Development of a Framework for Stereo Image Retrieval With Both Height and Planar Features

Feifei Peng, Le Wang, Jianya Gong, and Huayi Wu

***Abstract*—The wide availability and increasing number of applications for high-resolution optical satellite stereo images (HrosSIs) have created a surging demand for the development of effective content-based image retrieval methods. However, this is a challenge for existing stereo image retrieval methods since they were designed for stereo images collected from close-range imaging sensors. Thus, successful retrieval of images is not assured given the mismatch between existing methods and the characteristics of HrosSIs. Moreover, none of the existing remote sensing image retrieval methods takes account of the specific characteristics of HrosSIs such as the viewing number and multiview angles. This paper proposes a generic framework to exploit the unique characteristics of HrosSIs data so as to allow efficient and accurate content-based HrosSI retrieval. HrosSIs retrieval is executed by similarity matching between the features obtained from digital surface models (DSMs) and orthoimages, both extracted from the HrosSIs. In addition, the significance of height information for HrosSI retrieval was investigated. A prototype system was designed and implemented for method validation using the ISPRS stereo benchmark test dataset. Experimental results show that the proposed techniques are efficient for HrosSI retrieval. The proposed framework is efficient and suitable for spaceborne stereo images but might also be suitable for airborne stereo images as well. Experimental results also show that height information alone is inefficient and unstable for HrosSI retrieval; however, a combination of height information and planar information is efficient and stable.**

***Index Terms*—Digital surface model (DSM), fractals, height features, image retrieval, orthoimage, planar features, stereo imagery.**

List of Abbreviations

HrosSI	high-resolution optical satellite stereo image.
DSM	digital surface model.
GSD	ground sample distance.
D	fractal dimension.
RFD	regional fractal dimension.
RFI	regional fractal image.
NRFCM	normalized regional fractal cooccurrence matrix.
NDRI	normalized dissimilarity ranking index.
CS1	Cartosat-1.
WV1	Worldview-1.

Manuscript received March 20, 2014; revised September 09, 2014; revised June 19, 2014; accepted October 09, 2014. This work was supported by the National Key Basic Research and Development Program of China under Grant 2012CB719906. The work of L. Wang was supported in part by the National Science Foundation (NSF) under Grant DEB-0810933 and Grant BCS-0822489 and in part by the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University.

F. Peng, J. Gong, and H. Wu are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan 430079, China (e-mail: feifpeng@whu.edu.cn; gongjy@whu.edu.cn; wuhuayi@whu.edu.cn).

L. Wang is with the Department of Geography, The State University of New York at Buffalo, Buffalo, NY 14261 USA (e-mail: lewang@buffalo.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2014.2363953

I. Introduction

MANY SATELLITES with the capability to produce high-resolution optical satellite stereo images (HrosSIs) are currently available, such as IKONOS, QuickBird, WorldView-1/2, Cartosat-1/2, GeoEye-1, Pleiades-HR, ALOS (PRISM), and ZY3 [1]. As a result, large quantities of HrosSIs are now accessible to researchers and the general public. HrosSIs are widely applied in various fields, such as digital surface model (DSM) production [2]–[4], building reconstruction [5], [6], change detection [7]–[9], and hazard assessment [10]. In some applications, desired HrosSIs must first be located based on semantic information or content-based similarities between HrosSIs. HrosSIs covering human settlements [11] in China, for instance, must first be geographically located to assess human settlement areas. It is very difficult to infer image content from image metadata (e.g., rational polynomial coefficients, geographic coverage, and acquisition time), thus the desired images cannot be found by using image metadata alone. Content-based image retrieval expedites search and discovery among large quantities of images based on the content of those images; and therefore can meet the discovery requirements when locating these desired HrosSIs.

The unique characteristics of HrosSIs data, however, make it difficult to extract uniform features from various HrosSIs for content-based HrosSI retrieval. HrosSIs are acquired by diverse remote sensors under diverse acquisition conditions and vary significantly in many ways, such as in stereo acquisition modes [12], viewing number, viewing angles, convergence angle, base-to-height ratio, ground sample distance (GSD), radiometric resolution, and metadata structure. These unique characteristics of HrosSIs data are not exploited in existing content-based image retrieval methods; thus, their potential in feature extraction and retrieval is unknown.

The significance of height information needs to be investigated to obtain efficient and accurate HrosSI retrieval. Height information can easily be derived from stereo-extracted DSMs, one type of primary product created from HrosSIs data. However, height information is not taken into consideration in existing image retrieval methods. Height information may be efficient to distinguish between land cover types that are difficult to distinguish with planar features alone. For instance, forest land on a plain and forest land in mountains both

have similar texture characteristics, but have radically different height characteristics. Thus, height information might be helpful for content-based HrosSI retrieval.

Successful HrosSIs retrieval is not assured given the mismatch between existing stereo image retrieval methods and the characteristics of HrosSIs. Feng *et al.* [13] initially proposed a generic framework for content-based stereo image retrieval which refined results from conventional content-based image retrieval by employing depth cues extracted from stereo pairs in a reranking scheme. These stereo images were obtained by close-range imaging sensors, whereas HrosSIs are acquired by spaceborne sensors. These two types of images vary in many ways, such as in stereo acquisition modes, viewing number, and viewing angles. Thus, the generic framework proposed by Feng *et al.* [13] has limited applicability for HrosSI retrieval.

None of existing remote sensing image retrieval methods is specifically designed for HrosSIs. A number of prototype systems have been developed in recent years [14], such as KIM/KEO [15], I^3KR [16], S^3IR [17], GeoIRIS [18], and so on [19]–[23]. These studies mainly target multispectral images, SAR images, and multitemporal images, which are obtained at a single look angle. However, HrosSIs are a group of images with multiview angles [2], such as stereo pairs, triplets, or multiangular imagery. Therefore, a basic difference exists between HrosSI retrieval and conventional remote sensing image retrieval.

This paper proposes a generic framework for content-based HrosSI retrieval, taking height information into consideration. Specifically, we focus on: 1) what features can be uniformly extracted from various HrosSIs, which are obtained by diverse remote sensors under various acquisition conditions; and 2) determine if height information is helpful for HrosSI retrieval.

The rest of this paper is organized as follows. Section II introduces the technical background. Section III introduces the proposed retrieval scheme in detail. Section IV shows the experimental results. A discussion is presented in Section V. Section VI provides a summary and outlines the prospects for future research.

II. Technical Background

Four interrelated technical domains form the background for HrosSI retrieval and are discussed in detail in the following sections. These topics relate to remote sensing image retrieval in general, the characteristics of HrosSIs and stereo products as well as a methodological concern, fractal analysis for remotely sensed images. General issues in remote sensing image retrieval must be considered since HrosSIs are remotely sensed images. The characteristics of HrosSIs and stereo products (orthoimages and DSMs) are analyzed to develop an effective HrosSI retrieval method. Fractal techniques for feature extraction from stereo products are suitable and widely used, and are applied in this study. Consequently, fractal analysis for remotely sensed images must be discussed in detail.

A. Remote Sensing Image Retrieval

Remote sensing image retrieval helps expedite the discovery of desired images. Users can conveniently find desired images from a large image database when employing an image retrieval system. As a result, they no longer need to manually browse the whole image database. Since users want to obtain the desired images quickly, a rapid, efficient, and accurate image retrieval system is desirable.

There are some common techniques found in image retrieval regardless of the type of images collected from different remote sensors [14], [24]. These common techniques include data preprocessing, feature extraction, similarity matching, results ranking, retrieval evaluation, and relevance feedback. Features obtained from images through feature extraction processing are a significant asset for image retrieval.

Features used in image retrieval can be divided into two categories according to their source: planar features and height features. Planar features (such as texture, shape, and spectral signatures) are extracted from remotely sensed images, while height features are extracted from DSMs or digital elevation models. Planar features are frequently used in remote sensing image retrieval [14], while height features are rarely used except for similarity assessment between digital elevation models [25].

B. Characteristics of HrosSIs

The characteristics of HrosSIs vary with remote sensors and acquisition conditions. The viewing direction angles, convergence angle, and base-to-height ratio are constant for the standard simultaneous along-track systems, such as Cartosat-1, ALOS (PRISM), and ZY3. But these are inconsistent with the standard across-track or the agile single-lens systems [12], such as WorldView-1/2, IKONOS, QuickBird, and GeoEye-1. Some sensors only deliver mono-band panchromatic images [26], such as Cartosat-1 and WorldView-1, while other sensors deliver panchromatic images and multispectral images, such as IKONOS and WorldView-2. Regarding the radiometric resolution, most HrosSIs are stored at 8 or 10 or 11 bits/pixel [12]. The nominal GSD varies for HrosSIs obtained with different remote sensors or the same remote sensor with different viewing direction angles. Furthermore, the ground observed azimuth and elevation of the sun for each image observation varies during its acquisition, resulting in different shadows in the HrosSIs.

C. Characteristics of Stereo Products

Two primary stereo products can be generated from various HrosSIs. Regardless of how HrosSIs are acquired, by what sensors, and under what acquisition conditions—DSMs and orthoimages, the two primary stereo products, can be generated from the HrosSIs. Nevertheless, stereo products generated from various HrosSIs still vary in three important ways.

