values larger than the default of 1.6, twice fewer points were found than for  $\sigma$  equal to 1.0. On the other hand, the number of points matched was highest when the least initial image smoothing was applied. However, very low initial image smoothing increases the frequency of sampling in the image domain so that many keypoints located very close to each other are detected, while a uniform point distribution within an image is preferred for data registration purposes. Additionally, using the smallest values of  $\sigma$  significantly increased the number of false matches and the processing time. For these reasons, it was decided to use initial image smoothing values in the range [1.0:1.3] in the detailed analysis.

Although yielding the highest number of points, the minimum contrast threshold value (0.0) was also discarded from later analysis. This parameter influences the keypoint localization accuracy in the SIFT feature extraction process, and fitting a 3D quadratic function to the local pixel values in areas

with no contrast change (CT = 0.0) is likely to provide inaccurate keypoint localization and should be avoided, if possible. Because image matching with contrast threshold values larger than 0.03 resulted in significantly fewer points matched, higher values were excluded from detailed analysis (as indicated in Table 2).

**Figure 7.** Average number of points matched per single hyperspectral band for different initial image smoothing (sigma) levels, plotted against the nearest-neighbor ratio (nnRatio) for different contrast thresholds and edge thresholds. Graph created for number of scales in each octave (nScales) equal to 3. Black cross indicates results obtained using default SIFT parameters.

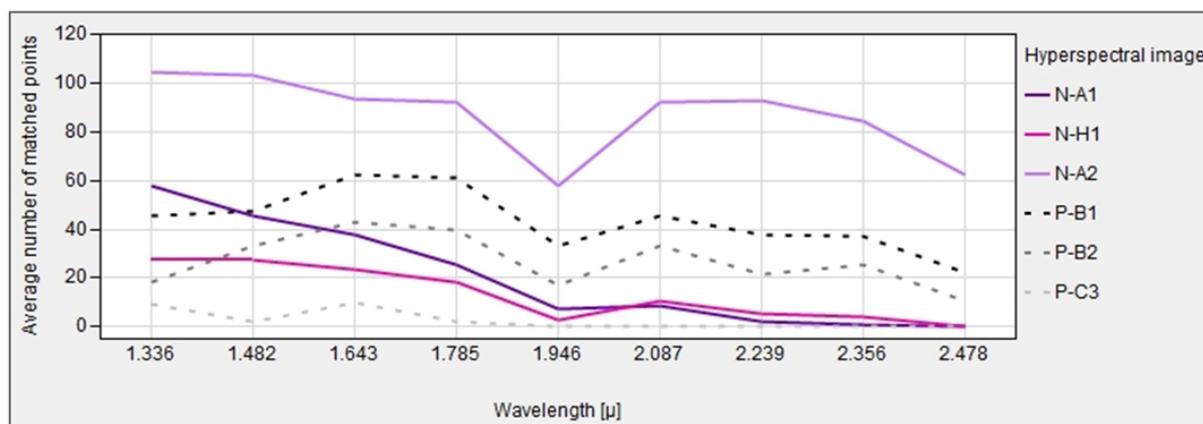


### 6.2. Parameter Optimization Analysis

As expected, the number of points matched differed between the hyperspectral images and between the study areas (Figure 8). The fewest points were matched in image P-C3, with a maximum of 24 points (for 3 scales in an octave) and an average of 2.5 points per single band. Very few points were found in the fresh cut surfaces, characterized by low image texture, and representing almost 90% of the image area. However, increasing the number of scales in an octave (nScales) to 6 increased the number of matched points to 70 (average of 5.5 per band; Figure 8). It requires a clarification here, that the average number of points per single hyperspectral band was computed as an average of all the points matched with all the tested parameter configurations, including all the unsuccessful matching approaches when no points were matched. Because all images were processed with identical SIFT

parameter configurations the average can be treated as a good indicator of differences between the obtained results. Many more points were found in the other Pozalagua quarry images, on average 44 and 27 per band for images P-B1 and P-B2 respectively (Figure 8).

