A Two-Step Descriptor-Based Keypoint Filtering Algorithm for Robust Image Matching

Vahid Mousavi, Masood Varshosaz, Fabio Remondino, Saied Pirasteh* and Jonathan Li

***Abstract*— Finding robust and correct keypoints in images remains a challenge, especially when repetitive patterns are present. In this paper, we propose a universal two-step filtering method to solve the mismatch problem in repetitive patterns. Having applied a mean-shift clustering algorithm to remove obvious mismatches, the proposed Confusion Reduction (CR) method uses a novel confusion index in a gridding schema to identify and filter out the remaining confusing keypoints. In both steps, the descriptors' statistical properties are evaluated using kernel density estimation. Various synthetic and real stereo pairs, along with multi-view image blocks were used to assess the performance of the presented algorithm. The results were also compared to those obtained by several state-of-the-art mismatch removal methods. The experiments showed that, on average, the proposed strategy improves the accuracy of matching by 10% and the accuracy of photogrammetric blocks by 20%-30%.**

***Index Terms*— Keypoint filtering, Image matching, Descriptor, Kernel density estimation, Mean-shift**

I. Introduction

Numerous photogrammetry and computer vision applications, such as 3D reconstruction [1], image registration [2], [3], change detection [4] and object recognition [5] require automatic keypoint detection and description. Various descriptors and feature matching comparison methods, such as brute force and Fast Library for Approximate Nearest Neighbors (FLANN), have been developed over the years. To construct feature matches appropriately, keypoints must be repeatable, discriminative, geometrically invariant and insensitive to brightness changes in the scene. In addition, computational efficiency should be considered in order to minimize memory consumption [6].

As an alternative to traditional hand-crafted methods, researchers have proposed learning-based descriptors using Convolutional Neural Networks (CNNs) [7]. LIFT (Learned Invariant Feature Transform) [8], Hard-Net [9], LF-Net (Local Feature Network) [10], Super-Point [11] and reinforced Super-Point [12]are some examples of learning-based descriptors. However, learning-based methods usually lack invariance in many geometrical transformations like large rotation or low-overlapped image pairs, making them only suitable for limited applications [13].

In contrast, traditional hand-crafted descriptors vary greatly by application. They are classified as real (floating-points) or binary (bit-strings) based on processing and memory requirements [14]. Oriented FAST and Rotated BRIEF (ORB) [15], Robust Invariant Scalable Keypoints (BRISK) [16], Fast Retina Keypoint (FREAK) [17], Binary Online Learned Descriptor (BOLD) [18] are some popular examples of binary descriptors. Binary descriptors are a compact representation of an image patch or region in the form of a binary string. Floating descriptors, on the other hand, are mostly high-dimensional real-valued, which require extensive computation costs. Examples of well-known floating descriptors are SIFT [19], SURF [20], DAISY[21], Local Intensity Order Pattern (LIOP) [22], Distinctive Order Based Self-Similarity (DOBSS) [23], Adaptive Binning Scale-Invariant Feature Transform (AB-SIFT) [24] and Radiation-variation Insensitive Feature Transform (RIFT) [25]. Binary descriptors, unlike non-binary descriptors, are often based on simple procedures that require less memory and correspond faster, making them suitable for real-time applications. Floating descriptors are the most widely used descriptors in photogrammetry applications [26].

Further to descriptor computation, feature matching methods look for descriptors that are similar in both target and source images. The Euclidean distance is usually used to compare descriptors [19]. When comparing distances between nearest and second-nearest neighbors, a match is only accepted if the second-best descriptor is significantly farther away than the best one, thereby avoiding selecting the incorrect target point.

As the main concern of this paper, repetitive patterns are a common and troublesome problem when trying to match keypoints in different images. Keypoints extracted from such visual patterns have almost identical descriptors. As a result, even with advanced algorithms, many points are matched incorrectly. Fig1 shows an example of some of such mismatches occurred using SIFT.

*Manuscript received ……; revised ……; accepted ……. Date of publication ……; date of current version …….

Vahid Mousavi is with the Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, Sichuan, China, and the Department of Photogrammetry and Remote Sensing, Geomatics Engineering Faculty, K.N. Toosi University of Technology, Tehran 15433-19967, Iran (e-mail: vmoosavy@mail.kntu.ac.ir).

Masood Varshosaz is with the Department of Photogrammetry and Remote Sensing, Geomatics Engineering Faculty, K.N. Toosi University of Technology, Tehran 15433-19967, Iran (e-mail: varshosazm@kntu.ac.ir).

Fabio Remondino is with the 3D Optical Metrology Unit, Bruno Kessler Foundation, Trento, Italy (remondino@fbk.eu).

Corresponding author: Saied Pirasteh is with the Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, Sichuan, China (sapirasteh@swjtu.edu.cn)..

Jonathan Li is with department of Geography and Environmental Management, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada, (junli@uwaterloo.ca).

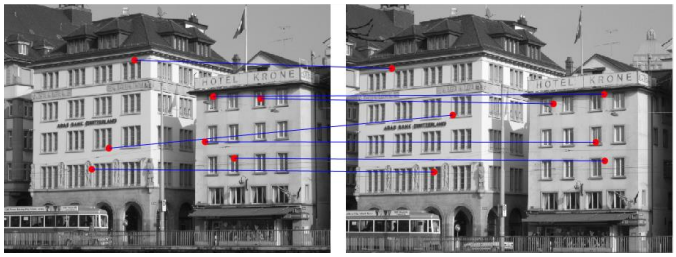


Fig. 1. An example of mismatches occurred in an image pair using the SIFT descriptor and due to repetitive patterns. (Other matches were omitted from the scene just to avoid clutter in the figure)

Up until now, different methods have been proposed to reduce ambiguity in repetitive patterns and find more reliable matches, which work either prior to the keypoint detection and description step [27]–[29] or after the matching process. As will be shown in the following section, most existing methods suffer from different shortcomings. These include sensitivity to the high number of outliers within the initial putative set [30], inefficiency in images with serious geometric deformation or with wide baselines [6] and poor distribution of the matched keypoints. To resolve these problems, in this paper, we propose a new technique, namely Confusion Reduction (CR), which formulates the repetitive keypoints filtering as a two-step classification problem. Our main goal is to select the subset of keypoints that, in addition to being highly distinctive, are also well-distributed across the images. First, we use a mean-shift clustering algorithm to remove obviously confusing keypoints. Then, using a novel Confusion Index (CI), we estimate how confusing each remaining keypoint is. The method is applied in a grid-based pattern to ensure a minimum of reliable keypoints are selected in each part of the image. As will be shown, our method can be used by any detector\descriptor to improve its matching results; it is not sensitive to the number of outliers and ensures proper distribution of the matched points over the images.

The main contributions of this study can be summarized as:

- Development of an effective keypoint filtering approach based on kernel density estimation. As will be shown, our technique works at the descriptor level and, thus, is not sensitive to the number of outliers. Therefore, it performs well even when excessively repeating patterns.
- Proposing a novel Confusion Index (CI), based on Probability Distribution Function, that indicates how confusing a keypoint and its equivalent descriptor are. CI is universal, i.e. it can be applied to different kinds of image pairs and can improve results obtained by any matching algorithm,
- Ensuring even distribution of the keypoints: Unlike current algorithms, we employ a gridding strategy that enables our method to achieve the proper distribution of matched points. This is a very important issue for the accurate adjustment of photogrammetric image blocks.
- Comprehensive testing of the proposed algorithm against several state-of-the-art mismatch removal methods. In our experiments, in addition to standard tests, we have evaluated our results in photogrammetric adjustment of different real image blocks.

The remainder of this paper is organized as follows: Section II provides a review of the related works dealing with the mismatch problems. Section III describes the proposed method and demonstrates how confusing keypoints are removed, and high-quality ones are selected using the explained confusion index. Section IV explains the proposed algorithm's evaluation framework, and Section V describes the used datasets. Section VI presents and discusses the results obtained using both synthetic, real image pairs and multi-view data, followed by the performance evaluation. Finally, Section VII concludes the paper and suggests some research lines for future studies.

II. Related Work

Mismatch removal can be classified as pre-processing, post-processing or in-processing algorithms. Pre-processing methods consider the mismatch problem before the matching process. For example, symmetry analysis could be a simple approach to filtering out repetitive patterns in each image [27], [28], [31]–[33]. Such methods frequently assume that repetitive patterns are aligned horizontally and vertically [31] or lie on a planar structure on the image [34]. In other methods, descriptors like SIFT are used to represent image information such as shapes or colors to find repetitive patterns [35]. Recently, pre-trained deep convolutional neural networks [27] with filters learned using natural images have been considered to find repeating patterns in images. Unfortunately, pre-processing approaches may not be suitable for photogrammetric purposes because they reduce image information contents and, thus, the number of extracted keypoints.

Post-processing methods, on the other hand, aim to remove false matches after the matching process. Some of these, which constitute a large number of techniques in the literature, rely on either global or local constraints. Others may involve clustering, graph matching, or learning-based approaches. The methods using global constraints, can roughly be divided into parameter estimation [36] and non-parametric interpolation methods [37]–[39].

Non-parametric methods learn a pre-defined model based on either prior knowledge or through a regression. As an example, Ma et al. [38], [39] proposed a non-parametric model for Vector Field Consensus (VFC) and applied it to mismatch removal. Their proposed Bayesian framework assigns each sample with a variable indicating whether it is an outlier or not. The Expectation-Maximization algorithm is used to solve the problem as a maximum posterior problem. Non-parametric methods use all of the points to identify the mismatches. Therefore, their precision decreases sharply when there are many outliers and/or independent moving structures in the point sets [30].

Parameter estimation methods attempt to obtain a subset of mismatch-free correspondence. The popular Random Sample Consensus (RANSAC) method [40] and its recent extensions such as Extreme Value RANSAC (EVSAC) [41], Graph Cut RANSAC(GC-RANSAC)[42] and Locality Preserving RANSAC (LP-RANSAC) [43] are widely used in this field. Parametric methods rely mainly on the hypotheses of sampling

consensus. This can be inefficient in matching image pairs having a wide baseline and a high percentage of outliers in the initial matching set [30].

Techniques that use local constraints perform a robust estimation of correct matching based on local structural consistency or piecewise consistency assumptions. Locality Preserving Matching (LPM) [45] is an example that uses the difference in local neighborhood structures of inliers and outliers to identify mismatches from a given putative set. This technique is sensitive to a high outlier proportion of the putative set due to the unreliability of neighborhood construction [46]. Moreover, Li et al. [47] proposed a support-line voting strategy based on the neighborhoods of correspondences and outliers filtering using affine-invariant ratios. The authors also proposed a local region descriptor based on a 4-point local structure [48]. These two methods consider both photometric and geometric properties inside a small local region, and their computational complexity is considerably high. In another effort, Li et al. [49] proposed a locality affine-invariant feature matching (LAM) method based on the concepts of local barycentric coordinates (LBCs) and matching coordinate matrices (MCMs). Similarly, Wang and Chen [50] proposed a Guided Local Outlier Factor (GLOF) algorithm for feature matching with gross mismatches under multi-granularity neighborhood structure-preserving. Generally, although methods based on local constraints are efficient, but their accuracy decreases when there are either local distortions or similar patterns in the scenes [30].

A recent approach has been to use clustering techniques to solve the matching problem when the putative matches include a large number of outliers. To eliminate the demand for the geometric constraints, Jiang et al. [51] proposed a DBSCAN-based iterative spatial clustering approach (RFMSCAN) to solve the matching problem when the putative matches suffer from a large number of outliers. In their method, feature matching is formulated as a spatial clustering problem with outliers. The main idea is to adaptively cluster the putative matches into several motion-consistent clusters together with an outlier/mismatch cluster. However, this method is limited because it is sensitive to the clustering parameters, and obvious outliers could be retained [46].

Graph matching is also a post-processing technique used to fix mismatches. Several studies in this field have been reported, including spectral matching [52]–[55] and ABPF (Adaptive & Branching Path Following) [56]. They adapt well to transformation models and obtain good matching results. However, they can be affected by drawbacks of their non-polynomial-hard nature that exponentially increases the required processing time when the dimension of the problem increases [46].

Learning-based approaches, which are usually combined with local neighborhood consensus, have recently been proposed as a type of post-processing method for mismatch removal. Ma et al. [57] developed the Learning for Mismatch Removal (LMR) approach, in which a general classifier is trained to evaluate the correctness of an arbitrary match. Then, using a multiple K-nearest neighbor strategy, match representation is obtained by exploiting the consensus of the local neighborhood structure. Pang et al. [58] proposed a new weakly supervised Graph Convolutional Siamese Network Matcher (GCSNMatcher) for feature matching. It can work directly on unstructured keypoint sets and exploit geometric information within sparse keypoints. The method builds dynamic neighborhood graph structures to improve the feature representation of each keypoint. A similar approach to GCSNMatcher is documented by Li et al., 2019 [46]. Overall, learning-based methods can usually be used in image pairs with slight geometrical deformations, such as medical image registration and binocular stereo matching. Nevertheless, a better understanding of their performance in wide baseline stereo images or image registration with serious geometric deformations is still needed [6].

In contrast to the pre- or post- processing approaches, in-processing techniques try to filter out mismatches during the actual matching process. For example, Mortensen et al. [59] enriched the SIFT descriptor with information about the global context of the image, inspired by shape contexts. The SERP (Surf Enhancer for Repeated Pattern) descriptor [60] uses mean-shift clustering [61] on SURF descriptors, where repetitive features are grouped into a single cluster and non-repetitive features are given their own cluster. In order to detect Local Distinctive Features (LDFs), Chen et al. [62] proposed an interest point detector that considers both the geometric distinctiveness of an image pixel and the support region surrounding it. Unfortunately, such detectors prevent the use of other classic detectors like SIFT and SURF that, if used, better results in photogrammetric adjustments could be achieved.

An interesting in-processing algorithm is that by Royer et al. [14], which has been specifically developed to filter out confusing keypoints in repetitive patterns. Their method, so-called CORE, ignores the visual property of the keypoints, which can vary by detectors. Instead, the descriptor's statistical properties are analyzed using kernel density estimation. A numerical value, namely confusion risk, is tied to each descriptor. To extract the most distinctive subset of keypoints, a probability threshold value is computed for the confusion risk. Although CORE can be used by different algorithms, its threshold is highly affected by the acceptable confusion rate and probability of finding different descriptors between the images [14].