The two primary stereo products vary in GSD, geometric accuracy, and radiometric resolution. The GSD of stereo products varies according to that of the HrosSIs from which they are generated. The geometric accuracy of stereo products is related to some characteristics of the HrosSIs [4], [27]–[33], such as sensor type, off-nadir viewing angle, and sensor model used. Their geometric accuracy is also related to other factors, such as number and distribution of ground control points, morphology of the study area, and the matching algorithms used in

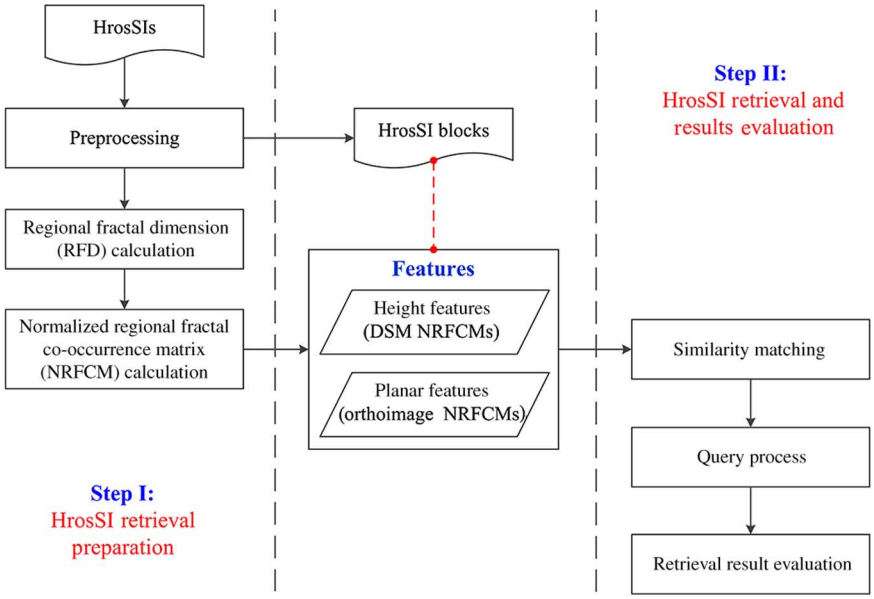


Fig. 1. Flowchart of HrosSI retrieval. HrosSI blocks for HrosSI retrieval are created from the raw complete HrosSIs through preprocessing. Features of the HrosSI blocks are obtained in Step I HrosSI retrieval preparation, and used in Step II HrosSI retrieval and results evaluation.

processing [4], [27]–[33]. The geometric accuracy of orthoimages is also influenced by the accuracy of DSMs employed in the orthorectification process [34]. Orthoimages vary according to radiometric resolution.

D. Fractal Analysis of Remotely Sensed Images

The fractal dimension (D) is widely used for characterizing the overall spatial complexity of remotely sensed images [35] or digital elevation models [36]. The fractal technique is suitable for orthoimages and DSMs and can be utilized for satellite image retrieval [23]. The D is a measurement of the roughness or textural complexity of land surface features [37]. The D values for an image range from 2.0 to 3.0. Irregular patterns always present higher D values but in any case, similar patterns present comparable D values. The fractal technique has limitations since it only yields a single D value for the entire image.

A local fractal dimension is an improved fractal dimension. The local D is widely used for image segmentation, image classification, and measurement of digital elevation models roughness [36], [38]. The local fractal technique yields many local D values with moving window techniques [39], [40]. If a kernel is moved over the image, a D value is calculated within the kernel, producing a local D value for the center pixel of the kernel. In this way, a local D is calculated for each pixel in the entire image except for the edges.

III. Methodology

A new generic framework is proposed for retrieval of various HrosSIs, obtained by diverse remote sensors under various acquisition conditions. HrosSI retrieval is executed by similarity matching between features extracted from DSMs and orthoimages, the two primary stereo products generated from HrosSIs. The proposed retrieval scheme involves two steps: Step I) HrosSI retrieval preparation and Step II) HrosSI retrieval and results evaluation, as shown in Fig. 1. Step I consists of preprocessing, regional fractal dimension (RFD) calculation, and normalized regional fractal cooccurrence matrix (NRFCM) calculation. Step II consists of similarity matching, a query process, and results evaluation. HrosSI blocks and their features are created in Step I, and brought into play in Step II.

A. Preprocessing

It takes several steps to distinguish between different HrosSIs with varied characteristics. Since similarity matching between HrosSIs is executed by similarity matching between stereo-extracted DSMs and orthoimages, the first step is to generate these two primary stereo products for various HrosSIs. Second, the rich and complex visual content of images at a large size make it difficult to make similar judgments about the image; therefore, these huge HrosSIs and stereo products must be decomposed into several blocks suitable for retrieval. A query template is advantageous when matching these smaller blocks [20].

Stereo products are generated to do similarity matching between HrosSIs. Since characteristics of stereo products may affect similarity matching between HrosSIs, some measures must be taken to minimize the impact of these characteristics. DSMs at the same GSD are generated from various HrosSIs using an identical matching algorithm. DSMs generated from various HrosSIs still vary in GSD and image details; therefore, DSMs are subsequently subjected to resampling and low-pass filtering. Moreover, with the aid of stereo-extracted DSMs, orthoimages are generated during orthorectification of the near-nadir image among the group of HrosSIs. Orthoimages generated from various HrosSIs still vary in the GSD, radiometric resolution, image intensity, and image details; therefore, orthoimages must be subjected to resampling,

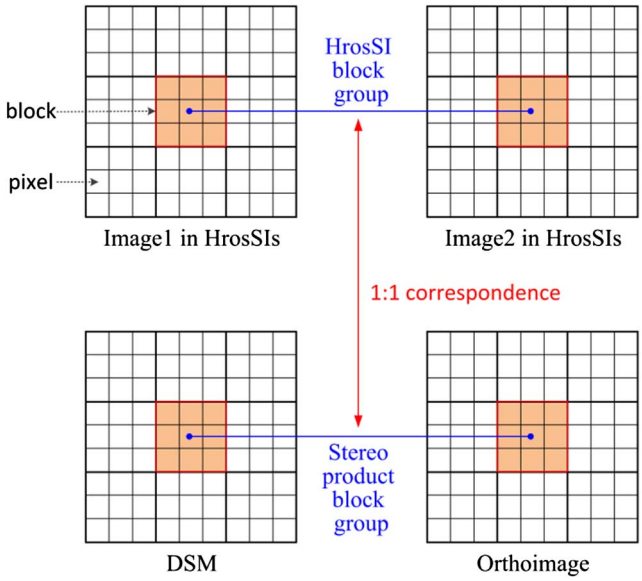


Fig. 2. Decomposition of HrosSIs and stereo products. Using the same setting, each image in the HrosSIs, DSM, and orthoimage is decomposed into equal sized nonoverlapping blocks, respectively. Although HrosSIs may have two, three, or more images, just two images in HrosSIs are shown for clarity in this figure. Through decomposition, a one-to-one correspondence between a HrosSI block group and its corresponding stereo product block group is built.

radiometric resolution normalization, histogram equalization, and low-pass filtering.

HrosSIs and stereo products at a huge size are decomposed into several blocks suitable for retrieval, as shown in Fig. 2. HrosSIs are decomposed into equal sized nonoverlapping HrosSI block groups by using the block-oriented image decomposition method [41]. These HrosSI block groups, which can be seen as HrosSIs at a smaller size, are used for HrosSI retrieval. The same decomposition setting for HrosSIs is used for their stereo products, thus producing several stereo product block groups. In this way, a one-to-one correspondence between a HrosSI block group and its corresponding stereo product block group is built, so that the features extracted from a stereo product block group can be seen as features of the corresponding HrosSI block group.

B. RFD Calculation

Fractal techniques cannot be directly used for HrosSI retrieval. Two images with the same single fractal dimension (D) value may vary significantly, due to the fact that ranges of D values derived from different land cover types may overlap each other [42]. Furthermore, since different numbers of local D values are created for different sizes of HrosSIs, local D values are not suitable for similarity matching between them. Thus, the RFD is proposed for HrosSI retrieval with an improved D and local D.

RFD values are calculated in stereo product blocks, as shown in Fig. 3. The process involves two substeps: regional window decomposition and fractal dimension calculation. In the regional window decomposition substep, the image—the DSM block or orthoimage block obtained through previous preprocessing—is divided into several nonoverlapping regional windows at a low level. In the fractal dimension calculation substep, a D value for each regional window (the RFD value) is calculated. In this way, a regional fractal image (RFI) is generated from the raw image; the RFI is a point-by-point image of the estimated RFD values of all regional windows. For instance, a DSM block RFI with $M/rw \times M/rw$ pixels size is generated from a DSM block with $M \times M$ pixels size by using the RFD calculation method with a regional window $rw \times rw$ pixels size. Stereo product block RFIs (the DSM block RFI and the orthoimage block RFI) are generated from stereo product blocks during RFD calculation.

In the regional window decomposition substep, nonoverlapping regional windows at a low level are obtained through the decomposition of the DSM or orthoimage block using the block-oriented image decomposition method [41]. During decomposition, the regional window size needs to be determined based on certain conditions. Minimization of regional window size is required to capture local variations effectively. However, a larger regional window size is required to reduce the impact of geometric accuracy differences in stereo products, generated from different remote sensors, on HrosSI retrieval. Moreover, the image size must be divisible by the regional window size. Several regional window sizes may be available for the same DSM or orthoimage block, producing differently sized RFIs. The same size of RFIs can be generated from different sizes of DSM or orthoimage blocks, by using corresponding regional window sizes.