**Figure 8.** Average (from all the tested parameter configurations) number of points matched per single hyperspectral band depending on the band wavelength.



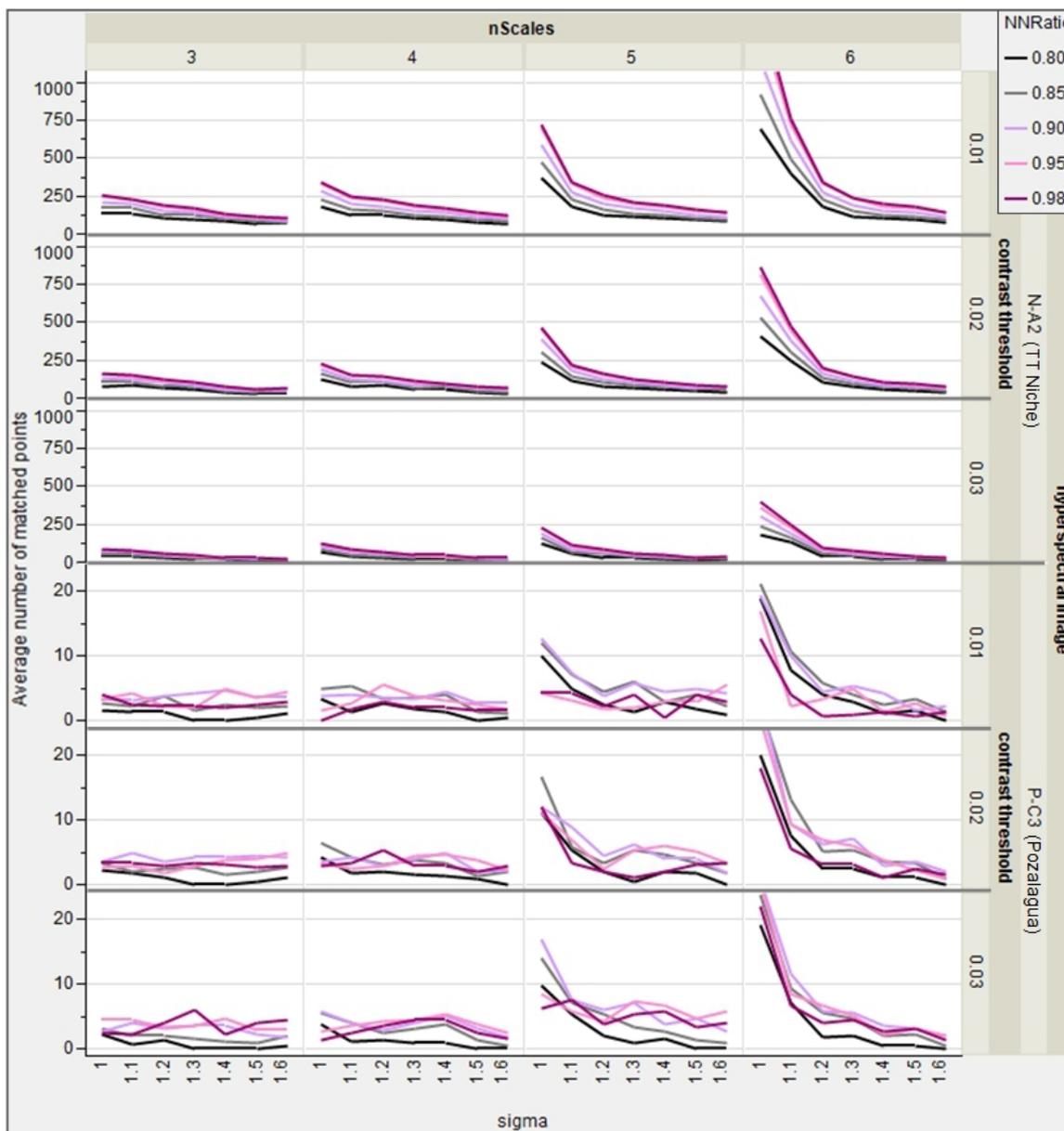
In the TT Niche dataset, as expected, more points were found in image N-A2 covering the upper tunnel face, where fewer changes were apparent between acquisition times of the SWIR and VIS images (see Section 4.1). Although the fewest points were matched in image N-H1 (12 per band on average), this result is very satisfying considering the limited image extents compared to the D200 image (Figure 5(a)).