Looking at the above techniques, we can see that most post-processing algorithms are ineffective for photogrammetric applications. Those that use a global/local pre-defined model have low performance in cases that are not characterized by the pre-defined transformation. Since global techniques use almost all of the points, they work poorly where there are many outliers in the initial matching set. Furthermore, methods based on local information are also susceptible to local distortions or similar patterns in the scenes. Regarding the learning-based approaches, precision can drop dramatically when the baselines are wide or when the images include large geometric deformations. Furthermore, the more recent in-processing CORE algorithm has inconsistent correspondence performance in some cases. The main reason is that finding the best settings for the confusion risk threshold is completely difficult since it depends on the image context. More importantly, to the best of our knowledge, the spatial distribution of keypoints has not been adequately considered in current methods, which is crucial for accurate feature correspondence and successful image orientation. As a result, despite the promising results achieved by some of the existing algorithms, they cannot effectively match images with repetitive patterns for photogrammetric

applications. Thus, developing an effective and generic method to filter out confusing keypoints is still an important requirement in the photogrammetry pipeline.

In this paper, we present a new Confusion Reduction (CR) method that is universal, is not subject to the number of outliers and ensures proper distribution of the keypoints across the images matched. Our solution adopts a two-step filtering approach that recognizes and classifies keypoints with similar vectors along a grid over the images. A confusion index is computed for each keypoint, using statistical properties of the associated descriptor to identify and filter out weak points in repetitive patterns. The proposed algorithm is described in detail below.

III. Methodology

Filtering repetitive keypoints is performed using a hierarchical method in two main steps: (1) removing obviously confusing keypoints and (2) selecting a subset of mismatch-free keypoints. The algorithm starts with keypoint extraction and description. Then it removes largely confusing keypoints using a mean-shift clustering algorithm. The idea behind descriptor clustering is that repeated descriptors be grouped in clusters relatively close to each other, and non-repeating descriptors form a separate cluster with fewer elements. Also, we use the mean-shift clustering since it does not require prior knowledge about the number of clusters, nor is constrained by their shape. However, it may be difficult to tune the required parameters and thresholds in different images. Also, since mean-shift clustering does not control the distribution of removed keypoints, controlling the amount of eliminated keypoints and their distribution is impossible. Therefore, the proposed CR algorithm applies a selection strategy to the remaining keypoints (Step 2) to find a subset of high-quality keypoints that include fewer mismatches. A confusion index computed from Reiny entropy for each remaining keypoint controls the number of removed keypoints.

In the following, these steps are described in detail.

A. Outline of the proposed method

Fig.2 shows the process of the proposed keypoint selection algorithm. Considering the initial keypoint location and scale, which are extracted using a keypoint detector in both reference and target image, the proposed strategy can be explained as follows:

1) The initial keypoints are extracted using a detector. Then one of the floating descriptors (i.e., SIFT) is computed for each keypoint in both reference and target images.
2) Mean-shift clustering is performed on all descriptors, and keypoints with close cluster centers (modes) are ignored based on the details explained in Section III-B.
3) The Confusion value for each of the remaining keypoints is computed based on the probability density function and Renyi entropy, which is fully explained in Section III-C.
4) The input reference image is divided into regular grid cells.
5) The number of competences keypoint in each grid cell is computed, as described in Section III-D.
6) Finally, for each grid cell, the initial keypoints are ordered based on their confusion value, and then the required number of keypoints with the lowest confusion value is selected within each grid cell.

B. Mean-shift clustering

As a non-parametric clustering technique, the mean-shift algorithm [61], [63] makes no assumptions about the distribution's shape or the number of clusters. It is based on an accurate analysis of feature spaces, which have different classes of shapes throughout the density estimation. In mean-shift clustering, cluster centers are equivalent to the modes obtained from an estimated density. In our method, mean shift is performed on all extracted descriptors and the modes to which each descriptor converges are determined.

Analyzing the keypoints in the repeating pattern areas of images shows that the cluster centers with repetitive patterns are close together and contain several descriptors. Others are distant and contain fewer descriptors. As a result, computing similarity between modes (Euclidean distance) can separate clusters that contain repetitive pattern descriptors. Therefore, obviously, confusing keypoints could be eliminated.

Let *I* be the image resulting from a specific scene, and *I'* to be another observation of the same scene which resulted from various transformations such as rotation, perspective changes, light modifications and so on. Supposing an input vector $u_i = [u^1.u^2.u^3 \ldots .u^d\ ]$ be *d*-dimensional descriptor computed for a keypoint. Now Let $u_i.i\ \epsilon\{1.2 \ldots .r\}$ be descriptor vectors computed on *r* keypoints of the image *I* and let $u_j.j\ \epsilon\{1.2 \ldots .s\}$, be the descriptor vectors in the image *I'*. The multivariate kernel density estimate obtained with kernel *K(u)* and window radius *h* is as follows:

$$\tilde{f}_{h.k}(u) = \frac{C_{k.d}}{Nh^d}\sum_{q=1}^{N} \mathrm{k}\left(\left\|\frac{u-u_q}{h}\right\|^2\right) \qquad (1)$$

where, $C_{k.d}$ is the normalization constant, *N* is the complete set of all descriptors in both images, *h* is the kernel window size, *k(u)* is the kernel function, and *u* is the *d*-dimensional vector of descriptors.

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3188931

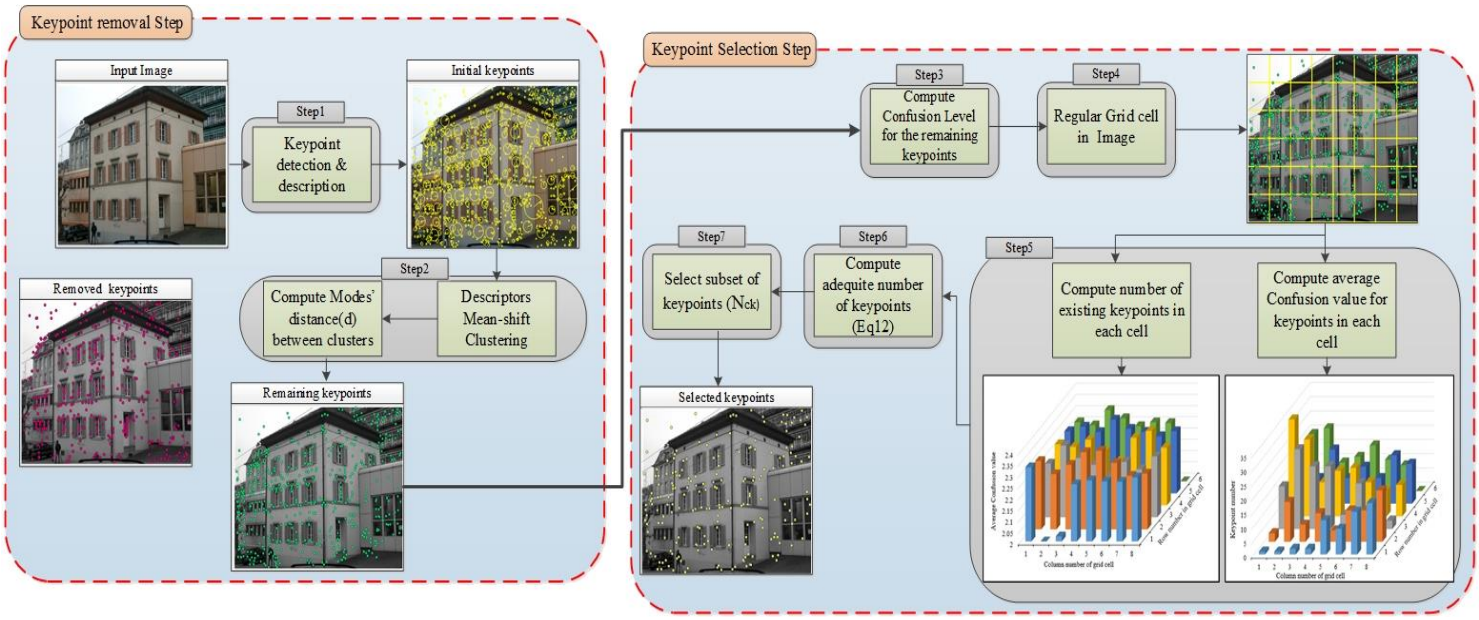


Fig. 2. Flowchart of the proposed CR algorithm.

The modes of the density function are located at the zeros of the gradient function (i.e., $\nabla f(u)=0$), so:

$$\nabla f(u) = \frac{2C_{k.d}}{Nh^{n+2}} \sum_{q=1}^{N} (u - u^q)\, g\left(\left\|\frac{u-u_q}{h}\right\|^2\right) = \frac{2C_{k.d}}{Nh^{n+2}} \left[\sum_{q=1}^{N} g\left(\left\|\frac{u-u_q}{h}\right\|^2\right)\right] \cdot \left[\frac{\sum_{q=1}^{N} u_r\, g\left(\left\|\left(\frac{u-u_q}{h}\right)\right\|^2\right)}{\sum_{q=1}^{N} u_r\, g\left(\left\|\left(\frac{u-u_q}{h}\right)\right\|^2\right)} - u\right] \quad (2)$$

where $g(u) = -k'(u)$. The first term is proportional to the density estimate at u computed with kernel $G(u) = C_{g.d} \cdot g(\|u\|^2)$ and the second term is the mean-shift. The mean-shift vector always points towards the direction of the maximum increase in density. Then, the coordinates of the cluster centers *m(u)* are calculated by:

$$m_h(u) = \frac{\sum_{q=1}^{N} u_r\, g\left(\left\|\left(\frac{u-u_q}{h}\right)\right\|^2\right)}{\sum_{q=1}^{N} u_r\, g\left(\left\|\left(\frac{u-u_q}{h}\right)\right\|^2\right)} - u \quad (3)$$

As previously mentioned, the similarity between modes is determined by measuring the Euclidean distance (d) between cluster centers as follows:

$$d(m(u_k).m(u_w)) = \sqrt{\sum_{i=1}^{d} (m(u_k^i) - m(u_w^i))^2} \quad (4)$$

A threshold value can be considered for the cluster centers Euclidean distance (d) to extract descriptors in repetitive pattern areas. Our experiments show that the threshold value would be different based on the descriptor type and a distance threshold equal to 0.7 and 0.02 is suitable for SIFT and SURF descriptors, respectively (see *Parameter Analysis* Section).

Unfortunately, mean-shift clustering method is strongly affected by kernel window size [60]. This sensitivity can be problematic when either all the descriptors are accumulated in a single cluster, or each descriptor becomes a separated cluster. Therefore, the mean-shift method suffers from the limitation that the number of clusters depends on the selection of the kernel size. Therefore, we develop a new confusion index to solve this problem in the second step of our method (selection step), which is explained in the next section.

C. Keypoints confusion index (CI) computation

It can be assumed that each descriptor vector in an image is subject to slight variations that could be assimilated as randomness in the image. Doing so, u_i could be assumed as random vectors, and an index associated with each keypoint of the image *I* could be defined that characterizes the confusion level. For each keypoint *i* of image *I*, we define an index H_i, called "Confusion Index" which indicates how distinctive the selected keypoint is. This index is computed using Euclidean distance between descriptors, a Probability Distribution Function (PDF) and Rényi's entropy and performs as an efficient tool for separating high confusion risk keypoints.

As explained, u_i and u_j are considered as the descriptors of the extracted keypoints in the first and second image, respectively. For each extracted descriptor of keypoint *i* on the first image, the Euclidean distance between descriptors in the second image can be written as follows:

$$X_{r\times s} = \sqrt{\sum_{k=1}^{d} \left(u_i^k - u_j^k\right)^2} \qquad \forall\, i \in \{1.2 \ldots . r\}; \forall\, j \in \{1.2 \ldots . s\} \quad (5)$$

where $X_{r\times s}$ is the vector of Euclidean distance between *d*-dimensional descriptors of u_i and u_j, respectively.

For each $X_{q\times s}.q = 1.2.3.\ldots r$ vector, the Probability Distribution Function (PDF) can be calculated using K_h kernel as follows:

$$f_q(x) = \frac{1}{r}\sum_{p=1}^{r} K_h\left|x - x_p\right| \quad (6)$$

where $f_q(x)$ is the PDF of the Euclidean distance vector given the keypoint number. It is assumed that numerous reasons cause to vector $|x - x_p|$ variation, which is either of natural origin or may be regarded as such [14]. Therefore, it makes sense to consider the vector $X_{q\times s}$ behavior to be Gaussian. With this assumption, K_h can be defined as the classical D-dimensional Gaussian Kernel:

$$K_h|x - x_p| = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^D exp(\frac{-|x - x_p|}{2\sigma^2})^2 \quad (7)$$

where x_p is each element of $X_{D\times K}$ vector and σ determine the width of the Gaussian kernel. Given the above, the PDF for each descriptor vector is written as:

$$f_q(x) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^D exp(\frac{-|x - x_p|}{2\sigma^2})^2 \quad (8)$$

In order to evaluate the amount of information constituted by the probability distribution function computed using Eq. (8), for all calculated descriptors on the image, a measure of entropy and, more specifically Rényi's entropy is conventionally used [64]. This measure could be used to assess the degree of confusion present in the set of descriptors. The Rényi's entropy is detailed below.

Assuming X is a random variable with constant distribution as $f_x(x)$, the Rényi's entropy of order α, where $\alpha \geq 0$ and $\alpha \neq 1$; $\alpha \neq 0$ as discussed in [64] can be calculated as follows:

$$H(X) = \frac{1}{1-\alpha} log\left(\int (f_x(x))^\alpha dx\right) \quad (9)$$

The entropy of a probability distribution can be interpreted as a measure of both uncertainty and information content. As α approaches zero, the Rényi entropy increasingly weighs all possible events equally, regardless of their probabilities. In the limit for α→0, the Rényi entropy is just the logarithm of the size of the support vector of X. The limit for α→1 is the same Shannon entropy. As α approaches infinity, the Rényi entropy is increasingly determined by the events of the highest probability. Rényi entropy of order α provides more weight for data with lower probability and could be used as a practical tool to measure the information content in each record. Therefore, it can be used to distinguish special descriptors from normal ones in our matching task. In fact, this entropy reveals the distinction between descriptors derived from repeating elements and other descriptors.