In the fractal dimension calculation substep, the RFD value for each regional window is calculated by the blanket method [43], [44]. Each regional window is seen as an image at a small size in the fractal dimension calculation. An image surface $f(i,j)$ is covered by a blanket of thickness 2ε. The covering blanket is defined by its upper surface u_ε and its lower surface b_ε [37]. Initially, $u_0(i,j) = b_0(i,j) = f(i,j)$ and the blanket surfaces are defined as follows:

$$\begin{aligned} u_\varepsilon(i,j) &= \max\left\{ u_{\varepsilon-1}(i,j)+1, \max_{|(m,n)-(i,j)\leqslant 1|} u_{\varepsilon-1}(m,n) \right\} \\ b_\varepsilon(i,j) &= \min\left\{ b_{\varepsilon-1}(i,j)-1, \min_{|(m,n)-(i,j)\leqslant 1|} b_{\varepsilon-1}(m,n) \right\}. \end{aligned} \tag{1}$$

The image points (m,n) with distance less than one from the point (i,j) are taken to be the four immediate neighbors of the point (i,j). Then, the area of blanket $A(\varepsilon)$ is calculated by

$$A(\varepsilon) = \frac{\sum_{i,j}\left(u_\varepsilon(i,j) - b_\varepsilon(i,j)\right)}{2\varepsilon}. \tag{2}$$

The behavior of a fractal surface is defined [45] as follows:

$$A(\varepsilon) = F\varepsilon^{2-D} \tag{3}$$

where F is a constant and D is the fractal dimension of the surface. When plotting $A(\varepsilon)$ versus ε on a log–log scale, a straight line of slope $2-D$ is achieved.

Stereo product block RFIs are generated for stereo product blocks through the RFD calculation. However, it is still difficult to directly distinguish between RFIs. On the other hand, their extracted features can distinguish them.

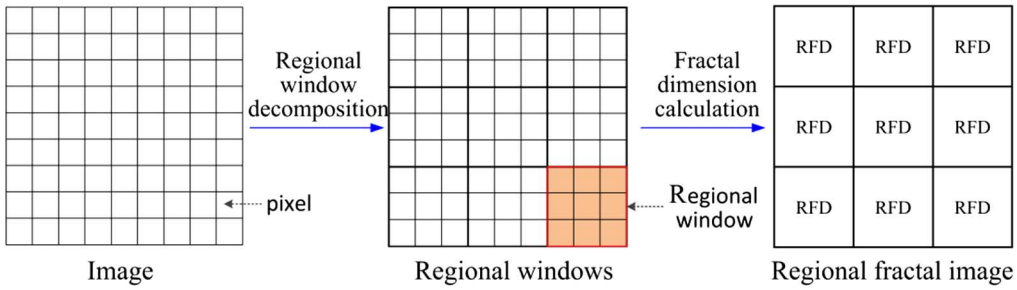


Fig. 3. RFD calculation. The process involves two substeps: regional window decomposition and fractal dimension calculation. A RFD value is calculated for each regional window obtained in the regional window decomposition substep.

C. NRFCM Calculation

The gray-level cooccurrence matrix measures the distribution of cooccurring gray values with the displacement vector in an image. Since the gray-level cooccurrence matrix is a good discriminator between images, the NRFCM modeled after the gray-level cooccurrence matrix is proposed as a feature to distinguish between RFIs. The NRFCM measures the occurrence frequency of the combination of two neighboring pixels (the RFD values) in the RFI. The RFD values in the RFI are decimals, and thus are inconvenient when calculating occurrence frequency.

RFD values must first be normalized to produce a normalized RFI prior to occurrence frequency calculation. RFD values in the RFI are decimals that range from 2.0 to 3.0. They are converted into integers ranging from 1 to N using a linear extension method thus obtaining normalized RFD values and the normalized RFI. The size of the normalized RFI is the same as the RFI.

The NRFCM is a matrix that measures the distribution of cooccurring normalized RFD values with a displacement vector in a normalized RFI. Mathematically, an NRFCM is defined using a normalized RFI I at $m \times n$ pixels size, and parameterized by the displacement vector $(\Delta x, \Delta y)$ as

$$NRFCM_{(\Delta x,\,\Delta y)}(i,j) = \sum_{p=1}^{m}\sum_{q=1}^{n}\begin{cases}1, & \text{if } I(p,q)=i \text{ and } I(p+\Delta x, q+\Delta y)=j\\ 0, & \text{otherwise}\end{cases} \tag{4}$$

where i and j are the row number and column number in the NRFCM, respectively, i and j are also the normalized RFD values in the normalized RFI I, p, and q are the spatial positions in I. The size of the NRFCM is $N \times N$, where N is the upper limit of normalized RFD values.

The displacement vector measures the relative relationship of the two neighboring pixels in the normalized RFI. Different NRFCMs are generated from the same normalized RFI by using different displacements vectors. There are four different displacement vectors in this study, as shown in Fig. 4. If a displacement vector is rotated by 180°, the two NRFCMs generated with the raw and new displacement vector are transposed each other, and are equivalent in similarity matching. These four displacement vectors represent all different types of the combination of two neighboring pixels in the normalized RFI.

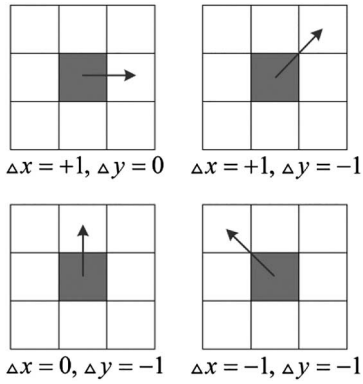


Fig. 4. Displacement vectors in the normalized RFI for calculating an NRFCM. The displacement vector $(\Delta x, \Delta y)$ is displayed with an arrow. The reference is the gray pixel in the center. The displacement vector indicates the relative relationship between the reference pixel and the specific neighboring pixel.

In this way, a group of NRFCMs, composed of four NRFCMs created with the four displacement vectors, is created from the normalized RFI.

Features are extracted from stereo product block RFIs through NRFCM calculation. A group of DSM NRFCMs is created from the DSM block normalized RFI. A group of orthoimage NRFCMs is created from the orthoimage block normalized RFI. According to the source of the NRFCMs, the DSM NRFCMs belong to height features, while the orthoimage NRFCMs belong to planar features. In this way, both height features and planar features are obtained from stereo product blocks.

D. Similarity Matching

Similarity matching between HrosSI blocks is executed by similarity matching between features extracted from their corresponding stereo product blocks. Since the features are composed of NRFCMs, similarity matching between NRFCMs must be determined first. Similarity matching between HrosSI blocks is done with height features, planar features, or a combination of both.

Similarity matching between NRFCMs is also done based on certain conditions. Two NRFCMs are comparable only if they belong to the same category (DSM NRFCMs or orthoimage NRFCMs), and created from the same size normalized RFIs with the same displacement vector. The generated normalized RFIs for HrosSI blocks must be at the same size in order to do

similarity matching between them. The similarity between two NRFCMs, $M1$ and $M2$ is calculated by (5).

Similarity matching between HrosSI blocks can be done with height features or planar features. Considering that the height features or the planar features are composed of NRFCMs, similarity matching between HrosSI blocks is done with a combination of similarities between NRFCMs as (6). The *vn* is the total number of NRFCMs with different displacement vectors. The *vn* is four for orthoimage NRFCMs or DSM NRFCMs.

Similarity matching between HrosSI blocks can also be done by the integration of height features and planar features. The similarity is measured with a combination of orthoimage NRFCMs or DSM NRFCMs as in (7). The parameter d indicates the weight values of the height features and the planar features.

$$\mathrm{SmMatrix} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\min(M1_{i,j}, M2_{i,j})}{\sum_{i=1}^{N}\sum_{j=1}^{N} M1_{i,j}} \tag{5}$$

$$\mathrm{SmUnion} = \sum_{v=1}^{vn}\frac{\mathrm{SmMatrix_v}}{vn} \tag{6}$$

$$\mathrm{SmIntegrated} = d \times \mathrm{SmUnion_{DSM}} + (1-d) \times \mathrm{SmUnion_{Ortho}}, \quad 0 \le d \le 1. \tag{7}$$

E. Query Process

A retrieval system deals with user query requests with a specific retrieval method. Three HrosSI retrieval methods reflect the three ways of similarity matching between HrosSI blocks. Classified according to the category of features used in similarity matching, the three retrieval methods are identified as the DSM-NRFCMs-based retrieval method, the ortho-NRFCMs-based retrieval method, and the integrated-NRFCMs-based retrieval method, respectively. In one query process, a retrieval system deploys only one specific retrieval method.

A query process refers to the interaction between users and the retrieval system, as shown in Fig. 5. At first, users submit a query by selecting an interesting HrosSI block group, and a specific retrieval method to be deployed in the retrieval system. Then, the similarity between the query HrosSI block group and each HrosSI block group in the dataset is calculated in turn. At last, a ranking list sorted in a descending order is returned to the users.