As also reported in [15,23], with a general increase in hyperspectral wavelength, fewer points were matched between the image band and the VIS images. In the Pozalagua quarry images, the highest number of points was matched in bands 62 (1.634 μm) and 93 (1.785 μm) (Figure 8, dashed lines), with relatively few points found at the start of the HySpex camera spectral range. In the TT Niche dataset the highest number of matches was found in the bands representing the shorter wavelengths of the camera spectral range (Figure 8, solid lines). This difference is most likely caused by a much shorter distance between the hyperspectral scanner and the tunnel face and therefore weaker disturbing influence of atmosphere and lower general data noise level. Except for the excluded bands containing only random noise (Section 4.1), in the Pozalagua dataset more bands, especially between 1.35 μm and 1.42 μm (bands 4–18), were affected by the atmospheric absorption (water vapor) than in the second dataset acquired in a ventilated tunnel. In both datasets the fewest points were found in band 124 (1.936 μm), covering another peak of the atmospheric absorption curve.

In both datasets the number of scales in an octave (nScales) had the strongest influence on the matching results. This trend can be observed in Figure 9 by comparing graphs in different columns, where matching results for different numbers of scales in an octave (nScales) are presented. For initial image smoothing with lower sigma values ( $\sigma < 1.2$ ), a one-step increase in the number of scales in an octave (for  $nScales \geq 4$ ), almost doubles the number of matched points. In contrast, an increase in number of scales in each octave had little impact for images initially blurred with high sigma. This interaction between the two parameters results from the fact that the more scales are sampled in each octave the more times the initially smoothed image is blurred. With high initial sigma, the frequency of

sampling in the image domain is reduced and very close keypoints are not detected or are merged. In this case, a further increase of image blur by increasing nScales does not produce more interest points.

**Figure 9.** Average number of points matched for a single hyperspectral band, showing influence of all the tested parameters for N-A2 (TT Niche top three rows) and P-C3 (Pozalagua quarry, bottom three rows). Note different scales on left vertical axis.