By substituting Eq. (8) into Eq. (9), the Renyi entropy for each $X_{D\times K}$ vector could be calculated by

$$H(X) = \frac{1}{1-\alpha} log\left[\int\left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^D exp\frac{-|x - x_i|^2}{2\sigma^2}\right)^\alpha dx\right] \quad (10)$$

The entropy function is considered with $\alpha = 2$. In this way, the value of the function is increased for small probabilities, and a higher weight is assigned. Our final "confusion index" H_i, could be computed by

$$H(X) = -log\left[\int\left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^D exp\frac{-|x - x_i|^2}{2\sigma^2}\right)^2 dx\right] \quad (11)$$

As mentioned before, the higher the "Confusion value" for each keypoint, the higher the probability of mismatching in the image matching process. By labeling each keypoint with its H_i value, the prerequisites for the proposed Confusion Reduction (CR) algorithm are easily achieved.

D. Keypoint selection strategy

Since a numerical Confusion index is associated with each keypoint, a quick method to extract a subset of keypoints could be to sort them according to their Confusion value and only keep the n_{th} first. However, such a solution lacks the capability to control the number and distribution of the remaining keypoints, which is crucial in photogrammetric applications. Therefore, a regular gridding strategy is applied to achieve an even distribution of keypoints in the spatial space to control the number and distribution of remaining keypoints. In the previous section, the "Confusion value" was computed for each keypoint. As shown in Fig.2, the input image is firstly divided into regular grid cells. The existing initial keypoints are then determined for each cell, and therefore the number of required keypoints in each grid cell N_{ck}, is computed by:

$$N_{ck} = \left[\left(\left(1 - \frac{\bar{H}_k}{\sum_{k=1}^{T_{cells}} H_k}\right) + \frac{N_k}{\sum_{k=1}^{Tcells} n_k}\right).n_k\right] \quad (12)$$
$$k = 1, 2, 3, \ldots, Tcells$$

where T_{cells} is the number of regular grid cells, $\bar{H}_k$ is the average of the Confusion values of all initial extracted keypoint in the k_{th} cell. Finally, the initial available keypoints of each grid cell are ordered based on their Confusion values measures, and the N_{ck} of the keypoints with the lowest confusion values are selected within each grid cell.

IV. Evaluation Framework of the Proposed Algorithm

The performance of the proposed CR method was tested on both synthetic and real datasets. The CR method was first evaluated with synthetic images to test the matching efficiency. Additional assessments were also conducted using real image datasets. Then the SfM pipeline and image blocks were used to evaluate the effect that the CR method can have on the orientation results.

In the following, the evaluation methodology adopted in both synthetic (Section IV-A) and real images (Section IV-B) is described. Then, the quality measures used to evaluate the results are explained in (Section IV-C).

A. Evaluations using synthetic data

For the first phase of evaluation, a synthetic dataset was created. The synthetic dataset was designed to remove the influence of image content and texture quality on matching the

results of the proposed algorithm. Also, the stability of the CR keypoint selection algorithm could be tested under similar imaging conditions since radiometric differences are ignored. One image is used as a reference in this dataset, and another is generated using a known geometric transformation. To generate the synthetic dataset, two geometric transformations, including rotation and scale, were applied according to the following equation:

$$\begin{bmatrix} x_k \\ y_j \end{bmatrix} = \begin{bmatrix} s_x \\ s_y \end{bmatrix} \cdot \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \cdot \begin{bmatrix} u_q \\ v_p \end{bmatrix} \quad (13)$$

where, u_q and v_p are the coordinate of a pixel in the input image and the x_k and y_j are the coordinates of the output pixel. s_x and s_y are positive-valued scaling coefficient and θ is the counter-clockwise rotation angle with respect to the horizontal axis of the input image.

After keypoint extraction using SIFT, four well-known descriptors, including SIFT, SURF, DAISY and LIOP, were computed for each of them. After that, the proposed CR algorithm was performed to select high-quality keypoints. Finally, to determine the performance of the CR methods, selected keypoints were matched using Euclidean distance.

B. Evaluations using real images

In the second phase of the evaluation, real image pairs are used to evaluate matching results.

First of all, some representative pictures are plotted from real image pairs datasets to visualize the performances of our algorithm. Then, to investigate the quantitative evaluation of our method, three tests are carried out as follows:

1) Comparison with CORE algorithm:

The result of the CR method is compared to the original descriptor and also the CORE algorithm. For this test, image keypoints are extracted from each test image pair using SIFT and SURF descriptors. The keypoints are then filtered using the proposed CR method and also the CORE algorithm. Finally, a brute-force matcher using Euclidean distance is used to match the filtered descriptors and results for both methods are compared.

2)Comparison with other mismatch removal algorithms:

In this test, the performance of the proposed CR-SIFT method is compared to the other mismatch removal methods. The traditional and basic techniques of RANSAC [40], as well as four state-of-the-art methods, including LPM [45], LMR [57], RFMSCAN [51] and GLOF [50], are chosen for comparison. In particular, RANSAC is a classical resampling method, LPM and GLOF are neighborhood preserving methods, LRM is a learning method, and RFMSCAN is a clustering-based technique. We tried to pick an algorithm from each type of state-of-the-art mismatch removal technique presented in the Related Work Section to have a comprehensive comparison. These algorithms are implemented based on their publicly available codes, with their parameters set according to their literature's suggestion.

3) Evaluating the CR method using multi-view real images:

This test evaluates the performance of the proposed method using multi-view image blocks and the SfM pipeline. These tests evaluate the influence of CR method on image orientation results. As previously explained, the proposed algorithm affects the number and distribution of matches. This phase of the evaluation examines how the CR method affects the image orientation results. The first step is to extract and describe the keypoints using SIFT. Then the CR algorithm is used to filter the extracted keypoints. Afterwards, the filtered keypoints are matched, and finally, the bundle adjustment is started. The results are then compared to the original SfM pipeline. Fig.3 summarizes the whole process.

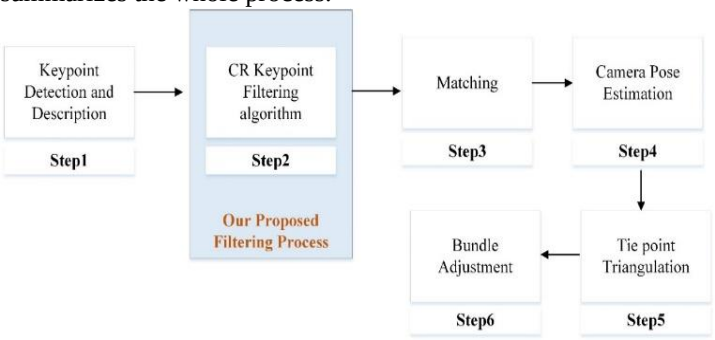


Fig. 3. The keypoint filtering module in the SfM pipeline

C. The quality measures

Five criteria, including recall, precision, positional accuracy, number of correct matches (NCM) and spatial distribution quality, are used to evaluate the capability of the keypoint selection method. The precision and recall criteria are computed using the following equations:

$$Precision = \frac{CM}{TM} \quad (14)$$

$$Recall = \frac{CM}{CM + FN} \quad (15)$$

where Correct Match (*CM*) and False Match (*FM*) are the numbers of correctly and falsely matched keypoint pairs in the matching results, respectively, and TM is the number of total matches. False Negative (*FN*) is the number of existing correctly matched pairs which are incorrectly rejected.

The relationship between each image pair should be known in order to compute the CM, FM and FN parameters. The CM, FM and FN values could be automatically computed in the synthetic image dataset due to the known geometrical relationship between each image pair. We used a spatial threshold equal to 1.5 pixels to separate correct matches from false matches, as suggested by [76].

For the real image dataset, the fundamental matrix is computed to find the relationship between two images. To this end, an expert operator manually selected 40–60 control points for each image pair and calculated the fundamental matrix. Similarly, a spatial threshold equal to 1.5 pixels is used to detect false matches.

Since the location accuracy of the matched keypoints is critical in most photogrammetric computations, the positional

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3188931

accuracy of each method is computed using the Root-Mean-Square Error (RMSE). The RMSE value is calculated using the location of correctly matched keypoints and their computed location determined by the known transformations. To evaluate the distribution quality, the global coverage index (α), which is based on Voronoi diagrams, is computed by:

$$\alpha = \frac{\sum_{i}^{n} = 1 A_i}{A_{Total}} \qquad (16)$$

where A_i is the area of the i_{th} Voronoi cell, n is the number of Voronoi cells and A_{Total} is the area of the whole image. The larger the α value, the better the spatial distribution of the matched pairs.

As no ground control/truth was available, following [73] and [74], four criteria at the end of bundle adjustment were analyzed for comparisons of the real multi-view image orientation results as follows:

a) *Average re-projection of the bundle adjustment*: This criterion expresses the re-projection error of all computed 3D points
b) *Average number of rays per 3D point*: It shows the redundancy of the computed 3D object coordinates.
c) *Visibility of 3D points in more than three images*: It indicates the number of the triangulated points which are visible in at least three images in the block.
d) *Average intersection angles per 3D points*: This criterion shows the intersection angle of 3D points, which are determined by triangulation. A higher intersection angle of homologous rays provides more accurate 3D information.

V. Description of Datasets

As mentioned above, we have used synthetic and real images to evaluate our algorithm. Fig.4 shows the synthetic images (S2 to S6) created using scale and rotation transformations. In this dataset, scale coefficients vary from 1.2 to 2 and rotation angles changes from 15° to 55° by a difference of 10° at each step.

As for the real image datasets, 12 image pairs were selected from three image databases for testing, as shown in Fig.5. Eight stereo image pairs (RP1 to RP8) with the size of 640×480 are used from the Zurich image datasets. These images were selected as their filtering results using CORE were also available [18]. Furthermore, three image pairs (RP9 to RP11) are selected from the PSU database [79] with a size of 1024×768, which contain completely regular and near-regular textures. Finally, one image pair (RP12) is selected from the VGG image database [67] with different viewpoints.

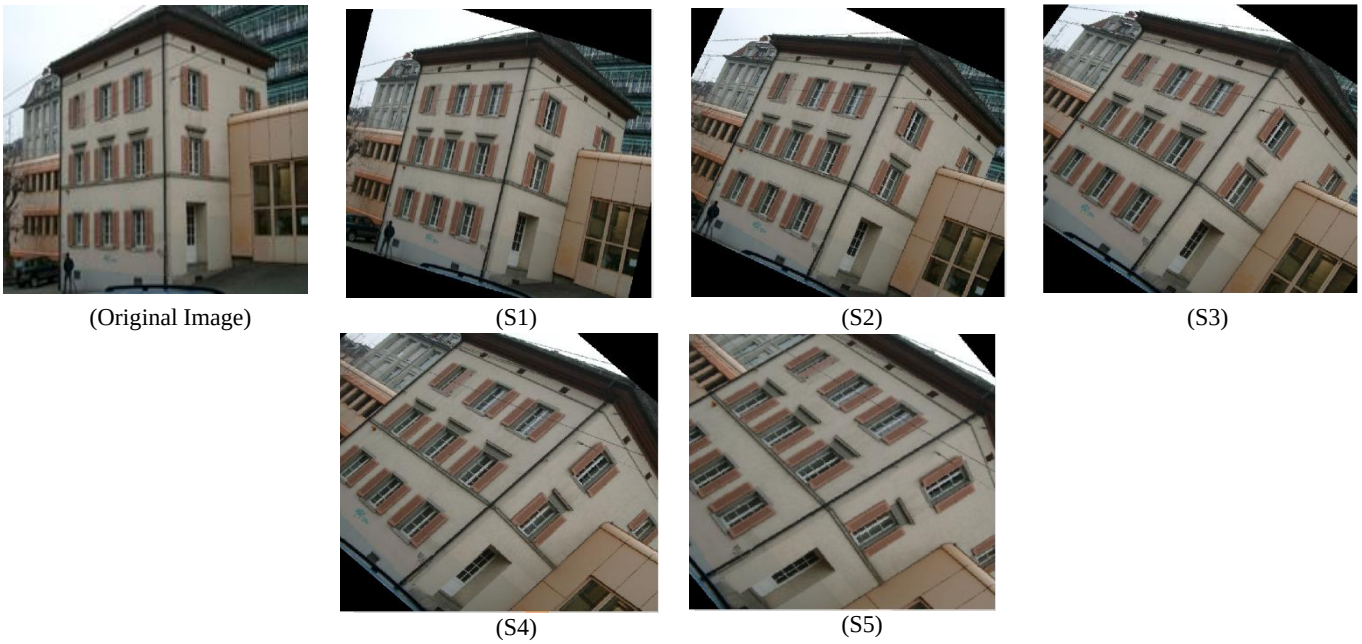


Fig. 4. The synthetic image dataset: the original image and the generated images applying five different transformations (S1 to S5).