F. Retrieval Result Evaluation

1) Retrieval Accuracy: The accuracy of the retrieval method is calculated based on the ranking list that responds to a query with the retrieval system. The recall, precision, and F-measure [21] are calculated as follows:

$$\begin{aligned}\mathrm{Recall} &= \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}\\ \mathrm{Precision} &= \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}\\ \mathrm{F-measure} &= \frac{2\times\mathrm{Recall}\times\mathrm{Precision}}{\mathrm{Recall}+\mathrm{Precision}}\end{aligned} \tag{8}$$

where TP is the number of true positives, FN is the number of false negatives, FP is the number of false positives, F-measure is a harmonic mean of recall and precision. In other words, TP is the number of returned HrosSI block groups that are relevant to the query HrosSI block group. FN is the number of relevant HrosSI block groups that are not returned. FP is the number of returned HrosSI block groups that are not relevant.

The retrieval accuracy can be calculated for the entire dataset as well as a subdataset of the entire dataset. With the same entire ranking list, the accuracy for the subdataset can be measured according to the number of corresponding HrosSI block groups in (8). For instance, the entire dataset is divided into the Worldview-1 and the Cartosat-1 subdatasets according to the sensor type of HrosSI blocks. If a Worldview-1 HrosSI block group is queried with the prototype system, it returns a ranking list for the entire dataset. With the same entire ranking list, retrieval accuracy can be calculated for the entire dataset, the Worldview-1 and the Cartosat-1 subdatasets, respectively. In this way, three retrieval accuracy modes are calculated.

2) Retrieval Efficiency: The retrieval efficiency is measured with the time from the query to the retrieval results return to users. The retrieval efficiency may be affected by some factors, such as the total number of HrosSI block groups in the entire dataset, the query HrosSI block group, and the experimental environments. Thus, these factors must be the same in order to objectively compare the retrieval efficiency of different retrieval methods.

3) Normalized Dissimilarity Ranking Index: The sequence order in the ranking list is crucial. Some indices, such as the tau index [46], are used for measuring the sequence order with the help of a desirable ranking list. However, because of the rich and complex visual contents of remotely sensed images and user experiences, it is very difficult for users to generate a unified ranking list. Nevertheless, two HrosSI block groups covering the same scene from different sensors should have a comparable ranking, regardless of their contents and a users' subjective judgment. The Normalized Dissimilarity Ranking Index (NDRI) is proposed to objectively measure the ranking difference between these two HrosSI block groups with no need for a desirable ranking list.

The NDRI for two HrosSI block groups covering the same scene from different sensors is calculated based on the ranking list. The NDRI is measured as follows:

$$\mathrm{NDRI} = \frac{|\mathrm{R1}-\mathrm{R2}|-1}{\mathrm{TR}}, \quad 0 \le \mathrm{NDRI} \le 1 \tag{9}$$

where R1 is the ranking of a HrosSI block group in the ranking list, R2 is the ranking of the other HrosSI block group in the ranking list, and TR is the total number of HrosSI block groups in the ranking list. The NDRI value ranges from 0 to 1, and indicates the stability of the retrieval method for the two HrosSI block groups. The NDRI value close to 0 indicates that the retrieval method is stable for the two groups, while the NDRI value close to 1 indicates that the retrieval method is unstable for the two groups. Further, with the same ranking list, an NDRI value can be calculated for each scene that is covered by two sensors in the entire dataset. The total number of NDRI values for the entire dataset is derived from the total number of scenes in the dataset.

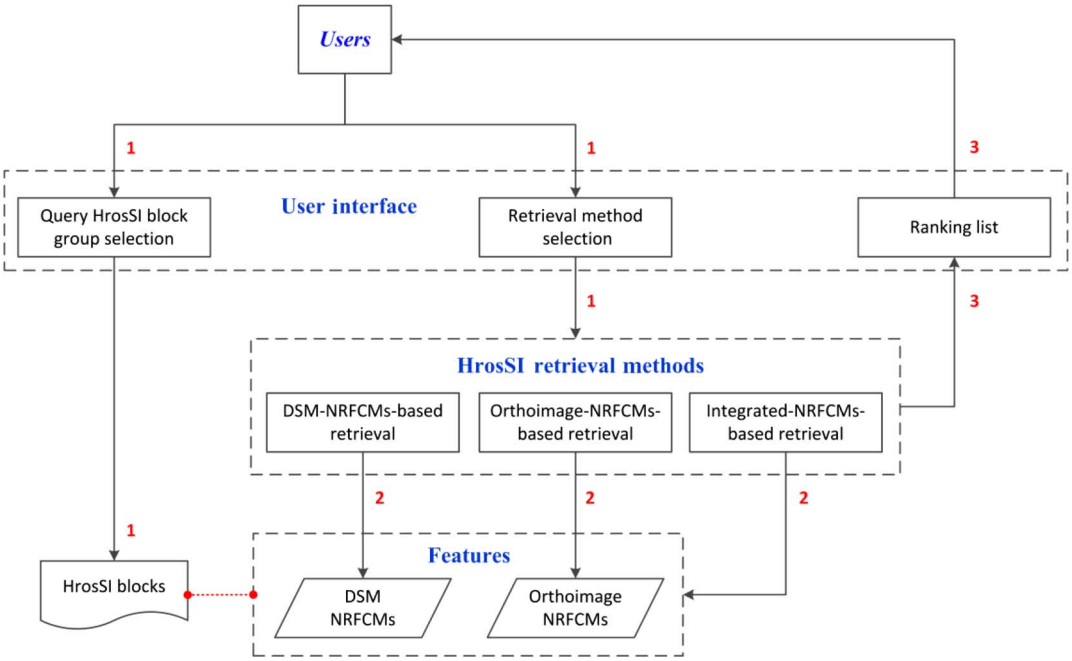


Fig. 5. Query process. Through the user interface, users interact with the retrieval system, deploying a retrieval method based on the corresponding features of HrosSI blocks. The number of the arrows indicates the order of each specific task in the entire query process. The number "1" indicates how users initialize the query process. The number "2" indicates similarity matching between HrosSI blocks with their features. The number "3" indicates results ranking and return to users.

TABLE I
PROPERTIES OF THE TEST AREAS AND THEIR STEREO PAIRS

Area name	Land cover types in the area	Area size	Stereo pairs	
			Cartosat-1 stereo pair	Worldview-1 stereo pair
Vacarisses	wooded hills, quarry, waste dump	4 × 4 km^2	• Acquisition time: March 2008 • Bands: mono-band panchromatic • Radiometric resolution: 10 bits • GSD: 2.5 m • Looking angles: +26° /–5° • Convergence angle: 31°	• Acquisition time: August 2008 • Bands: mono-band panchromatic • Radiometric resolution: 11 bits • GSD: 0.5 / 0.66 m • Looking angles: +4° /–33° • Convergence angle: 34°
Lamola	steep mountainous terrain, forests	4 × 4 km^2		
Terrassa	urban, industrial, residential	4 × 4 km^2		

The distribution of NDRI values for the entire dataset indicates the stability of the retrieval method. In order to objectively analyze the distribution without the impact of the selected query scene, all scenes covered by one remote sensor are queried with the retrieval system in turn, producing large quantities of NDRI values. Then, the distribution of all these NDRIs can be visualized graphically as a boxplot. The stability of the retrieval method decreases with the increase in the median and the quartiles of the boxplot.

IV. EXPERIMENTAL RESULTS

A. Study Data

The ISPRS stereo benchmark test dataset [47], [48] was used in this paper. The test dataset involved three test areas with distinct properties in Catalonia, Spain [49]. The three test areas were Vacarisses, Lamola, and Terrassa, respectively. Each test area was covered by two stereo pairs with rational polynomial coefficients: one Cartosat-1 (CS1) stereo pair and one Worldview-1 (WV1) stereo pair. Tables I shows the properties of the test areas and stereo pairs that covered the test areas. The ground control points, and the azimuth and elevation of the sun during acquisition of stereo pairs were not provided.

B. Experimental Setting

The prototype system was implemented using MATLAB R2012b. The testing platform had two Intel(R) Core (TM) i3-2310M 2.10 GHz CPUs with 2-GB memory.

DSMs and orthoimages were generated from the six stereo pairs in the ISPRS stereo benchmark test dataset. Stereo products were generated with ENVI 4.8 software [50]. At first, stereo products at 2.5-m GSD were produced from CS1 stereo pairs, while stereo products at 0.5-m GSD were produced from WV1 stereo pairs. Then, stereo products generated from WV1 stereo pairs were resampled to 1, 2, and 2.5 m. All orthoimages generated from WV1 stereo pairs at 11 bits were normalized to 10 bits. All orthoimages generated from all stereo pairs were

TABLE II
TYPES OF STEREO PRODUCTS AND DECOMPOSITION SETTINGS

Stereo pair type	Radiometric resolution	GSD	Complete size (pixels)	Block size (pixels)
Cartosat-1	10 bits	2.5 m	1536×1536	256×256
	1 type (1 type of orthoimage and 1 type of DSM)		1 scene	36 scenes
Worldview-1	10 bits, 11 bits	0.5 m	7680×7680	1280×1280
		1 m	3840×3840	640×640
		2 m	1920×1920	320×320
		2.5 m	1536×1536	256×256
	8 types (8 types of orthoimages and 4 types of DSMs)		1 scene	36 scenes

TABLE III
REGIONAL WINDOW SIZE IN RFD CALCULATION: REGARDLESS OF STEREO PRODUCT BLOCK SIZE, FOUR SIZES OF RFIS ARE CREATED FROM EACH STEREO PRODUCT BLOCK: 64×64, 32×32, 16×16 AND 8×8 PIXELS

Block size (pixels)	Regional window size (pixels)			
	64×64	32×32	16×16	8×8
1280×1280	20×20	40×40	80×80	160×160
640×640	10×10	20×20	40×40	80×80
320×320	5×5	10×10	20×20	40×40
256×256	4×4	8×8	16×16	32×32

subjected to histogram equalization. All stereo products generated from all stereo pairs were subjected to low-pass filtering with the 3×3 pixels kernel size. In this way, different types of stereo products were generated from HrosSIs. The type of stereo products can be labeled with the number of bits per pixel and GSD of stereo products. Tables II shows the types of stereo products. Different types of stereo products were created for different types of stereo pairs. One type of stereo product was created for CS1 stereo pairs, while eight types were created for WV1 stereo pairs. Since the types of DSMs were only related to the GSD, there were only four different types of DSMs for WV1 stereo pairs.