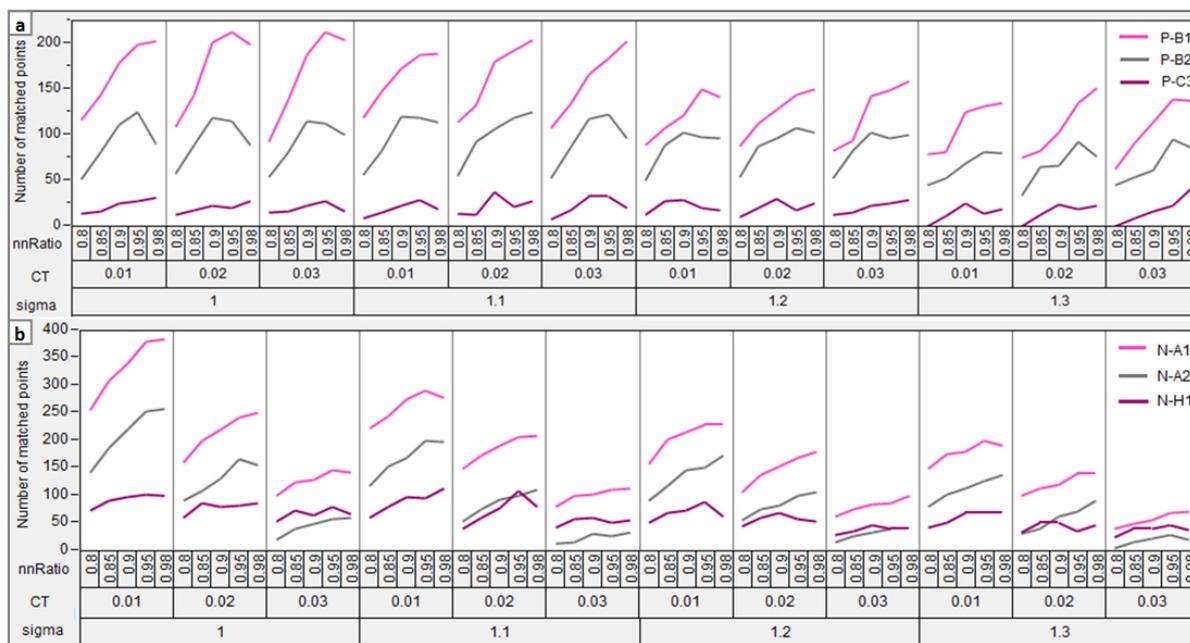


The influence of the contrast threshold varied in each of the datasets (compare rows related to each image in Figure 9), from very strong in the TT Niche dataset to very weak in the data from the Pozalagua quarry. This behavior is related to the differences in texture and appearance of the images for each study site. As previously mentioned (Section 2), the contrast threshold controls extraction of keypoints in low contrast areas. When using low CT values (<0.03) in low contrast areas, representing most of the TT Niche dataset, a small increase in CT in the SIFT algorithm can eliminate many points.

In the Pozalagua dataset most of the points are found in areas covering weathered outcrop surfaces, which have high contrast and image texture. In these areas, an increase in CT values has less influence.

According to Lowe [21], when matching keypoints using values of the nearest neighbor ratio (nnRatio) larger than 0.85 the probability of incorrect matching will be severely increased. His empirical tests show that using a ratio of 0.95 eliminates only 25% of incorrect matches (compared to 85% when nnRatio equaled 0.85). In the current analysis, using RANSAC to eliminate incorrect matches allowed for application of higher nnRatio values. Increasing the nearest neighbor ratio value to 0.95 increased the number of points matched in cases where a high number of homologous points was found. Generally if fewer keypoints are found, also fewer keypoints have their second-nearest neighbor very similar to the nearest neighbor, and the influence of nnRatio is very weak. This can be observed in Figure 10 by comparing inclination of plots for images with high and low numbers of points matched.

**Figure 10.** Total number of points matched in nine bands of the hyperspectral images from the Pozalagua quarry (a) and the TT Niche tunnel (b), depending on chosen SIFT parameters: the initial smoothing for the first level of each octave (sigma), contrast threshold and nearest neighbor ratio (nnRatio) values. Number of scales in each octave (nScales) equal to 3.