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3188931

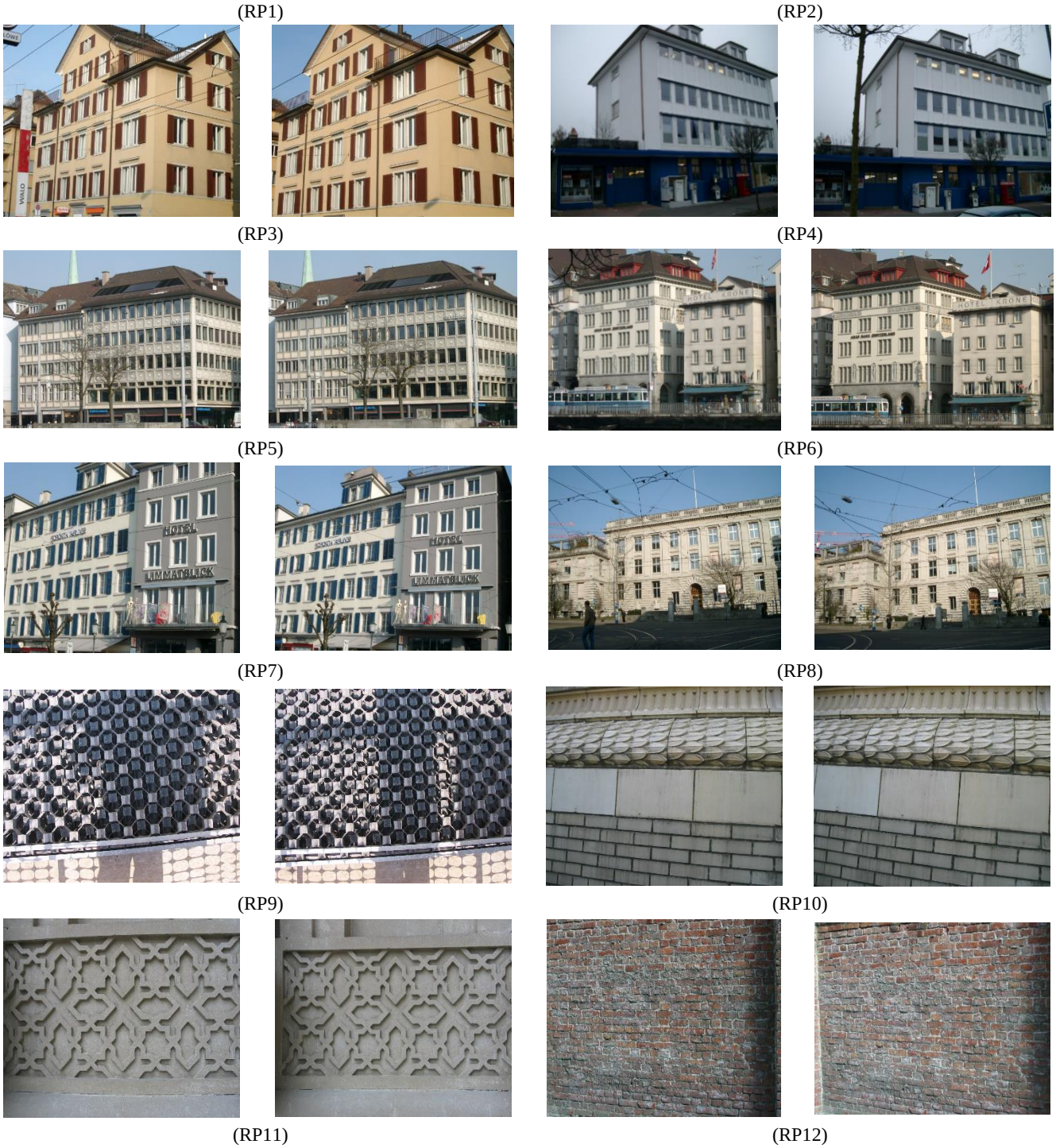


Fig. 5. The employed real image pairs datasets

Moreover, five multi-view image blocks were used, which were captured with different cameras at different locations (Fig. 6 and Table 1). These datasets are characterized by different image resolutions, varying overlap and a number of images.

TABLE I
THE SPECIFICATIONS OF THE MULTI-VIEW IMAGE DATASETS

Dataset	No. of Images	Camera Model	Sensor size (mm)	Resolution (pixel)	Pixel size (μm)	Focal length (mm)
RMV1	6	Canon Power Shot SX50 HS	6.17×4.55	4000×3000	1.50	4.3
RMV 2	10	Canon EOS 30D	22.5×15	3504×2336	6.41	28.0

RMV 3	27	KODAK M590	6.23×4.68	2880×2160	2.16	6.4
RMV 4	29	KODAK M590	6.23×4.68	2880×2160	2.16	6.4
RMV 5	25	KODAK M590	6.23×4.68	2880×2160	2.16	6.4

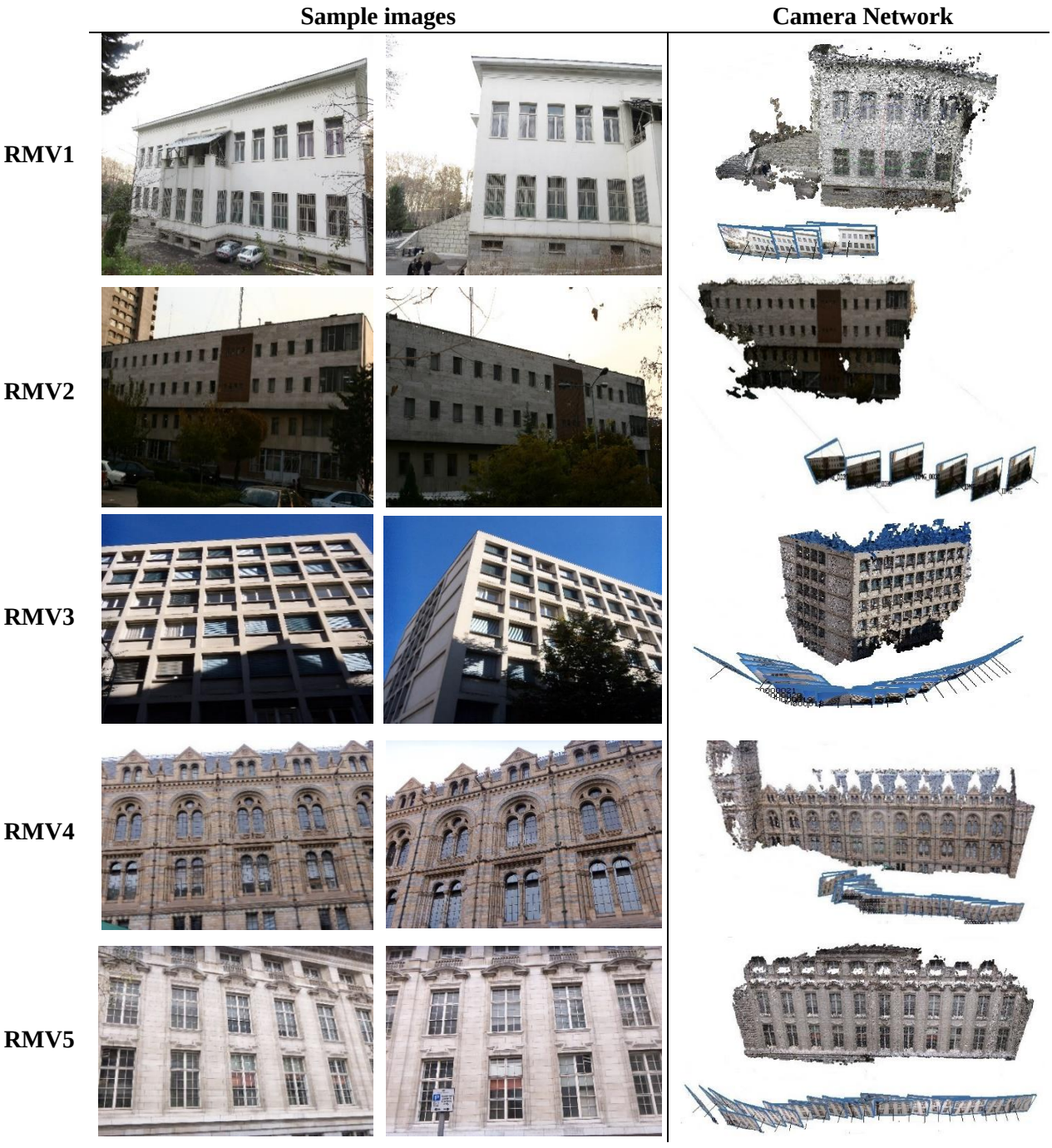


Fig. 6. Real multi-view datasets with their sample images (left) and the corresponding camera network (right)

VI. Experimental Results and Discussion

This section presents the results of evaluating the proposed method's performance on synthetic and real image datasets. To produce these results, SIFT detector was computed using the VLFeat toolbox [68] and SURF, DAISY and LIOP descriptors were implemented in MATLAB (R2017b). The CORE algorithm is also implemented in python. For the image block analysis, the open-source toolbox DBAT was used to run the photogrammetric image orientation process in MATLAB,

applying the Levenberg-Marquardt algorithm.

In order to perform the tests, the CR thresholds/parameters regarding each of the above descriptors need to be set up. For this, we performed a parameter analysis. In the following, first, this parameter analysis is described. Then, the results of each test on synthetic and real images are shown and discussed.

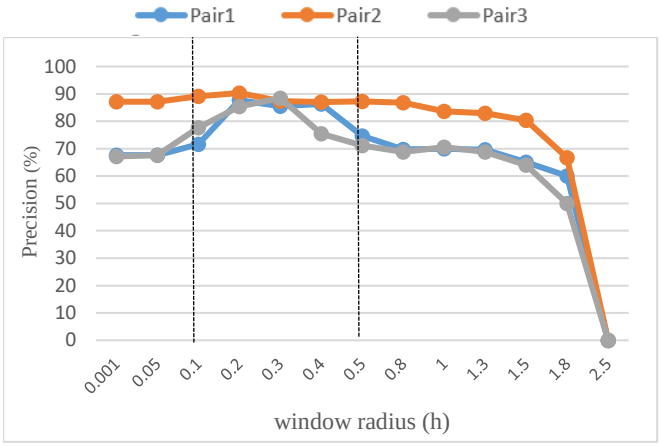
A) Parameter Analysis:

There are three main parameters in our algorithm: window radius (h), clusters Euclidean distance (d) and size of the regular grid cells (g), which need to be set. This section describes how we did this. The parameter study and sensitivity analysis were performed on three real image pairs. Three independent experiments were designed to learn parameters h, d, and g for both SIFT and SURF descriptors. In each experiment, only one parameter was considered as a variable, with the others fixed. The experimental setup details are summarized in Table II. For each parameter, precision and NCM are considered as the evaluation metrics. The experimental results are reported in Fig7~Fig9.

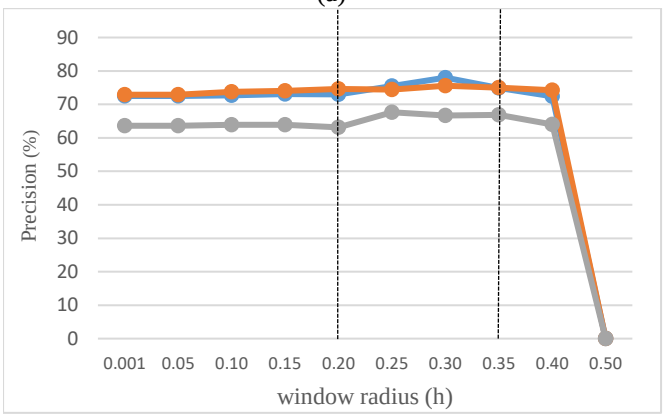
TABLE II: THE DETAILS OF PARAMETER SETTINGS FOR EXPERIMENTS

Experiment	Variable	Fixed parameters
Parameter *h*	**Sift:** h=[0.001,0.05,0.1,0.2,0.3,0.4,0.5,0.8,1,1.3,1.5,1.8,2.5] **Surf:** h=[0.001,0.05,0.1,0.15.0.2,0.25,0.30,0.350.40,0.50]	Sift: d=0.7 Surf: d=0.02 g=40pixels
Parameter *d*	**Sift**: d=[0.1,0.3,0.5,0.7,0.9,1.1,1.3,1.5,1.7,1.9] **Surf:** d=[0.009,0.01,0.02,0.03,0.04,0.05,0.06,0.07,0.08,0.1]	Sift: h=0.3 Surf: h=0.3 g=40pixels
Parameter *g*	g=[10,20,40,60,80,100,120]	Sift: d=0.7 h=0.3 Surf: d=0.02 h=0.3

1) Window radius (h): as can be seen in Fig 7, the results of mean-shift clustering is strongly affected by kernel window radius. Depending on the window radius, the resulting clusters could be quite different. An extremely small radius (smaller than 0.1 for SIFT and 0.2 for SURF) will result in each point having its own cluster. So no keypoints are removed in the removal step, and the results of image matching do not change. On the other hand, a large value for window radius (larger than 1.5 for SIFT and 0.4 for SURF) will result in a limited number of the cluster containing the data points. Therefore, a larger number of keypoints are removed in the removal step, and image matching results significantly degrade. Therefore, a suitable window radius parameter is very important. Considering these results, which are almost similar in different datasets, we set h around 0.3 for both SIFT and SURF.



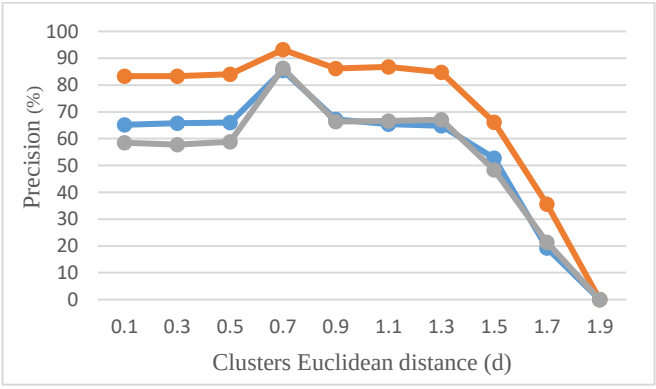
(a)



(b)

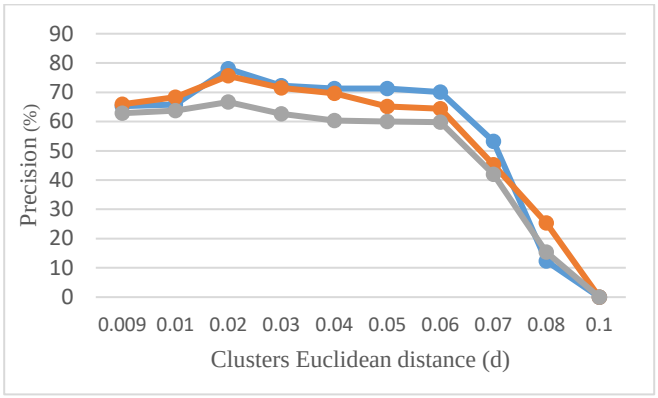
Fig. 7. The results of window radius parameter, (a): SIFT, (b): SURF

2) Clusters Euclidean distance (d): This parameter computes the similarity between cluster centers and can separate clusters that contain repetitive pattern descriptors. A large threshold for clusters Euclidean distance (larger than 1.3 for SIFT and 0.06 for SURF) removes more clusters and, thus, will result in fewer NCM. On the other hand, a small threshold will remove a limited number of clusters and does not improve the precision. Therefore, we set d to 0.7 for SIFT and 0.02 for SURF.