Stereo pairs and their stereo products were decomposed into several blocks, as shown in Tables II. For each test area in the ISPRS stereo benchmark test dataset, the complete scene (stereo pair or stereo products) was divided into 36 scenes. The GSD of stereo pairs from different remote sensors varied; thus, the complete stereo pairs or the decomposed stereo pair blocks that covered the same scene by different remote sensors varied in the size. This also happened to stereo products and stereo product blocks. Moreover, the decomposition setting for CS1 stereo products was the same as that for CS1 stereo pairs. The decomposition setting for eight types of WV1 stereo products was different from that for WV1 stereo pairs, and varied with the GSD. Eight types of WV1 stereo products were split into different decomposition sizes according to their GSD, respectively, as shown in Tables II. Considering that stereo products have different types, stereo product blocks also have types. The types of stereo product blocks were determined by types of stereo products.

Stereo product block RFIs were created from stereo product blocks through RFD calculation. RFD values were estimated with the blanket's scale range from 1 to 100. Determining the regional window size for all stereo product blocks was significant for the RFD value calculation. The regional window size for each stereo product block itself was determined under the conditions as previously described. The RFIs generated from various sizes of stereo product blocks must be at the same size in order to do similarity matching between HrosSI blocks. For these reasons, four sizes of RFIs were chosen to be generated from all stereo product blocks, as shown in Tables III. In this way, various types of RFIs were created for different sensor types of stereo pair blocks. The number of types of RFIs was equal to four times of the number of types of stereo product blocks.

Fig. 6 shows stereo products and their corresponding RFIs of the three complete test areas in the ISPRS stereo benchmark test dataset. The orthoimages and the DSMs, which were at 10 bits and at 2.5-m GSD, were generated from CS1 stereo pairs covering the three complete test areas. The orthoimage RFIs and the DSM RFIs were created with the 8×8 pixels regional window size in the RFD calculation.

Features of stereo pair blocks were created from stereo product block normalized RFIs. Normalized RFIs ranging from 1 to 100 were produced by normalizing RFIs with the linear extension method. Then, DSM NRFCMs and orthoimage NRFCMs were created from DSM and orthoimage block normalized RFIs, respectively. Since various types of RFIs were created for the stereo pair blocks of different sensor types, different groups of height features or planar features were created for them. The number of groups of height features or planar features was determined by the number of types of RFIs.

The dataset of stereo pair blocks and their features were prepared for HrosSI retrieval, as shown in Tables IV. Since 36 scenes were created from one complete test area through decomposition, 36×3 scenes were created from the three complete test areas. Each scene were covered by a CS1 stereo pair block group as well as a WV1 stereo pair block group, therefore the entire dataset for HrosSI retrieval composed of 216 stereo pair block groups. The entire dataset can be divided into the WV1 and the CS1 subdatasets according to the sensor type of stereo pair blocks. Both the WV1 subdataset and the CS1 subdataset composed of 108 stereo pair block groups of the corresponding sensor type. Furthermore, since different types of stereo product blocks and RFIs were produced for different sensor types of stereo pair blocks, various groups of features were created for them.

The integrated-NRFCMs-based retrieval method is based on a combination of planar features and height features. The weight values for planar features and height features were determined by the parameter d in (7). In this study, three different weight values were assigned to the parameter: 0.25, 0.5, and 0.75. Moreover, planar features and height features combined in the retrieval method were created from the same size of normalized RFIs.

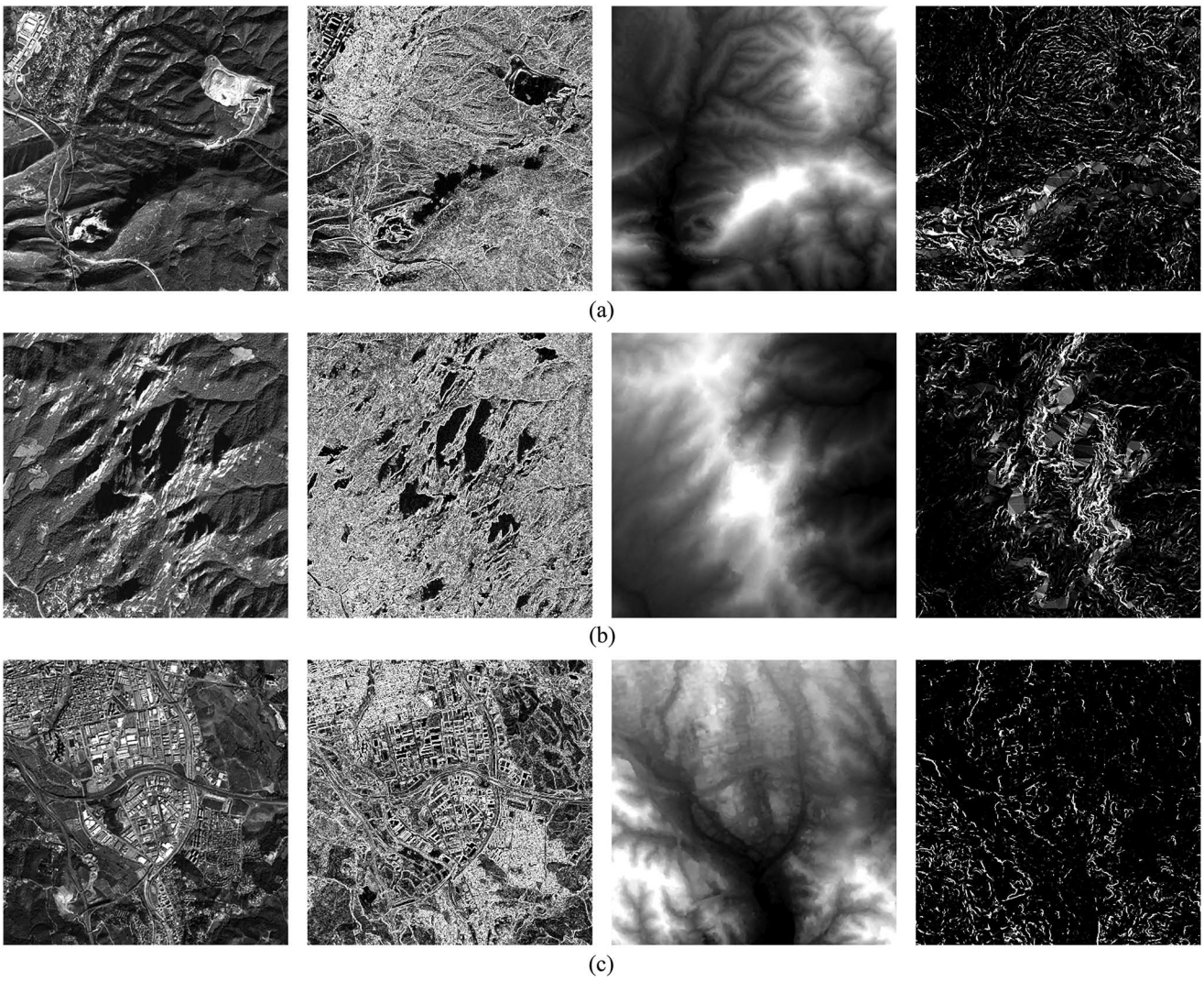


Fig. 6. Stereo products and corresponding RFIs for the three test areas. (a) Vacarisses. (b) Lamola. (c) Terrassa. The orthoimage, orthoimage RFI, DSM, and DSM RFI are arranged from left to right for each test area. All the images were produced from CS1 stereo pairs for the three complete test areas in the ISPRS stereo benchmark test dataset. All the images are stretched for visual clarity. Light pixels indicate high RFD values in orthoimage RFIs and DSM RFIs, while light pixels indicate high elevation values in DSMs.