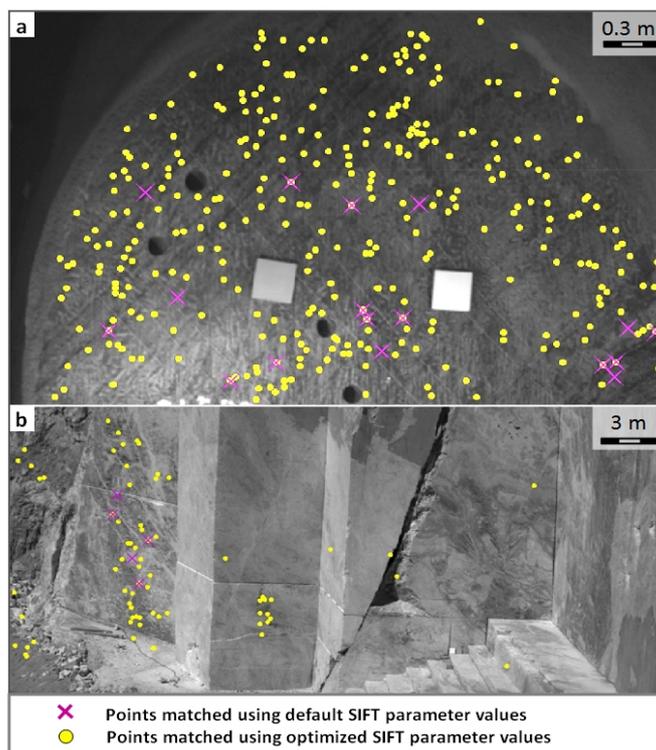
As previously mentioned, in order to maximize the number of matched points for integration of the images with the 3D data, not only the SIFT parameters were tuned but also multiple spectral bands used. All points matched in more than one hyperspectral band were classified as a single (unique) point if the point position varied by less than 0.1 pixels between bands and matched to a single point in the Nikon image. The total number of unique points matched in nine bands of the hyperspectral images is shown in Figure 10. Because the optimal parameters values are very similar for each image within one dataset, one set of parameters per site can be used. The range of optimal SIFT parameter values resulting from this analysis is summarized in Table 4.

Table 4 also reports the maximal total number of unique points matched for the nine bands of each hyperspectral image with the optimized and default SIFT parameters. Not only the quantity but also the spatial distribution of extracted points is crucial for the image registration accuracy. Using SIFT with optimized parameter values resulted in a high number of well-distributed points (Figure 11(a)) in all images except P-C3, where for nScales equal to 3 most of the 24 extracted points were clustered in the left part of the image. Increasing the number of scales in each octave to six resulted in 70 homologous points matched, with improved, but still non-optimal point distribution (Figure 11(b)). Decreasing the prior image smoothing (sigma) value did not result in more matches in the low-textured areas. Including more bands of the hyperspectral image slightly improves the point distribution, though at the cost of increased processing time.

**Table 4.** Optimized SIFT parameters with corresponding number of points matched in nine bands of the hyperspectral images.

Site	Image	nScales	Sigma	Contrast Threshold	Edge Threshold	nnRatio	Maximal Total Number of Matched Points	
							Optimized Parameters	Default Parameters
Pozalagua	P-B1	3	1.0–1.1	0.01–0.03	10	0.9–0.95	214	31
	P-B2	3	1.0–1.1	0.01–0.03	10	0.9–0.95	132	30
	P-C3	6	1.1	0.03	10	0.9–0.95	70	5
TT Niche	N-A1	3	1.0	0.01	10	0.95–0.98	258	12
	N-A2	3	1.0	0.01	10	0.95–0.98	383	17
	N-H1	3	1.0	0.01	10	0.95–0.98	101	14

**Figure 11.** Points matched using SIFT with default and optimized parameter values shown in Table 3. (a) Image N-A2. (b) Image P-C3, 6 scales per octave sampled.



Optimization approaches such as the one presented here or in [31] are time consuming and might take up to several days of mostly CPU time. On the other hand in most cases one optimization analysis per entire dataset is sufficient and the automatically matched points allow for significant time gain in the process of data registration. For example registration of ten hyperspectral images from the Pozalagua quarry, described in [30], required approximately two days of work using a manual approach, whereas a similar results was obtained in less than three hours of CPU time when using automatic image matching.

## 7. Conclusions

In this paper the role of five scale invariant feature transform (SIFT) parameters was explored, and optimization of their values for matching multi-wavelength imagery was documented. The SIFT interest operator was employed to find corresponding points between images acquired in significantly different spectral ranges—visible light (Nikon D200 photos) and short wave infrared (HySpex SWIR-320m imagery)—acquired with different resolutions, geometric projections, and from different viewing positions.