(a)

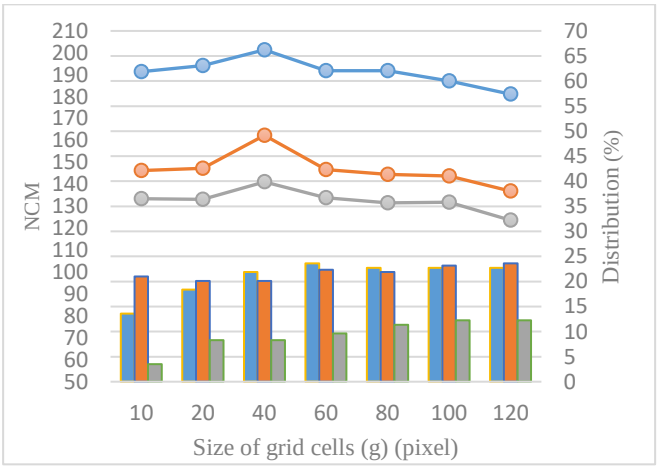
This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3188931



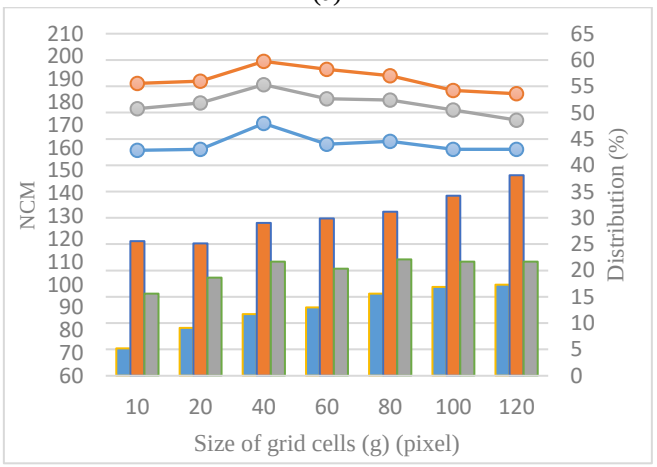
(b)

Fig. 8. The results of clusters Euclidean distance, (a): SIFT, (b): SURF

.



(a)



(b)

Fig. 9. The results of grid cell size, (a): SIFT, (b): SURF

3) Size of grid cells(g): this parameter indicates the size of regular gridding which is applied to achieve an even distribution of keypoints in the spatial space. As our experiments show in Fig9, if the grid cell is too small (smaller than 20 pixels), the NCM is relatively low. In contrast, if the grid cell is large, NCM is relatively higher but the distribution of selected keypoints degrades. Therefore, to take into account both the matching performance and distribution of keypoints in CR, we set *g* to 40 pixels.

Similar experiments were also conducted for other descriptors, and parameters were also set for them. Table III summarizes all parameters set.

TABLE III. THE PARAMETERS SET FOR EACH DESCRIPTOR IN CR METHOD

	SIFT	SURF	DAISY	LIOP
Clusters Euclidean distance *(d)*	0.7	0.02	1.6	0.4
Window Radius *(h)*	0.3	0.3	0.5	0.3
Regular grid Size (*g*) (Pixel)	40	40	40	40

B. Evaluations using synthetic images

This section describes the proposed method's capabilities for five levels of geometric transformation. Fig.10 shows the experiment's precision, recall, RMSE, and α for all five synthetic image cases and descriptors. As can be seen, the DAISY descriptor is not invariant to scale and rotation changes and thus fails as the scale and rotation values in images change. In all other cases, the proposed CR algorithm outperforms the original SIFT, SURF, DAISY, and LIOP descriptors for precision and recall criteria in all levels of transformations.

As shown in Fig.10, the capability of all of the descriptors is degraded with increasing in the geometric difference level, especially for DAISY descriptor, which is not invariant to scale and rotation changes. As shown in Fig. 10 (a), when the proposed CR algorithm is used, the performance of descriptors increases in terms of precision and recall criteria. Selecting keypoints based on the CR method enhances the performance of SIFT descriptor with an average increase of 12% in terms of precision. The average precision increase for CR-DAISY, CR-SURF and CR-LIOP descriptors are 9.4%, 10,4% and 9.25%, respectively.

As shown in Fig.10(b), the applied filtering method outperforms the original descriptor in all experiments. The recall results for the CR-DAISY and CR-LIOP descriptors are improved around 6% and 11%, respectively comparing to the original ones. CR-SIFT descriptor performs the best, with an average recall of 96.2%.

As illustrated in Fig.10(c), the average RMSE value of the matched keypoints for all descriptors is very close together; however, the accuracy of the CR-DAISY is slightly better than the other methods. Therefore, we can reveal that the positional accuracy of the matched keypoints extremely depends on the type of detector applied to extract keypoints, and descriptors are not specifically effective in this regard.

As shown in Fig.10(d), the spatial distribution of the selected matched keypoints confirms the capability of the proposed method to find well-distributed matched points. It should be noted that although the proposed CR algorithm removes highly confusing keypoints at the first step and selects a subset of high-quality keypoints from the remaining. However, the distribution of the selected keypoint is not degraded in comparison to the conventional method. The results indicate that the distribution

of the selected keypoints is averagely 2.21% decreased when the proposed CR algorithm is applied to the dataset.

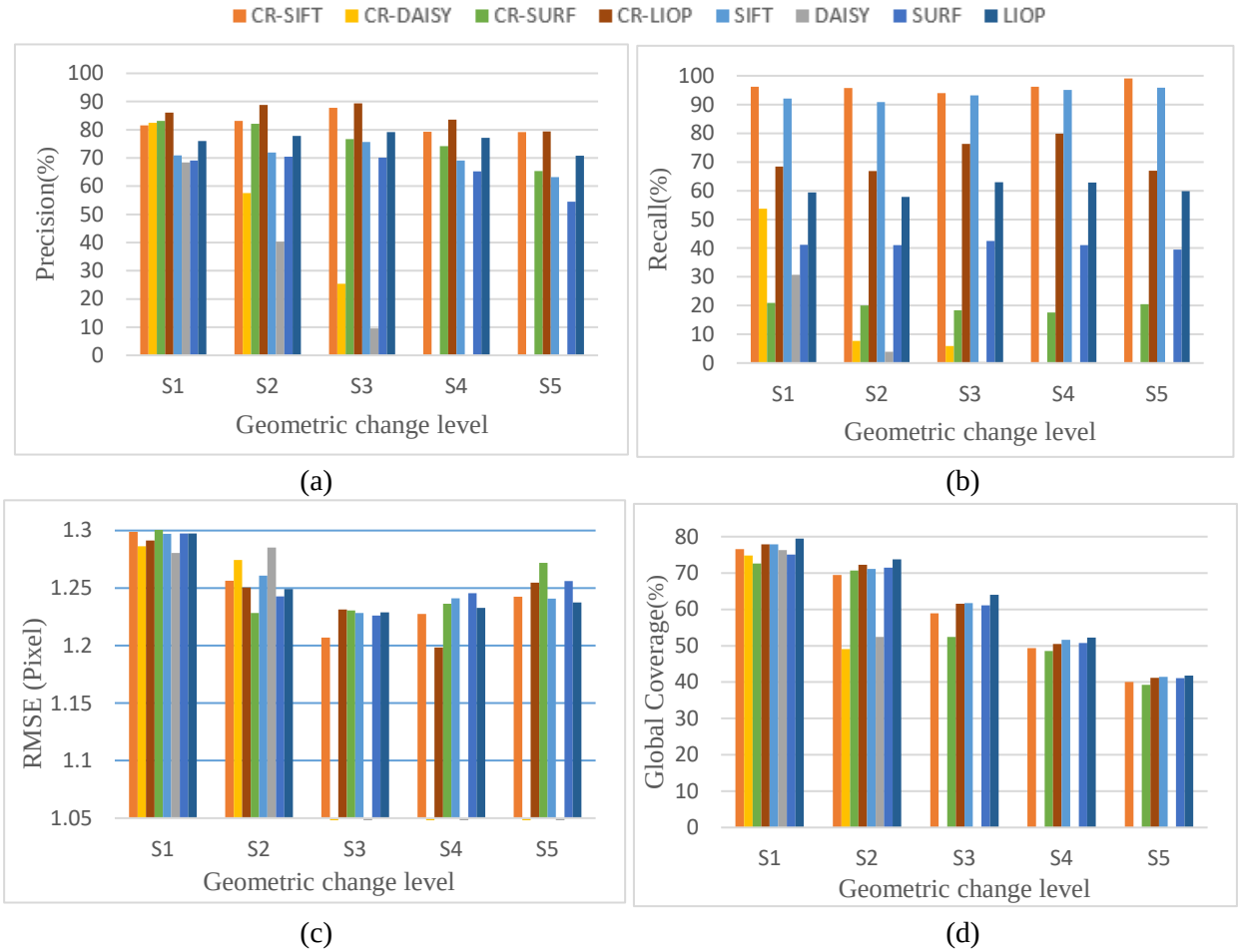
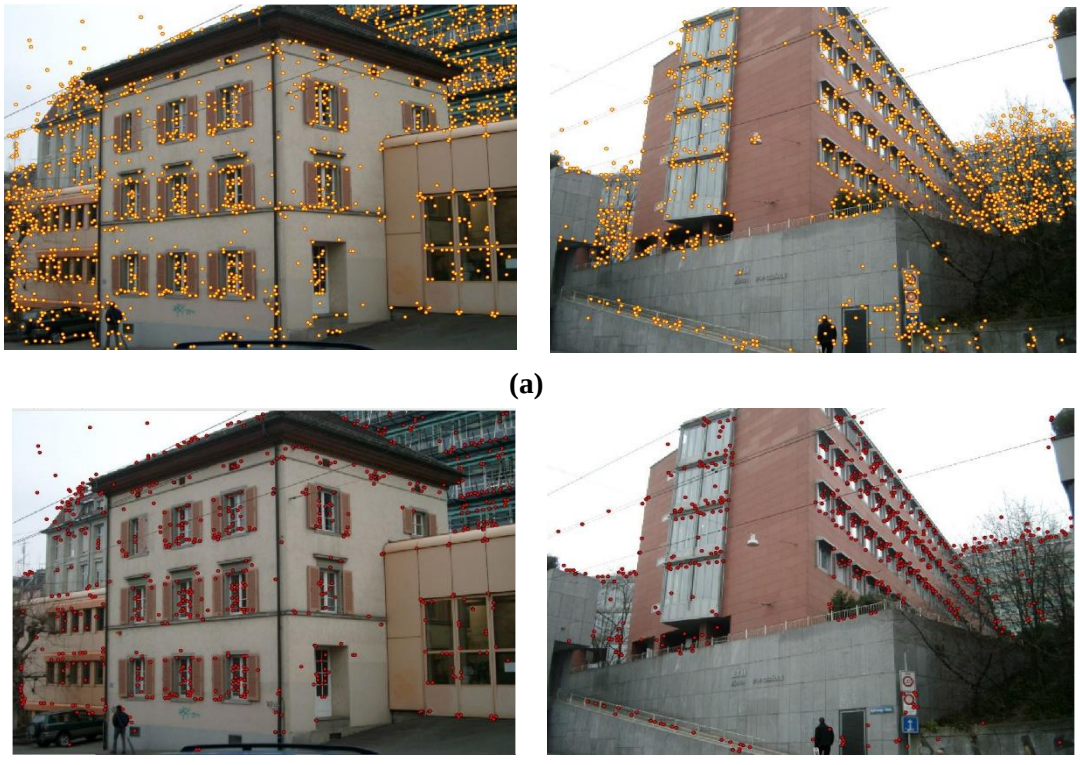


Fig. 10. Matching results for different descriptors in synthetic dataset: (**a**) Precision, (**b**) Recall, (**c**) RMSE and (**d**) Global coverage.



(a)

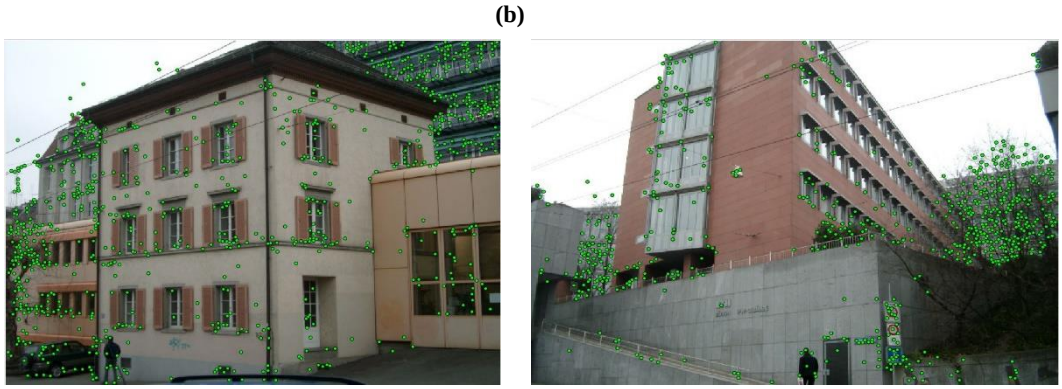


Fig. 11. Examples of the proposed filtering results in real image pairs: (a) initial extracted keypoints (in yellow), (b) removed keypoints (in red), (c) kept keypoints (in green).

C. Evaluations using real images

The performance evaluation of the proposed CR method on real image datasets is presented in this section. In the first part of the experiments, the results of the CR method against the original descriptor and the CORE algorithm are reported. Then the performance of our proposed method is compared with other mismatching removal methods.

Prior to discussing the results, Fig.11 illustrates a direct application of the proposed CR filtering algorithm for SIFT detector and descriptor. The initial extracted keypoints are shown in yellow. Keypoints that have been removed from images are highlighted in red, while the ones that have been selected are highlighted in green. The results indicate that the majority of keypoints (in red) that have been removed are located in areas with repetitive patterns.

It can also be noted that despite the removal of keypoints in the sky using the CR filtering algorithm, some of them still are remained. This is because these points are not eliminated during the removal process but are retained during the selection process to maintain the keypoint distribution. However, as can be seen, the CR method still has outperformed SIFT in areas with homogeneity in texture. This is an important issue that needs to be tackled in future studies.

1) Comparison with CORE algorithm:

Comparisons of the results for three different approaches of original SIFT descriptor, SIFT+CORE algorithm and CR-SIFT algorithm are summarized in Table IV (left columns). Furthermore, the results for three different approaches of original SURF, SURF+CORE algorithm and CR-SURF method are also illustrated in Table IV (right columns). The best results for each image pair are highlighted.

As it is shown in Tables IV, the proposed reduction method globally improves the correct matching ratio in almost all experiments. Comparing the precision results of CR-SIFT algorithm to the original SIFT descriptor, an increased average precision value of 17.8% and recall of 10.2% for the real image datasets is obtained. Similarly, the CR-SIFT algorithm outperforms the CORE algorithm with an average precision and recall increase of 15.3% and 12.8% respectively.