TABLE IV
NUMBER OF STEREO PAIR BLOCKS AND FEATURE TYPES

Stereo pair block type		Cartosat-1	Worldview-1
Scene number		36 × 3	36 × 3
Stereo product block type	DSM	1	4
	Orthoimage	1	8
RFI size type	DSM	1 × 4	4 × 4
	Orthoimage	1 × 4	8 × 4
Feature type	Height	4	16
	Planar	4	32

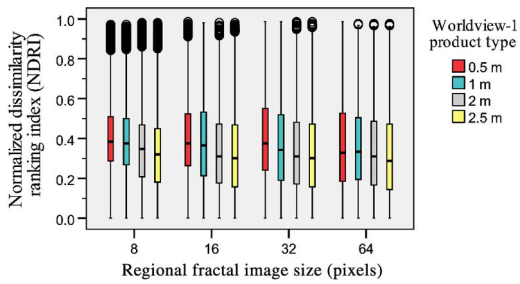


Fig. 7. Boxplots for NDRI values using the DSM-NRFCMs-based retrieval method. A boxplot represents the distribution of NDRI values for the entire dataset, using a retrieval method and a group of features. Specifically, features of HrosSI blocks used in the retrieval method are created from the DSM block RFIs in four sizes. CS1 DSM block RFIs are generated from DSMs at 2.5-m GSD, while WV1 DSM block RFIs are generated from DSMs at four different GSD.

C. NDRI Analysis

In this experiment, the influence of key parameters in the feature extraction processing on HrosSI retrieval was investigated with NDRI analysis; the GSD of stereo products, the radiometric resolution of orthoimages, and the RFI size.

Boxplots for NDRI values using the three retrieval methods were created as follows. The prototype system deployed a retrieval method. Then, 108 CS1 stereo pair block groups in

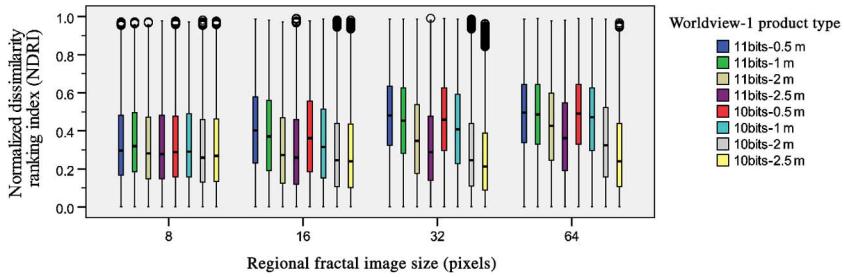


Fig. 8. Boxplots for NDRI values using the ortho-NRFCMs-based retrieval method. The features of CS1 HrosSI blocks are created from DSMs and orthoimages at 2.5-m GSD and at 10 bits. The features of WV1 HrosSI blocks are created from DSMs and orthoimages with eight types of GSD and radiometric resolutions. Other settings are the same as those in Fig. 7.

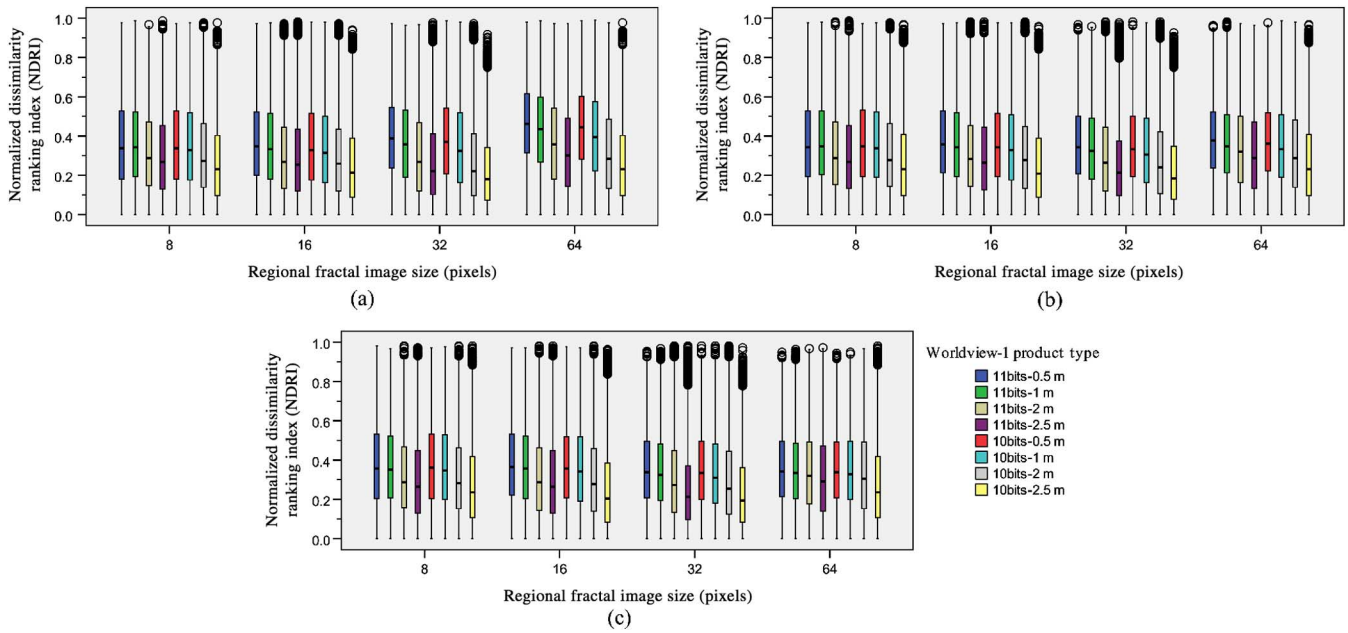


Fig. 9. Boxplots for NDRI values using integrated-NRFCMs-based retrieval method. (a) $d = 0.25$. (b) $d = 0.5$. (c) $d = 0.75$. The other settings are the same as those in Fig. 7.

the entire dataset were queried with the prototype system in turn. Since 108 NDRI values can be created with one query, 108×108 NDRI values were obtained and displayed as a boxplot for the retrieval method. The number of boxplots for a retrieval method was determined by the number of groups of features used in this retrieval method. Therefore, there are 16 boxplots by using the DSM-NRFCMs-based retrieval method, as shown in Fig. 7. There are 32 boxplots by using the other two retrieval methods, as shown in Figs. 8 and 9.

The influence of the GSD and radiometric resolution of stereo products on HrosSI retrieval can be investigated with Figs. 7–9. Considering that the GSD of CS1 stereo products was constant at 2.5-m GSD, Figs. 7–9 show that the stability increases with the GSD of WV1 stereo products toward 2.5 m for the three retrieval methods at the four RFI sizes. Considering that the radiometric resolution of CS1 orthoimages was constant at 10 bits, Figs. 8 and 9 show that 10 bits is more stable than 11 bits for the three retrieval methods at four RFI sizes. Therefore, stereo products generated from various HrosSIs must be at the same GSD and radiometric resolution in order to obtain the most stable HrosSI retrieval.

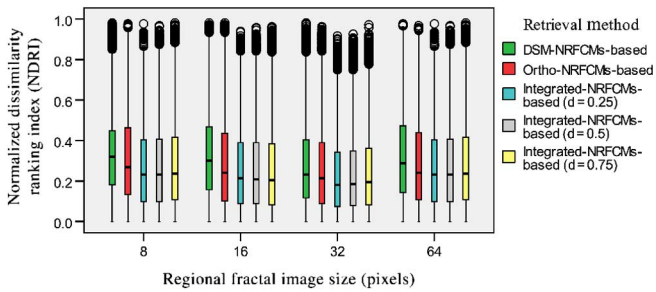


Fig. 10. Boxplots for NDRI values using the three retrieval methods. The features of both CS1 HrosSI blocks and WV1 HrosSI blocks are created from DSMs and orthoimages at 2.5-m GSD and at 10 bits. Other settings are the same as those in Fig. 7.

Fig. 10 shows boxplots for NDRI values using the three retrieval methods. Since features extracted from stereo product blocks at 2.5 m and at 10 bits were the most stable for HrosSI retrieval, they were used in the analysis. The stability of height information and the RFI size can be derived from boxplots of the three retrieval methods with Fig. 10.

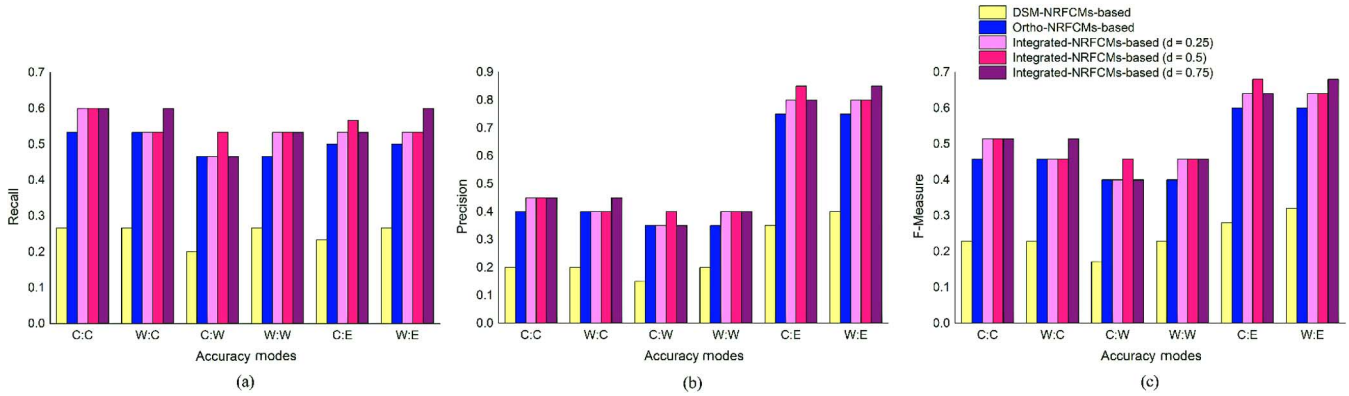


Fig. 11. Retrieval accuracy of the three retrieval methods. (a) Recall. (b) Precision. (c) F-measure. By using one retrieval method, a ranking list responds to one sensor type query. The retrieval accuracy for the sensor type query is calculated on the entire dataset, the CS1 and the WV1 subdatasets, respectively. In the accuracy mode labels, the part to the left before the colon represents the sensor type query: "C" indicating the CS1 query HrosSI blocks, and "W" indicating the WV1 query HrosSI blocks. The part to the right after the colon represents the evaluated dataset: "C" indicating the CS1 subdataset, "W" indicating the WV1 subdataset, and "E" indicating the entire dataset.