In the two examined geological datasets, using the SIFT operator with optimized parameters and additional outlier elimination allowed between 4 and 22 times more homologous points to be found than when using the default parameter values. This increase has positive implications for registration of multi-wavelength imagery, which is becoming more relevant as new sensors are developed.

Among the five analyzed SIFT parameters, the initial image smoothing ( $\sigma$ ), the number of scales sampled in each octave ( $nScales$ ) and the contrast threshold (CT) had the strongest influence on the number of points matched. In the studied datasets, using less initial image smoothing ( $\sigma \sim 1.0$ ) than suggested in [21] brought a higher number of matched points. Decreasing this value even further can yield a higher number of points matched, though more clustered, and at the cost of longer processing time. Similarly, increasing the number of scales sampled in each octave helped to find more homologous points, also at the cost of increased processing time. For one image, where very few points were matched using the default  $nScales$  (equal to 3), increasing this value to 6 resulted in nearly 2.9 times more homologous points matched. Adjustment of the contrast threshold value, from the default 0.03 to 0.01, was especially important for images with very homogenous texture and little contrast. In both datasets the edge threshold parameter had the least impact on the matching results and the value (10) suggested in [21] was maintained. Because an increased number of points matched between the SWIR and VIS images resulted in a large number of false matches, the remaining incorrect matches were eliminated using RANSAC. This in turn allowed higher nearest neighbor ratio ( $nnRatio$ ) values to be used (0.9–0.98 as opposed to the default of 0.8), resulting in a higher number of points matched.

Despite diverse image characteristics within the two datasets, the optimized values are valid for all images within a single dataset. The significant increase in number of homologous points found between images acquired in non-overlapping parts of the electromagnetic spectrum suggest that better understanding and adjustment of the SIFT parameters can be beneficial for studies where the default values bring an unsatisfactory point quantity and distribution. The latter statement can also be extended to other image matching algorithms where tuning is conducted at the algorithm validation stage for a specific image content (e.g., covering man-made objects).

Although empirical SIFT parameter selection proved to provide satisfactory results for the assessed datasets, more work is necessary to evaluate the efficiency of RANSAC in false match detection, as well as to estimate the overall matching accuracy. Applicability and effectiveness of the methodology suggested in [34] for discriminating between correct and incorrect matches should be explored. To facilitate suitability assessment of different interest operators for matching multi-wavelength imagery, further work is required to increase the level of automation in the parameter optimization process, e.g., by using genetic algorithms [34].

## Acknowledgments

This research was supported by the Research Council of Norway and FORCE (project number 193059). Tobias Kurz and John Howell are thanked for assisting with collection of the hyperspectral data. Statoil ASA is acknowledged for supporting the data collection in Spain. NAGRA, Swisstopo and the Mont Terri Consortium are thanked for financial support in data collection in Switzerland. Jens K. Becker (NAGRA) is thanked for logistical assistance and project organization. The authors acknowledge Norsk Elektro Optikk AS for providing hardware support.

## Conflict of Interest

The authors declare no conflict of interest.

## References

1. Tuytelaars, T.; Mikolajczyk, K. Local invariant feature detectors: A survey. *Found. Trends Comput. Graph. Vis.* **2007**, *3*, 177–280.
2. Harris, C.; Stephens, M. A Combined Corner and Edge Detector. In Proceedings of 4th Alvey Vision Conference, Alvey, UK, 31 August–2 September 1988; pp. 147–151.
3. Meierhold, N.; Spehr, M.; Schilling, A.; Gumhold, S.; Maas, H.-G. Automatic feature matching between digital images and 2D representations of a 3D laser scanner point cloud. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2010**, *38*, 446–451.
4. Rosten, E.; Drummond, T. Machine Learning for High-speed Corner Detection. In Proceedings of European Conference on Computer Vision, ECCV 2006, Graz, Austria, 7–13 May 2006; pp. 1–14.
5. Calonder, M.; Lepetit, V.; Strecha, C.; Fua, P. BRIEF: Binary robust independent elementary features. *Lect. Note. Comput. Sci.* **2010**, *6314*, 778–792.
6. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An Efficient Alternative to SIFT or SURF. In Proceedings of IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, 8–11 November 2011; pp. 2564–2571.
7. Matas, J.; Chum, O.; Urban, M.; Pajdla, T. Robust wide-baseline stereo from maximally stable extremal regions. *Image Vision Comput.* **2004**, *22*, 761–767.
8. Morel, J.-M.; Yu, G. ASIFT: A new framework for fully affine invariant image comparison. *J. Imaging Sci.* **2009**, *2*, 438–469.