Findings in Table IV show that the CR-SURF algorithm significantly outperforms SURF and SURF+CORE algorithms in all real image cases. Furthermore, the matching results on the real image cases are superior to the other methods. As can be seen, the CR-SURF algorithm with an average increase of 5.0% in precision and 20% for recall outperforms the original SURF. Similarly, the proposed method with an average increase of 6.3% in precision and 30.23% for recall value surpasses the SURF+CORE algorithms.

The results of the spatial distribution (α) analysis in Tables IV confirm that keypoint removal and selection processes in the proposed CR algorithm do not negatively influence the distribution of the matched keypoints. As can be seen, for the CR-SIFT algorithm, the distribution of matched keypoints is almost equivalent to the original SIFT results, with an average decrease of 2% to 5%. Similarly, the distribution of matched keypoints using the CR-SURF algorithm slightly differ from the original SURF algorithm.

Based on the experimental results, it can be seen that the proposed CR algorithm increases matching performance and improves spatial distribution. The results of α indicate that the distribution of the selected keypoints is effectively controlled during the selection process, along with a better precision performance of the proposed keypoint filtering method. Filtering the confusing keypoints and selecting a subset of well-distributed keypoints increases matching accuracy and robustness in images with repetitive patterns effectively.

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3188931

TABLE IV
EVALUATING PERFORMANCE OF CR METHOD AGAINST THAT OF ORIGINAL DESCRIPTOR AND CORE ALGORITHM

Dataset		Original SIFT descriptor	SIFT descriptor +CORE	CR-SIFT	Original SURF descriptor	SURF descriptor +CORE	CR-SURF
RP1	NCM	133	108	100	108	52	78
	Precision (%)	65.04	62.67	85.47	72.48	71.23	78.00
	Recall (%)	54.06	62.06	65.78	66.10	45.44	96.18
	α (%)	65.72	66.06	66.23	49.01	42.07	47.95
RP2	NCM	114	90	96	148	71	127
	Precision (%)	79.17	76.92	93.20	72.90	63.39	75.59
	Recall (%)	77.43	78.31	81.35	71.43	23.96	86.39
	α (%)	49.52	49.16	49.16	64.93	61.80	59.74
RP3	NCM	94	127	69	154	56	110
	Precision (%)	56.97	80.37	88.46	63.63	51.37	66.66
	Recall (%)	42.35	53.21	58.47	74.76	21.27	82.91
	α (%)	41.40	61.42	39.87	61.28	58.49	55.31
RP4	NCM	113	69	113	136	47	86
	Precision (%)	63.48	71.13	86.25	53.12	77.14	74.78
	Recall (%)	52.36	62.41	65.69	73.37	36.72	91.57
	α (%)	78.30	75.69	78.40	69.17	66.28	68.87
RP5	NCM	435	307	291	342	254	287
	Precision (%)	76.72	78.92	97.32	63.92	63.65	64.78
	Recall (%)	68.35	72.74	77.64	53.54	35.48	77.50
	α (%)	68.49	67.57	67.13	71.50	68.04	69.50
RP6	NCM	812	489	462	639	453	509
	Precision (%)	89.53	90.89	97.88	86.35	82.06	87.75
	Recall (%)	82.14	85.13	91.18	76.96	80.32	96.65
	α (%)	76.28	73.75	77.21	72.40	74.72	72.15
RP7	NCM	449	69	357	370	342	300
	Precision (%)	82.69	71.13	97.54	77.24	78.28	83.57
	Recall (%)	78.36	80.42	83.71	36.17	53.15	69.33
	α (%)	79.87	75.69	79.84	76.61	74.35	72.04
RP8	NCM	310	282	202	153	125	137
	Precision (%)	83.33	85.47	95.28	67.70	73.65	77.84
	Recall (%)	71.35	75.32	87.44	32.62	43.58	59.21
	α (%)	43.94	42.38	40.15	52.20	50.25	51.52
RP9	NCM	35	35	37	17	15	13
	Precision (%)	12.28	24.38	32.45	8.46	7.35	18.68
	Recall (%)	33.25	36.19	42.04	19.77	20.42	25.12
	α (%)	47.03	40.58	47.83	33.90	32.28	37.12
RP10	NCM	12	13	15	1436	1258	744
	Precision (%)	7.36	2.34	41.66	95.23	93.57	95.75
	Recall (%)	25.41	27.52	37.50	39.14	53.47	72.52
	α (%)	49.33	50.32	54.65	85.70	81.76	81.92
RP11	NCM	459	405	397	587	325	284
	Precision (%)	83.91	86.50	98.75	93.32	92.37	98.95
	Recall (%)	79.24	74.28	86.68	44.70	57.85	75.15
	α (%)	78.67	77.21	76.55	78.00	71.37	75.68
RP12	NCM	1143	753	803	318	217	104
	Precision (%)	99.48	99.21	99.62	77.18	71.35	78.82
	Recall (%)	90.28	93.24	99.50	75.00	81.27	83.20
	α (%)	75.34	73.43	75.93	82.22	78.69	75.39

2) Comparison with other mismatch removal algorithms:

In this section, the performance of the proposed CR method is tested and compared with other feature matching methods.

The number of inliers, precision and recall statistics, and RMSE and spatial distribution of the five algorithms are reported in Table V, and the best results are shown in **bold**. As it shows, the proposed CR algorithm can effectively remove the mismatches before the matching stage and achieve an effective result even when the repeating areas are covered all over the image and the image quality is low.

The precision and recall of the proposed method have always been the highest in all experiments. Using our proposed CR method to filter out the mismatches, the highest precision and recall are achieved. In addition, when there are a few outliers in the initial matching set, like in RP1 to RP8, all algorithms can achieve good results. However, when there are many outliers, such as in RP 9 and RP 10, the results of other methods will deteriorate sharply. The accuracy of RANSAC will decrease to below 15% because it estimates the transformation matrix and is efficient with a large number of outliers. It is worth noting that although the accuracy and recall of RANSAC are not as well as other methods, its RMSE is smaller than others.

The LPM and LMR methods are both based on the spatial consistency among the putative matches. The accuracy of LPM and LMR is not satisfactory in images with lots of outlier points (RP 9 and RP10) because these methods are very suitable for the situation with a few outlier points and sensitive to a large proportion of outliers; and thus, the neighborhood construction will be unreliable.

In the case of high outliers, the recall of RFMSCAN is better than other algorithms. The reason is that as the outlier ratio increases, a small part of outliers may have weak motion consistency and then form one or more false inlier clusters, leading to a decrease in precision. In contrast, the inliers in general, always have motion consistency that is seldom affected by outliers, and hence it can achieve a large recall even in the case of a large outlier ratio.

From the statistic comparisons of LMR and GLOF in the above experiments, LMR outperforms many tested approaches since it is a learning-based matching method with many powerful machine learning models such as neural networks, and the trained classifier can get a good generalization ability even facing the complex image transformations in the training data.

TABLE COMPARISON OF THE NUMBER OF CORRECT MATCHES(NCM), PRECISION, RECALL, RMSE AND SPATIAL DISTRIBUTION OF THE TEST DATASETS

Dataset		RANSAC	LPM	LMR	RFMSCAN	GLOF	CR-SIFT
RP1	NCM	11	130	136	**160**	159	100
	Precision (%)	78.57	83.33	85	66.67	72.27	**85.47**
	Recall (%)	4.47	52.84	55.28	65.04	64.63	**65.78**
	RMSE (pix)	1.11	**1.059**	1.03	1.06	1.06	1.15
	α (%)	4.78	63.53	65.33	65.72	65.72	**66.23**
RP2	NCM	31	113	111	**114**	108	96
	Precision (%)	86.11	88.98	90.24	80.28	92.31	**93.20**
	Recall (%)	21.53	78.47	77.08	79.17	75.00	**81.35**
	RMSE (pix)	**0.45**	0.53	0.51	0.54	0.54	0.55
	α (%)	26.35	49.52	49.52	49.52	41.28	49.16
RP3	NCM	20	77	83	**94**	93	69
	Precision (%)	86.96	87.50	88.30	63.09	83.78	**88.46**
	Recall (%)	12.12	46.67	50.30	56.97	56.36	**58.47**
	RMSE (pix)	1.06	**1.03**	1.04	1.06	1.07	1.09
	α (%)	13.26	24.82	35.12	41.39	39.17	39.87
RP4	NCM	36	90	85	**113**	105	**113**
	Precision (%)	92.31	78.26	85.86	66.86	71.92	**86.25**
	Recall (%)	20.22	50.56	47.75	63.48	58.99	**65.69**
	RMSE (pix)	**0.59**	0.87	0.89	0.86	0.85	0.86
	α (%)	62.21	78.09	78.25	78.30	78.09	**78.40**
RP5	NCM	227	422	414	**435**	416	291
	Precision (%)	89.37	93.57	95.83	77.96	87.95	**97.32**
	Recall (%)	40.04	74.43	73.02	76.72	73.37	**77.64**
	RMSE (pix)	**0.30**	0.46	0.45	0.47	0.45	0.48
	α (%)	56.54	64.80	66.07	68.49	67.06	**67.13**
RP6	NCM	479	796	805	**812**	801	462
	Precision (%)	96.96	96.72	96.64	89.82	97.33	**97.88**
	Recall (%)	52.81	87.76	88.75	89.53	88.31	**91.18**
	RMSE (pix)	**0.89**	0.95	0.95	0.95	0.95	0.92
	α (%)	68.99	74.61	76.25	76.28	75.67	**77.21**
RP7	NCM	159	429	435	**449**	434	357
	Precision (%)	97.55	97.72	97.53	90.16	95.81	**97.54**

	Recall (%)	29.28	79.01	80.11	82.69	79.93	**83.71**
	RMSE (pix)	**0.22**	0.44	0.43	0.47	0.43	0.47
	α (%)	70.85	79.32	79.68	79.77	79.77	**79.84**
RP8	NCM	134	299	295	**310**	284	202
	Precision (%)	93.05	94.32	93.65	88.57	93.73	**95.28**
	Recall (%)	36.02	80.38	79.30	83.33	76.34	**87.44**
	RMSE (pix)	**0.31**	0.43	0.43	0.46	0.41	0.43
	α (%)	36.95	43.62	43.24	**43.94**	41.21	40.15
RP9	NCM	14	24	22	45	43	**47**
	Precision (%)	13.34	18.18	14.81	13.01	16.02	**32.45**
	Recall (%)	1.40	4.91	4.21	12.28	11.58	**42.04**
	RMSE (pix)	**0.60**	1.27	1.01	1.13	1.15	1.07
	α (%)	5.86	11.75	8.51	47.03	44.07	**47.83**
RP10	NCM	13	17	17	22	21	**25**
	Precision (%)	9.67	31.82	17.95	8.05	13.25	**41.66**
	Recall (%)	0.61	4.29	4.29	7.36	6.75	**37.50**
	RMSE (pix)	2.67	**1.38**	1.59	1.49	1.41	1.27
	α (%)	12.47	22.52	28.28	49.33	49.33	**54.65**
RP11	NCM	249	453	457	**459**	456	397
	Precision (%)	96.13	97.00	97.03	87.93	97.02	**98.75**
	Recall (%)	45.52	82.82	83.55	83.91	83.36	**86.68**
	RMSE (pix)	**0.67**	0.90	0.89	0.89	0.89	0.91
	α (%)	78.32	78.67	78.47	78.57	**78.66**	76.55
RP12	NCM	1044	1134	**1145**	1062	1142	803
	Precision (%)	99.42	99.47	99.50	99.44	99.48	**99.62**
	Recall (%)	90.86	98.69	99.21	92.43	99.39	**99.50**
	RMSE (pix)	0.35	0.36	0.33	**0.33**	0.37	0.37
	α (%)	74.78	75.02	75.85	66.23	75.34	**75.93**

3) Results of image block orientation

In this section, the proposed CR method is applied to blocks of real images. As the experiments in previous sections have shown, the number of extracted keypoints is decreased using the proposed keypoint filtering method. The reduction of keypoints can lead to either orientation failure due to the inadequate number of tie points or incorrect orientation results. Therefore, the experiments in this section evaluate the impact of the CR algorithm on the results of the image orientation as follows.

1) Average re-projection error of the bundle adjustment:

The re-projection error of all computed 3D points is shown in Fig.12(a) for all datasets. This metric is not only affected by the matching accuracy but also by the accuracy of the external parameters. As shown, the proposed CR algorithm has a better performance compared to the original SIFT in each dataset. The proposed CR-SIFT algorithm averagely decreases 20% to 30% of the re-projection error. However, the re-projection error in the proposed CR-SIFT method is still large. Although the CR-SIFT tries to select well-distributed keypoints, the selected keypoints are still at large scales, leading to lower spatial resolution. The lower spatial resolution of the keypoints causes large re-projection errors in the bundle adjustment.

2) Average angles of intersection:

Since 3D points are calculated by triangulation, a higher angle of intersection of similar rays provides more accurate 3D details. Fig.12(b) shows that the intersection angles in the R2 dataset do not change significantly using the proposed CR algorithm; however, an average improvement of 14% in the intersection angles is achieved. Thus, for R1 and R4 datasets with higher overlapping images, the average intersection angles are relatively smaller. In this respect, the CR-SIFT is slightly performing better because the keypoints are better distributed.

3) Average rays per 3D point:

As the number of images and their overlap increases in a dataset, more accurate 3D object coordinates are expected. As shown in Fig.12(c), higher average multiplicity for the tie-points is achieved in the R4 dataset with higher overlapping images. The larger multiplicity belongs to the original SIFT for all datasets. However, the average multiplicity of the proposed CR algorithm, despite keypoint removal in the process, is close to the original SIFT with an average of 2% to 4% decrease.