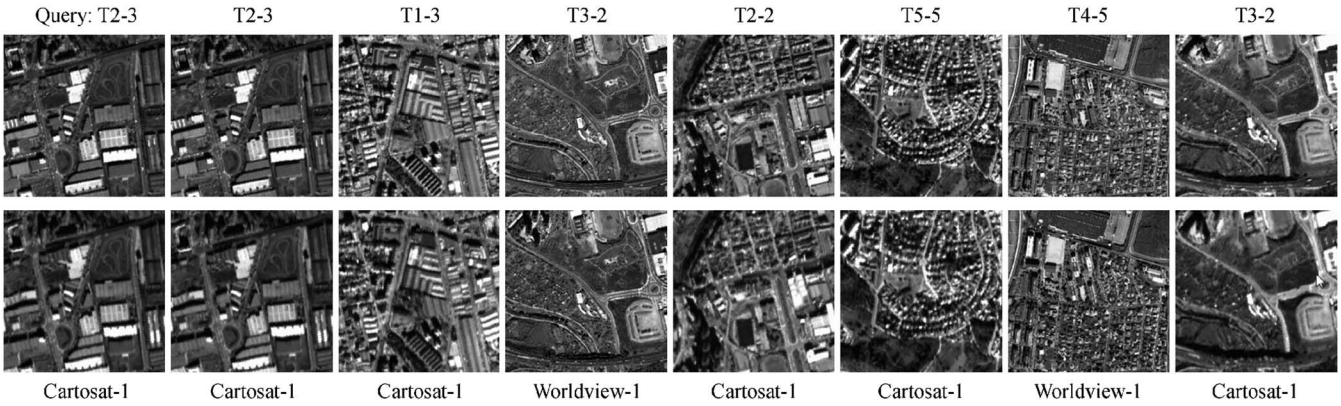


Fig. 12. Results by the DSM-NRFCMs-based retrieval method. The first column represents the query stereo pair block group. The second to eighth columns show the seven most similar stereo pair block groups with descending similarity to the query stereo pair block group, arranged from left to right. Each stereo pair block group is labeled with the scene name above it and sensor type under it.

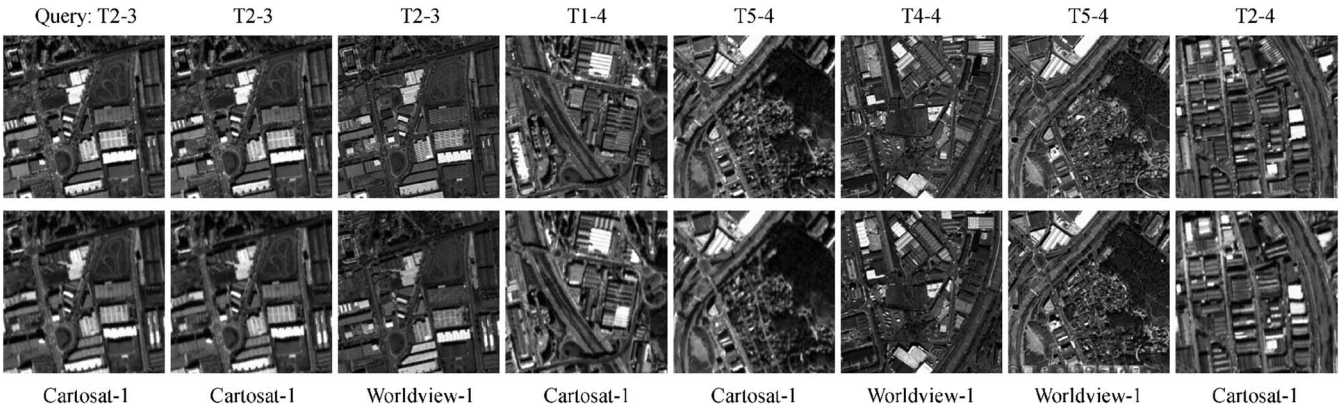


Fig. 13. Results by the ortho-NRFCMs-based retrieval method. Other settings are the same as those in Fig. 12.

The stability of height information for HrosSI retrieval can be investigated with Fig. 10. On the four RFI sizes, the integrated-NRFCMs-based retrieval method with three different weight values for planar features and height features was more stable than the DSM-NRFCMs-based retrieval method using only height features. Thus, height information alone is unstable for HrosSI retrieval, while a combination of height information and planar information is stable.

The stability of the RFI size is illustrated in Fig. 10. The 32×32 pixels size was slightly more stable than other three RFI sizes for all the three retrieval methods. Thus, features at the 32×32 pixels size, which were created from stereo

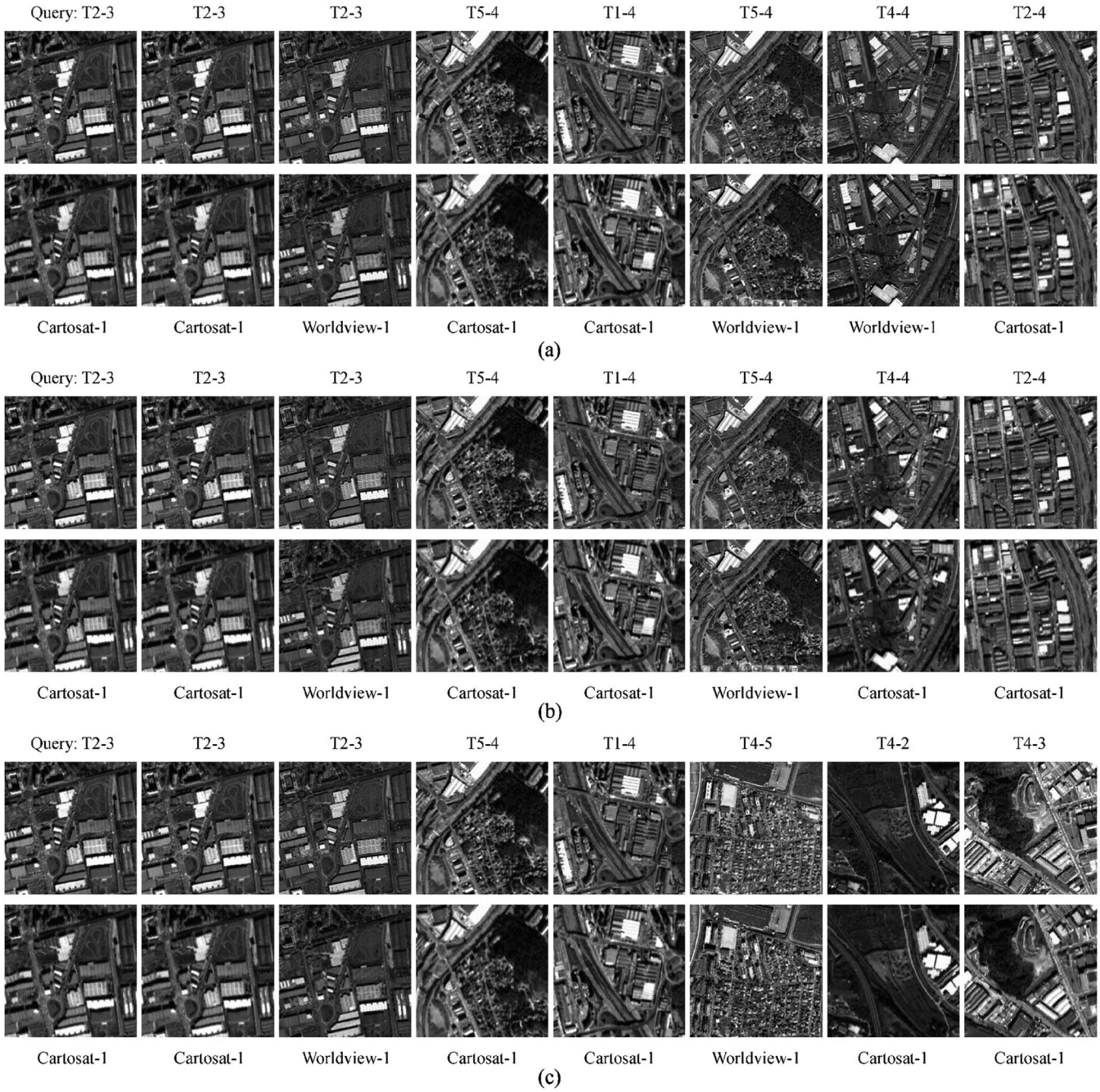


Fig. 14. Results by the integrated-NRFCMs-based retrieval method. (a) $d = 0.25$. (b) $d = 0.5$. (c) $d = 0.75$. Other settings are the same as those in Fig. 12.

products at 10 bits and at 2.5-m GSD, were chosen for retrieval accuracy analysis and retrieval efficiency analysis.