9. Alahi, A.; Ortiz, R.; Vandergheynst, P. FREAK: Fast Retina Keypoint. In Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2012, Providence, RI, USA, 16–21 June 2012; pp. 510–517.
10. Lowe, D. Object Recognition from Local Scale-Invariant Features. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; Volume 2, pp. 1150–1157.
11. Wessel, B.; Huber, M.; Roth, A. Registration of near real-time SAR images by image-to-image matching. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2007**, *36*, 179–184.
12. Khan, N.Y.; McCane, B.; Wyvill, G. SIFT and SURF Performance Evaluation against Various Image Deformations on Benchmark Dataset. In Proceedings of 2011 International Conference on Digital Image Computing: Techniques and Applications, Noosa, Australia, 6–8 December 2011; pp. 501–506.
13. Lingua, A.; Marenchino, D.; Nex, F. Performance Analysis of the SIFT Operator for Automatic Feature Extraction and Matching in Photogrammetric Applications. In Proceedings of the 10th CESCAG Conference, Budmerice Castle, Slovakia, 20–22 April 2009; Volume 9, pp. 3745–3766.
14. Nieto, J.I.; Monteiro, S.T. 3D Geological Modelling Using Laser and Hyperspectral Data. In Proceedings of the IGARSS 2010 Symposium, Honolulu, HI, USA, 25–30 July 2010; pp. 4568–4571.
15. Sima, A.A.; Buckley, S.J.; Kurz, T.H.; Schneider, D. Semi-automatic integration of panoramic hyperspectral imagery with photorealistic lidar models. *Photogramm. Fernerkund. Geoinf.* **2012**, *2012*, 443–454.
16. De Matías, J.; de Sanjosé, J.J.; López-Nicolás, G.; Sagüés, C.; Guerrero, J.J. Photogrammetric methodology for the production of geomorphologic maps: Application to the Veleta Rock Glacier (Sierra Nevada, Granada, Spain). *Remote Sens.* **2009**, *1*, 829–841.
17. Delponte, E.; Isgrò, F.; Odone, F.; Verri, A. SVD-matching using SIFT features. *Graph. Model.* **2006**, *68*, 415–431.
18. Yun, S.U.; Min, D.; Sohn, K. 3D Scene Reconstruction System with Hand-held Stereo Cameras. In Proceedings of 3DTV Conference, Kos Island, Greece, 3–7 May 2007.
19. Ostiak, P. Implementation of HDR Panorama Stitching Algorithm. In Proceedings of the 10th CESCAG Conference, Castá-Papiernicka, Slovakia, 23–26 April 2006.
20. Ogawa, Y.; Shimada, N.; Shirai, Y. Environmental Mapping for Mobile Robot by Tracking SIFT Feature Points Using Trinocular Vision. In Proceedings of the SICE Annual Conference, Takamatsu, Japan, 17–20 September 2007; pp. 1996–2001.
21. Lowe, D. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110.
22. Yi, Z.; Cao, Z.; Xiao, Y. Multi-spectral remote image registration based on SIFT. *Electron. Lett.* **2008**, *44*, 4–5.
23. Aguilera, C.; Barrera, F.; Lumbreras, F.; Sappa, A.D.; Toledo, R. Multispectral image feature points. *Sensors* **2012**, *12*, 12661–12672.
24. Buckley, S.J.; Kurz, T.H.; Howell, J.A.; Schneider, D. Terrestrial lidar and hyperspectral data fusion products for geological outcrop analysis. *Comput. Geosci.* **2013**, *54*, 249–258.
25. Goetz, A.F.H.; Vane, G.; Solomon, J.E.; Rock, B.N. Imaging spectrometry for earth remote sensing. *Science* **1985**, *228*, 1147–1153.