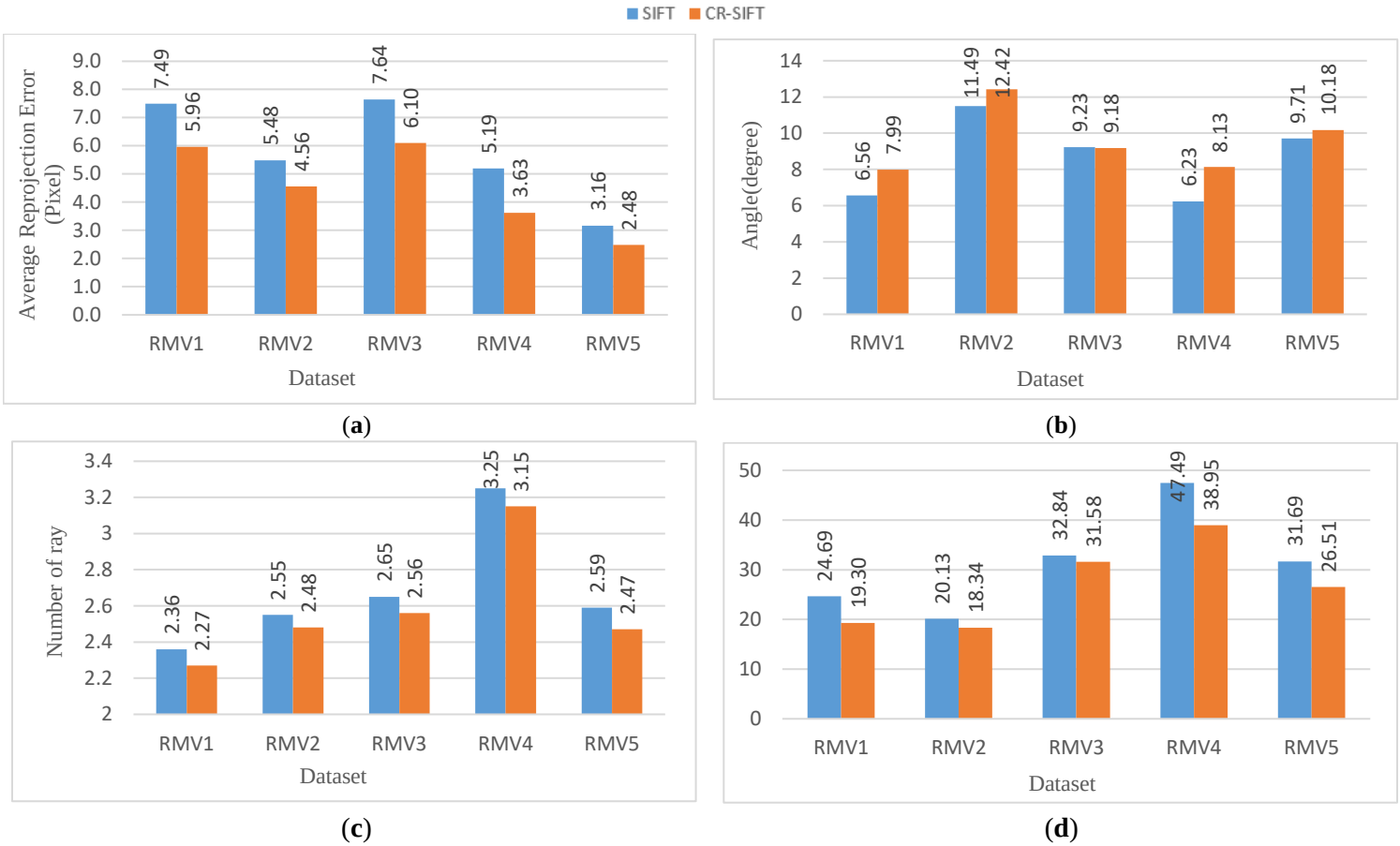


Fig. 12. Results for the real image block orientation. (a): Re-projection error of bundle adjustment for each dataset; (b): Average intersection angles; (c): Average rays per computed 3D points; (d): The visibility of the derived 3D points in more than 3 images.

4) Visibility of 3D points in more than 3 images:
Fig.12(d) shows the visibility of 3D points that are normalized with respect to all extracted points for each dataset. The visibility results indicate that an average of 30% of the triangulate points using the SIFT are visible in three images in all datasets. The proposed CR-SIFT has slightly weaker performance in this regard, with an average of 27% of visible points in more than three images.

As can be seen, our CR algorithm outperforms the original SIFT in terms of average re-projection error and average intersection angles but has a lower performance in visibility and average ray issues. However, it should be noted such issues are not as important as the first two issues. This is because average re-projection error and average intersection angles play an important role in defining the geometry of the image network, whereas visibility of 3D points in more than three images and average rays per 3D point only refers to the number of points. Obviously, compared to the geometry of the imaging network, this is less important, and thus, its value is of less concern.

D. Computational cost

The proposed CR algorithm seems to imply a significant computational cost to the matching process. Since the mean-shift clustering is used in the first step of the proposed algorithm, it can be computationally expensive for a large number of keypoints because it needs to iteratively follow the procedures for each descriptor vector in a given image. Therefore, a straightforward implementation of mean-shift should have a complexity of $O(N^2)$, where N is the number of keypoints in the image. Fig.13 shows the computational cost of the proposed algorithm with respect to the number of keypoints. The experiments were performed on a PC with a 2.6 GHz Intel Core i5-3230 processor and 6GB of RAM. As shown in Fig.13, the mean-shift algorithm used in the proposed method is very computation-intensive. It is clear that the processing time will grow quadratically as the number of points to be processed increases. However, since the removal stage of CR method is very suitable for parallel computing, an implementation based on parallel computing on a GPU architecture could significantly reduce the time complexity.

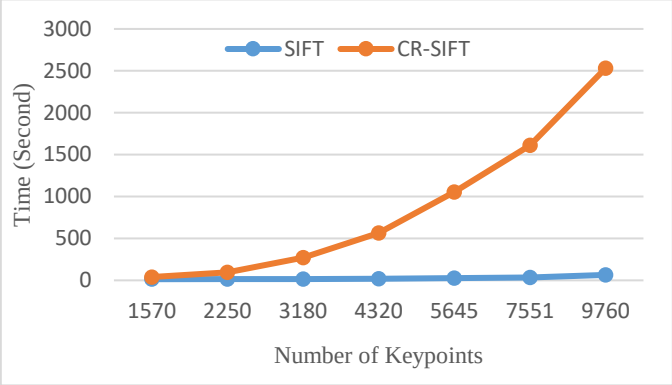


Fig. 13. The computation cost of the proposed approach by the number of keypoints

VII. Conclusion

Image matching is a critical task in a wide variety of photogrammetry applications [69]-[71]. This study introduced a novel filtering algorithm to address the keypoints confusion problem. By removing highly confusing keypoints, the proposed algorithm extracts a smaller subset of initial keypoints that are less prone to confusion. The mean-shift clustering algorithm and a novel confusion index with a gridding strategy are used to ensure the quality of the selected keypoints. The proposed method was implemented and compared with four conventional descriptors: SIFT, SURF, DAISY, and LIOP and also extensive experiments on different real image pairs were performed using several state-of-the-art mismatch rejection methods. The proposed method resulted in more discriminant keypoint subsets than the original descriptors and the CORE algorithm. Better results are obtained when compared with that of several other popular methods, especially in terms of robustness to outliers. Additionally, the proposed CR algorithm outperformed the CORE algorithm in nearly all experiments and significantly improved matching accuracy and robustness through the use of evenly distributed keypoints. Furthermore, compared to the state-of-the-art mismatch rejection methods, the proposed CR algorithm can effectively remove the mismatches before the matching stage and achieve an effective result even when the image quality is low and repeating areas are covered all over the image. As our experiments show, the thresholds in our method are independent of the image context. Besides, the results of the newly developed method on the accuracy of multi-view image pose estimation were compared to those of the conventional SIFT algorithm, which demonstrated an average improvement of 20% to 30% in image bundle adjustment results.

We will investigate how the proposed algorithm can be coupled to binary descriptors such as BRIEF in future work. Furthermore, the performance of the CR method in areas with homogeneity in texture (like sky) should be deeply analyzed. Additionally, the proposed method makes use of the mean-shift clustering with a significant time complexity; further research to speed up the clustering strategy based on parallel computing implementation is suggested as additional research work.

Acknowledgement

The authors appreciate the Faculty of Geomatics Engineering, K.N. Toosi University of Technology, Tehran, Iran, 3D Optical Metrology Unit, Bruno Kessler Foundation, Trento, Italy, and the Geospatial Sensing and Data Intelligence Lab, Department of Geography and Environmental Management, University of Waterloo, Canada, for collaboration and technical research support to the GeoAI, Smarter Map and LiDAR Lab of the Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University (SWJTU). This work is the outcome of one of the joint research studies with international University collaborators to support SDG-17 and the SWJTU internationalization.

References

[1] V. Mousavi *et al.*, "The performance evaluation of multi-image 3D reconstruction software with different sensors," *Meas. J. Int. Meas. Confed.*, vol. 120, pp. 1–10, 2018, doi: 10.1016/j.measurement.2018.01.058.

[2] W. Hartmann, M. Havlena, and K. Schindler, "Recent developments in large-scale tie-point matching," *ISPRS J. Photogramm. Remote Sens.*, vol. 115, pp. 47–62, 2016.

[3] A. Sedaghat and N. Mohammadi, "High-resolution image registration based on improved SURF detector and localized GTM," *Int. J. Remote Sens.*, vol. 40, no. 7, pp. 2576–2601, 2019, doi: 10.1080/01431161.2018.1528402.

[4] J. Xing, R. Sieber, and T. Caelli, "A scale-invariant change detection method for land use/cover change research," *ISPRS J. Photogramm. Remote Sens.*, vol. 141, pp. 252–264, 2018, doi: 10.1016/j.isprsjprs.2018.04.013.

[5] Z. Shangguan, L. Wang, J. Zhang, and W. Dong, "Vision-based object recognition and precise localization for space body control," *Int. J. Aerosp. Eng.*, vol. 2019, 2019, doi: 10.1155/2019/7050915.

[6] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image Matching from Handcrafted to Deep Features: A Survey," *Int. J. Comput. Vis.*, vol. 129, no. 1, pp. 23–79, 2021, doi: 10.1007/s11263-020-01359-2.

[7] F. Remondino, F. Menna, and L. Morelli, "Evaluating Hand-Crafted and Learning-Based Features for Photogrammetric Applications," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLIII-B2-2, pp. 549–556, 2021, doi: 10.5194/isprs-archives-xliii-b2-2021-549-2021.

[8] K. M. Yi, E. Trulls, V. Lepetit, and P. Fua, "LIFT: Learned invariant feature transform," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9910 LNCS, pp. 467–483, doi: 10.1007/978-3-319-46466-4_28.

[9] A. Mishchuk, D. Mishkin, F. Radenović, and J. Matas, "Working hard to know your neighbor's margins: Local descriptor learning loss," in *Advances in Neural Information Processing Systems*, 2017, vol. 2017-Decem, pp. 4827–4838.

[10] Y. Ono, P. Fua, E. Trulls, and K. M. Yi, "LF-Net: Learning local features from images," in *Advances in Neural Information Processing Systems*, 2018, vol. 2018-Decem, pp. 6234–6244.

[11] D. Detone, T. Malisiewicz, and A. Rabinovich, "SuperPoint: Self-supervised interest point detection and description," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2018, vol. 2018-June, pp. 337–349, doi: 10.1109/CVPRW.2018.00060.

[12] A. Bhowmik, S. Gumhold, C. Rother, and E. Brachmann, "Reinforced feature points: Optimizing feature detection and description for a high-level task," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4947–4956, 2020, doi: 10.1109/CVPR42600.2020.00500.

[13] T. Georgiou, Y. Liu, W. Chen, and M. Lew, "A survey of traditional and deep learning-based feature descriptors for high dimensional data in computer vision," *Int. J. Multimed. Inf. Retr.*, vol. 9, no. 3, pp. 135–170, 2020, doi: 10.1007/s13735-019-00183-w.

[14] E. Royer, T. Lelore, and F. Bouchara, "COnfusion REduction (CORE) algorithm for local descriptors, floating-point and binary cases," *Comput. Vis. Image Underst.*, vol. 158, pp. 115–125, 2017, doi: 10.1016/j.cviu.2017.01.005.

[15] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proceedings of the IEEE International Conference on Computer Vision*, 2011, pp. 2564–2571, doi: 10.1109/ICCV.2011.6126544.

[16] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary Robust invariant scalable keypoints," in *Proceedings of the IEEE International Conference on Computer Vision*, 2011, pp. 2548–2555, doi: 10.1109/ICCV.2011.6126542.

[17] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: Fast retina keypoint," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012, pp. 510–517, doi: 10.1109/CVPR.2012.6247715.

[18] V. Balntas, L. Tang, and K. Mikolajczyk, "BOLD - Binary online learned descriptor for efficient image matching," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015, vol. 07-12-June, pp. 2367–2375, doi: 10.1109/CVPR.2015.7298850.

[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004, doi: 10.1023/B:VISI.0000029664.99615.94.

[20] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, 2008, doi: 10.1016/j.cviu.2007.09.014.

[21] E. Tola, V. Lepetit, and P. Fua, "DAISY: An efficient dense descriptor applied to wide-baseline stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 815–830, 2010, doi: 10.1109/TPAMI.2009.77.

[22] Z. Wang, B. Fan, and F. Wu, "Local intensity order pattern for feature description," in *Proceedings of the IEEE International Conference on Computer Vision*, 2011, pp. 603–610, doi: 10.1109/ICCV.2011.6126294.

[23] A. Sedaghat and H. Ebadi, "Distinctive Order Based Self-Similarity descriptor for multi-sensor remote sensing image matching," *ISPRS J. Photogramm. Remote Sens.*, vol. 108, pp. 62–71, 2015, doi: 10.1016/j.isprsjprs.2015.06.003.

[24] A. Sedaghat and H. Ebadi, "Remote Sensing Image Matching Based on Adaptive Binning SIFT Descriptor," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5283–5293, 2015, doi: 10.1109/TGRS.2015.2420659.

[25] J. Li, Q. Hu, and M. Ai, "RIFT: Multi-Modal Image Matching Based on Radiation-Variation Insensitive Feature Transform," *IEEE Trans. Image Process.*, vol. 29, pp. 3296–3310, 2020, doi: 10.1109/TIP.2019.2959244.

[26] A. Sedaghat and N. Mohammadi, "Illumination-Robust remote sensing image matching based on oriented self-similarity," *ISPRS J. Photogramm. Remote Sens.*, vol. 153, pp. 21–35, 2019, doi: 10.1016/j.isprsjprs.2019.04.018.

[27] L. Lettry, M. Perdoch, K. Vanhoey, and L. Van Gool, "Repeated pattern detection using CNN activations," in *Proceedings - 2017 IEEE Winter Conference on Applications of Computer Vision, WACV 2017*, 2017, pp. 47–55, doi: 10.1109/WACV.2017.13.

[28] C. Rodriguez-Pardo, S. Suja, D. Pascual, J. Lopez-Moreno, and E. Garces, "Automatic extraction and synthesis of regular repeatable patterns," *Comput. Graph.*, vol. 83, pp. 33–41, 2019, doi: 10.1016/j.cag.2019.06.010.

[29] R. Huang, Y. Liu, Y. Zheng, and M. Ye, "Optical frequency and phase information-based fusion approach for image rotation symmetry detection," *Opt. Express*, vol. 28, no. 13, p. 18577, Jun. 2020, doi: 10.1364/oe.390224.

[30] Y. Wu *et al.*, "A Two-Step Method for Remote Sensing Images Registration Based on Local and Global Constraints," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 14, pp. 5194–5206, 2021, doi: 10.1109/JSTARS.2021.3079103.

[31] Y. Liu, R. T. Collins, and Y. Tsin, "A Computational Model for Periodic Pattern Perception Based on Frieze and Wallpaper Groups," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 3, pp. 354–371, 2004, doi: 10.1109/TPAMI.2004.1262332.

[32] G. Loy and J. O. Eklundh, "Detecting symmetry and symmetric constellations of features," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2006, vol. 3952 LNCS, pp. 508–521, doi: 10.1007/11744047_39.

[33] J. Pritts, O. Chum, and J. Matas, "Rectification, and segmentation of coplanar repeated patterns," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2973–2980, doi: 10.1109/CVPR.2014.380.

[34] H. Xiao, G. Meng, L. Wang, and C. Pan, "Facade repetition detection in a fronto-parallel view with fiducial lines extraction," *Neurocomputing*, vol. 273, pp. 435–447, 2018, doi: 10.1016/j.neucom.2017.07.040.

[35] J. Liu and Y. Liu, "GRASP recurring patterns from a single view," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2003–2010, 2013, doi: 10.1109/CVPR.2013.261.

[36] Y. Wu, W. Ma, Q. Su, S. Liu, and Y. Ge, "Remote sensing image registration based on local structural information and global constraint," *J. Appl. Remote Sens.*, vol. 13, no. 01, p. 1, Feb. 2019, doi: 10.1117/1.jrs.13.016518.

[37] G. Wang, Q. Zhou, and Y. Chen, "Robust non-rigid point set registration using spatially constrained Gaussian fields," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1759–1769, Apr. 2017, doi: 10.1109/TIP.2017.2658947.

[38] J. Ma, J. Zhao, J. Tian, X. Bai, and Z. Tu, "Regularized vector field learning with sparse approximation for mismatch removal," *Pattern Recognit.*, vol. 46, no. 12, pp. 3519–3532, 2013, doi: 10.1016/j.patcog.2013.05.017.

[39] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1706–1721, 2014, doi: 10.1109/TIP.2014.2307478.

[40] M. A. Fischler and R. C. Bolles, "Random sample consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981, doi: 10.1145/358669.358692.

[41] V. Fragoso, P. Sen, S. Rodriguez, and M. Turk, "EVSAC: Accelerating hypotheses generation by modeling matching scores with extreme value theory," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 2472–2479, doi: 10.1109/ICCV.2013.307.

[42] D. Barath and J. Matas, "Graph-cut RANSAC," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6733–6741.

[43] Y. Wang, X. Mei, Y. Ma, J. Huang, F. Fan, and J. Ma, "Learning to find reliable correspondences with local neighborhood consensus," *Neurocomputing*, vol. 406, pp. 150–158, 2020, doi: 10.1016/j.neucom.2020.04.016.

[44] S. Pang, J. Xue, Q. Tian, and N. Zheng, "Exploiting local linear geometric structure for identifying correct matches," *Comput. Vis. Image Underst.*, vol. 128, pp. 51–64, 2014, doi: 10.1016/j.cviu.2014.06.006.

[45] J. Ma, J. Zhao, J. Jiang, H. Zhou, and X. Guo, "Locality Preserving Matching," *Int. J. Comput. Vis.*, vol. 127, no. 5, pp. 512–531, 2019, doi: 10.1007/s11263-018-1117-z.

[46] Y. Li, Q. Huang, Y. Liu, Y. Huang, and X. Sun, "Efficient properties-based learning for mismatch removal," *IEEE Access*, vol. 7, pp. 149612–149622, 2019, doi: 10.1109/ACCESS.2019.2947178.

[47] J. Li, Q. Hu, M. Ai, and R. Zhong, "Robust feature matching via support-line voting and affine-invariant ratios," *ISPRS J. Photogramm. Remote Sens.*, vol. 132, pp. 61–76, Oct. 2017, doi: 10.1016/j.isprsjprs.2017.08.009.

[48] J. Li, Q. Hu, and M. Ai, "4FP-structure: A robust local region feature descriptor," *Photogramm. Eng. Remote Sensing*, vol. 83, no. 12, pp. 813–826, Dec. 2017, doi: 10.14358/PERS.83.12.813.

[49] J. Li, Q. Hu, and M. Ai, "LAM: Locality affine-invariant feature matching," *ISPRS J. Photogramm. Remote Sens.*, vol. 154, pp. 28–40, Aug. 2019, doi: 10.1016/j.isprsjprs.2019.05.006.

[50] G. Wang and Y. Chen, "Robust feature matching using guided local outlier factor," *Pattern Recognit.*, vol. 117, p. 107986, Sep. 2021, doi: 10.1016/j.patcog.2021.107986.

[51] X. Jiang, J. Ma, J. Jiang, and X. Guo, "Robust Feature Matching Using Spatial Clustering with Heavy Outliers," *IEEE Trans. Image Process.*, vol. 29, pp. 736–746, 2020, doi: 10.1109/TIP.2019.2934572.

[52] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," in *Proceedings of the IEEE International Conference on Computer Vision*, 2005, vol. II, pp. 1482–1489, doi: 10.1109/ICCV.2005.20.

[53] L. Torresani, V. Kolmogorov, and C. Rother, "Feature correspondence via graph matching: Models and global optimization," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2008, vol. 5303 LNCS, no. PART 2, pp. 596–609, doi: 10.1007/978-3-540-88688-4_44.

[54] H. Liu and S. Yan, "Common visual pattern discovery via spatially coherent correspondences," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 1609–1616, doi: 10.1109/CVPR.2010.5539780.

[55] K. Adamczewski, Y. Suh, and K. M. Lee, "Discrete tabu search for graph matching," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, vol. 2015 Inter, pp. 109–117, doi: 10.1109/ICCV.2015.21.

[56] T. Wang, H. Ling, C. Lang, and S. Feng, "Graph Matching with Adaptive and Branching Path Following," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 12, pp. 2853–2867, 2018, doi: 10.1109/TPAMI.2017.2767591.

[57] J. Ma, X. Jiang, J. Jiang, J. Zhao, and X. Guo, "LMR: Learning a Two-Class Classifier for Mismatch Removal," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 4045–4059, 2019, doi: 10.1109/TIP.2019.2906490.

[58] S. Pang, A. Du, M. A. Orgun, and H. Chen, "Weakly supervised learning for image keypoint matching using graph convolutional networks," *Knowledge-Based Syst.*, vol. 197, p. 105871, 2020, doi: 10.1016/j.knosys.2020.105871.

[59] E. N. Mortensen, H. Deng, and L. Shapiro, "A SIFT descriptor with global context," in *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, 2005, vol. I, pp. 184–190, doi: 10.1109/CVPR.2005.45.

[60] S. J. Mok, K. Jung, D. W. Ko, S. H. Lee, and B. U. Choi, "SERP: SURF enhancer for repeated pattern," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011, vol. 6939 LNCS, no. PART 2, pp. 578–587, doi: 10.1007/978-3-642-24031-7_58.

[61] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790–799, 1995.

[62] M. Chen, R. Qin, H. He, Q. Zhu, and X. Wang, "A local distinctive features matching method for remote sensing images with repetitive patterns," *Photogramm. Eng. Remote Sensing*, vol. 84, no. 8, pp. 513–523, 2018, doi: 10.14358/PERS.84.8.513.

[63] K. Fukunaga and L. D. Hostetler, "The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition," *IEEE Trans. Inf. Theory*, vol. 21, no. 1, pp. 32–40, 1975, doi: 10.1109/TIT.1975.1055330.

[64] P. Adriaans and P. E. Boas, "Computation, information, and the arrow of time," in *Computability in Context: Computation and Logic in the Real World*, 2011, vol. 1, pp. 1–17, doi: 10.1142/9781848162778_0001.

[65] V. Mousavi, M. Varshosaz, and F. Remondino, "Using information content to select keypoints for uav image matching," *Remote Sens.*, vol. 13, no. 7, 2021, doi: 10.3390/rs13071302.

[66] V. Mousavi, M. Varshosaz, and F. Remondino, "Evaluating Tie Points Distribution, Multiplicity and Number on the Accuracy of Uav Photogrammetry Blocks," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLIII-B2-2, pp. 39–46, 2021, doi: 10.5194/isprs-archives-xliii-b2-2021-39-2021.

[67] K. Mikolajczyk *et al.*, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, no. 1–2, pp. 43–72, Oct. 2005, doi: 10.1007/s11263-005-3848-x.

[68] A. Vedaldi and B. Fulkerson, "Vlfeat - An open and portable library of computer vision algorithms," *MM'10 - Proc. ACM Multimed. 2010 Int. Conf.*, pp. 1469–1472, 2010, doi: 10.1145/1873951.1874249.

[69] S. Badrloo, M. Varshosaz, S. Pirasteh, J. Li, "A novel region-based expansion rate obstacle detection method for MAVs using a fisheye camera," International Journal of Applied Earth Observation and Geoinformation, 2022, 108:102739. https://doi.org/10.1016/j.jag.2022.102739.

[70] M. Ghasemi, M. Varshosaz, S. Pirasteh, G. Shamsipour, "Optimizing Sector Ring Histogram of Oriented Gradients for human injured detection from drone images," Geomatics, Natural Hazards and Risk, 2021, 12:1, 581-604, DOI: 10.1080/19475705.2021.1884608.

[71] R. Yazdan, M. Varshosaz, S. Pirasteh, F. Remondino, "Performance of image classifiers for automatic segmentation of traffic signs," Geomatica, 2021, https://doi.org/10.1139/geomat-2020-0010.

DECLARATION OF INTEREST

The authors declare no conflict of interest. Data and codes are available upon request.

AUTHORS' BIOGRAPHY

Vahid Mousavi received B.Eng. degree in geomatics engineering from the University of Tafresh, Tafresh, in 2014 and a master's degree in photogrammetry from the K.N. Toosi University of Technology, Tehran, Iran, in 2016. He is currently working toward a Ph.D. degree in the Faculty of Geodesy and Geomatics Engineering, K.N. Toosi University of Technology. His research interests include image processing applied to close-range and UAV photogrammetry imagery, machine learning, deep neural networks and point cloud processing.

Dr. Masood Varshosaz is an associate professor in Photogrammetry at KN Toosi University of Technology, Iran (1999) from University College London, UK. He has published more than 100 papers in refereed journals and conference proceedings. Also, he has supervised more than 50 MSc and PhD students. His main interests are UAV and close range photogrammetry, computer vision, and laser scanning. He published the first book on UAV Photogrammetry principles which was soon recognised as one of the best three top-selling books in the field. He is a full-time member of the K. N. Toosi University of Technology, Iran. He has directed several national projects in the fields of Photogrammetry and GIS and has international ties with members from Canada and Europe.

Fabio Remondino got a PhD in Photogrammetry from ETH Zurich (2006) and is now the head of the 3D Optical Metrology research unit in FBK - Bruno Kessler Foundation (Italy). His main research interests are in the field of reality-based surveying and 3D modeling, sensor and data fusion and 3D data classification. He is working in all automation aspects of the entire 3D reconstruction pipeline for applications in the industrial, environmental and heritage fields. He is the author of more than 250 articles in journals and conferences. He organized more than 30 conferences, 20 summer schools and 5 tutorials. Fabio is now Vice-President of EuroSDR while he was serving as President of the ISPRS Technical Commission V and II (2012-2021) and vice-President of CIPA Heritage Documentation from 2015 to 2019.

Saeid Pirasteh received a Ph.D. degree in geology (remote sensing and GIS) from AMU, Aligarh, India, in 2004, and also a Ph.D. degree in geography (geomatics and LiDAR) from the University of Waterloo, Waterloo, ON, Canada, in 2018. He is currently an Associate Professor with the Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, China, scientist collaborator at the Geospatial Sensing and Data Intelligence Lab, University of Waterloo, Canada. He has co-authored over

200 publications. He invented and developed the Geospatial Infrastructure Management Ecosystem (GeoIME), which was one of the award winners for "Life Saving Solution of The Year" in 2022, Miami. His research interests include GeoAI, natural hazards & disasters and applications beyond, Remote Sensing (satellite, drone, LiDAR) data processing, GIS and Geospatial information analysis and management for specific purposes. Dr. Pirasteh is the UN-GGIM Academic Network Member Expert and the UN Open GIS WGs of GeoAI and Capacity Building. Since 2017, he has focused on the UN sustainable development goals (SDGs) 2030. His research interest migrated to integrating artificial intelligence, machine learning, deep learning, computer vision, development, and web app in geosciences and disaster applications.

Jonathan Li (Senior Member, IEEE) received the Ph.D. degree in geomatics engineering from the University of Cape Town, South Africa, in 2000. He is a Professor with the Department of Geography and Environmental Management and cross-appointed with the Department of Systems Design Engineering, University of Waterloo, Canada and a Fellow of the Engineering Institute of Canada. His main research interests include image and point cloud analytics, mobile mapping, and AI-powered information extraction from LiDAR point clouds and earth observation images. He has co-authored over 500 publications, including 300+ in refereed journals and 200+ in conference proceedings. Dr. Li is a recipient of the 2021 Geomatica Award, 2020 Samuel Gamble Award, and 2019 Outstanding Achievement Award in Mobile Mapping Technology. He is currently serving as the Editor-in-Chief of International Journal of Applied Earth Observation and Geoinformation, Associate Editor of IEEE Transactions on Intelligent Transportation Systems, IEEE Transactions on Geoscience and Remote Sensing, and Canadian Journal of Remote Sensing.