26. Kruse, F.A.; Bidell, R.L.; Taranik, J.V.; Peppin, W.A.; Weatherbee, O.; Calvin, W.M. Mapping alteration minerals at prospect, outcrop, and drill core scales using imaging spectrometry. *Int. J. Remote Sens.* **2012**, *33*, 1780–1898.
27. Kurz, T.H.; Dewit, J.; Buckley, S.J.; Thurmond, J.B.; Hunt, D.W.; Swennen, R. Hyperspectral image analysis of different carbonate lithologies (limestone, karst and hydrothermal dolomites): The Pozalagua Quarry case study (Cantabria, North-west Spain). *Sedimentology* **2012**, *59*, 623–645.
28. Clark, R.N.; King, T.V.V.; Klejwa, M.; Swayze, G.A. High spectral resolution reflectance spectroscopy of minerals. *J. Geophys. Res.* **1990**, *95*, 12653–12680.
29. Clark, R.N.; Roush, T.L. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *J. Geophys. Res.* **1984**, *89*, 6329.
30. Kurz, T.H.; Buckley, S.J.; Howell, J.A.; Schneider, D. Integration of panoramic hyperspectral imaging with terrestrial lidar data. *Photogramm. Rec.* **2011**, *26*, 212–228.
31. May, M.; Turner, M.J. Scale Invariant Feature Transform: A Graphical Parameter Analysis. In Proceedings of the BMVC 2010 UK. Postgraduate Workshop, Aberystwyth, UK, 31 August–3 September 2010; pp. 1–11.
32. Park, U.; Pankanti, S.; Jain, A.K. Fingerprint Verification Using SIFT Features. In Proceedings of SPIE Defense and Security Symposium 2008, Orlando, FL, USA, 16–20 March 2008; pp. 1–9.
33. Weinmann, M.; Jutzi, B. Fully automatic image-based registration of unorganized TLS data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2011**, *38*, 1–6.
34. May, M. Data Analytics and Methods for Improved Feature Selection and Matching. Ph.D. Dissertation, The University of Manchester, Manchester, UK, 2012; p. 232.
35. Maestas, D.R.; Lumia, R.; Starr, G.; Wood, J. Scale Invariant Feature Transform ( SIFT ) parametric optimization using taguchi design of experiments. *Design* **2010**, *6424*, 630–641.
36. Alonso-Fernandez, F.; Tome-Gonzalez, P.; Ruiz-Albacete, V.; Ortega-Garcia, J. Iris Recognition based on SIFT Features. In Proceedings of 2009 First IEEE International Conference on Biometrics, Identity and Security (BIDS), Tampa, FL, USA, 22–23 September 2009; pp. 1–8.
37. Schneider, D.; Maas, H.-G. A geometric model for linear-array-based terrestrial panoramic cameras. *Photogramm. Rec.* **2006**, *21*, 198–210.
38. Durda, F. *NTSC Primer*. Available online: [http://nemesis.lonestar.org/reference/internet/web/color/ntsc\\_primer](http://nemesis.lonestar.org/reference/internet/web/color/ntsc_primer) (accessed on 22 October 2012).
39. Fischer, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Graph. Image Process.* **1981**, *24*, 381–395.
40. Montgomery, D.C. *Design and Analysis of Experiments*, 6th ed.; Wiley: New York, NY, USA, 2004; p. 660.
41. Howell, D.C. *Statistical Methods for Psychology*, 4th ed.; Wadsworth: Belmont, CA, USA, 1997; p. 332.