D. Retrieval Accuracy Analysis

The retrieval accuracy of three retrieval methods was calculated as follows. A CS1 or WV1 stereo pair block group named T2-3 was queried with the prototype system that deployed a retrieval method. The recall, precision, and F-measure were calculated based on the first 20 stereo pair block groups in the ranking list. The retrieval accuracy was calculated on the entire dataset, the CS1 and the WV1 subdatasets, respectively. Three retrieval accuracy modes for the retrieval method were obtained with a sensor type query. In this way, six retrieval accuracy modes for each retrieval method were obtained with two sensor type queries. Fig. 11 shows the six accuracy modes for the three retrieval methods.

Fig. 11 indicates that accuracy varies with different retrieval methods. The integrated-NRFCMs-based retrieval method had similar accuracy in each accuracy mode when three different weight values were used. The order of accuracy of the three retrieval methods remained constant with the same sensor type of query stereo pair block group and evaluated dataset. The order of accuracy of the three retrieval methods from

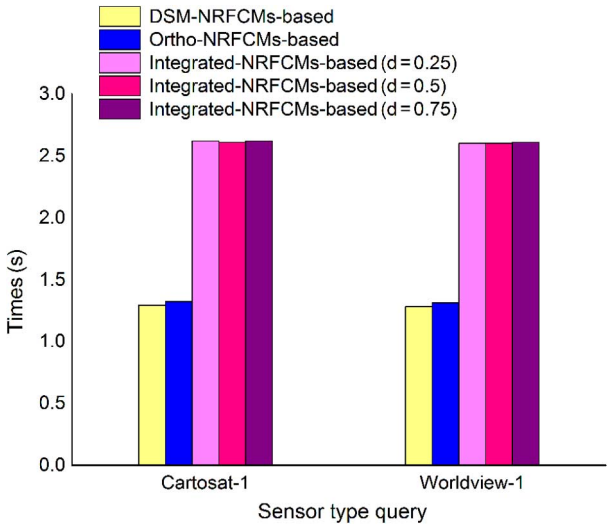


Fig. 15. Retrieval efficiency of the three retrieval methods. The retrieval efficiency of each retrieval method is calculated with two sensor type queries.

high to low was consistently: the integrated-NRFCMs-based retrieval method, the ortho-NRFCMs-based retrieval method, the DSM-NRFCMs-based retrieval method.

Figs. 12–14 show results of the three retrieval methods. Since similarity values between stereo pair block groups calculated by different retrieval methods were not comparable, similarity values between stereo pair block groups in the dataset and the query stereo pair block group are not given. The three retrieval methods can be compared with the visual results, especially with the most similar stereo pair block groups. For the three retrieval methods, the correlated degree of the first seven stereo pair block groups and the query group was in line with the retrieval efficiency. The first seven groups did not look so similar to the query group using the DSM-NRFCMs-based method, while they looked more similar using the ortho-NRFCMs-based method and the integrated-NRFCMs-based method. The WV1 stereo pair block group covering the same scene with the query CS1 stereo pair block group was not in the first seven similar groups using the DSM-NRFCMs-based method.

The significance of height information for HrosSI retrieval is evident in an efficiency comparison of the three retrieval methods. The integrated-NRFCMs-based retrieval method using planar features and height features was efficient. The accuracy of the DSM-NRFCMs-based retrieval method using only height features was inefficient. Thus, height information alone is inefficient for HrosSI retrieval, while a combination of height information and planar information is efficient.

E. Retrieval Efficiency Analysis

The retrieval efficiency of three retrieval methods was measured as follows. The CS1 or WV1 stereo pair block group named T2-3 was queried 10 times with the prototype system that deployed one retrieval method. The average time for the retrieval method with the sensor type query was then determined. In this way, the average time for the three retrieval methods with the two sensor types was calculated. Fig. 15 presents the average time for the three methods.

Fig. 15 shows that the time consumed by each retrieval method is independent of the sensor type query. With the same sensor type query, the ortho-NRFCMs-based method and the DSM-NRFCMs-based method consumed almost the same time, while the time consumed by the integrated-NRFCMs-based retrieval method was almost the sum of the time consumed by the other two methods. The integrated-NRFCMs-based retrieval method consumed almost the same time when three different weight values were used.

V. Discussion

The generic framework for content-based HrosSI retrieval is suitable for diverse HrosSIs with various characteristics. The proposed framework could be efficient for airborne as well for spaceborne stereo images used in this study. This is because retrieval is executed by similarity matching between the features extracted from the two primary stereo products, DSMs and orthoimages. Regardless of how stereo images are acquired, by what sensors, and under what acquisition conditions—the DSMs and orthoimages can be generated from the stereo images. In this study, features are NRFCMs created from the DSMs and orthoimages.

This study reveals the significance of height information for HrosSI retrieval. On one hand, height information alone is inefficient and unstable for HrosSI retrieval. The details of different land cover types vary in DSMs generated from stereo images of different remote sensors [48]. Some small land cover areas can only be extracted in some sensor types of stereo images. For instance, the residential area in Terrassa can be extracted from DSMs generated from Worldview-1 stereo pairs, but not from the Cartosat-1 stereo pairs [48]. For the same land cover area, height features extracted from different sensor types of stereo images may vary slightly. Moreover, different land cover types that vary significantly in a planar remotely sensed image may be very similar in the DSMs. For instance, both farmland and playgrounds have different texture characteristics but are very flat in the DSMs. Therefore, some land cover types are difficult to distinguish from DSMs alone. Height information is much less visual and intuitive than planar information when users distinguish between stereo images. On the other hand, a combination of height information and planar information is efficient and stable for HrosSI retrieval. Height features and planar features are complementary for each other in HrosSI retrieval. It may be efficient to use height features to distinguish between some land cover types, but inefficient to use planar features for discrimination, or the reverse.

This study has some disadvantages which can be divided into three categories according to the related issue: stereo products, features, and the prototype system. The stereo products are missing multispectral bands, and have no accuracy or filtering influence evaluation. In addition, HrosSI or stereo product characteristics have not been thoroughly evaluated. Considering that sensors with stereo capability often deliver panchromatic images, only the panchromatic images are used in the framework to support all HrosSIs. Multispectral bands

with multiview angles will be used in the future for refining the HrosSI retrieval accuracy. The accuracy of stereo products was not evaluated; however, our study clearly indicates that similarity matching between stereo products is efficient for HrosSI retrieval. The accuracy of stereo products and its influence on HrosSI retrieval will be investigated in future research. Stereo products were subjected to low-pass filtering with just one kernel size during preprocessing. We did not inspect the influence of the filtering method for stereo products or its parameters on HrosSI retrieval, which will be investigated further using other filtering methods with different parameters. We did not inspect the influence of HrosSI or stereo product characteristics on HrosSI retrieval, except issues related to GSD and radiometric resolution of stereo products. The influence of other HrosSI or stereo product characteristics, such as viewing angles and shadows, will be investigated later.

The feature types, and feature extraction opportunities in this study also have some limitations. Considering that RFD calculation and similarity matching are done based on certain conditions, HrosSIs are comparable only if HrosSIs and their stereo products are at the same size or at integer multiples of image size. Therefore, HrosSIs and their stereo products must be subjected to resampling. Other features that need not the resampling will be investigated. Both height features and planar features are NRFCMs in this study; that does not imply that they must be the same type of feature. They just need to be derived from DSMs and orthoimages, respectively. Since ranges of fractal dimension values derived from different land cover types may overlap each other, it may be difficult to distinguish between some HrosSIs using NRFCMs alone. This problem can be solved if other features are obtained from stereo products or from the HrosSIs themselves, and used in combination with the fractal dimension for HrosSI retrieval. Features were extracted offline and therefore feature extraction must be done in advance. This problem can be solved if features are extracted in real time.

There are also some limitations in the prototype system concerning the user interaction experience and test data. The prototype system is currently based on blocks, and requires users to search for query HrosSI block group. These problems and limitations can be solved if the system allows for user interaction, such as region-based HrosSI retrieval, query HrosSIs uploading, and relevance feedback. We examined only 216 stereo pairs from two sensors; more stereo images from more remote sensors will be involved to test whether or not the proposed techniques are robust.

The framework is specifically designed for HrosSIs retrieval. Existing remote sensing image retrieval methods concentrate on remotely sensed images with a single look angle, while this study concentrates on HrosSIs composed of a group of images with multiview angles. The characteristics of HrosSIs vary in tandem with diverse remote sensors and with acquisition conditions, such as viewing number and multiview angles. None of the existing remote sensing image retrieval methods is specifically designed for HrosSIs due to the fact that they do not take account or harness the special characteristics of HrosSIs. Our research exploits the unique characteristics of HrosSIs data so as to allow efficient and accurate content-based HrosSI retrieval. Our research takes height information into consideration, and reveals that height information is helpful for HrosSI retrieval. Height features must be involved to achieve high accuracy of HrosSI retrieval. Our research will provide important support for future work on HrosSI retrieval using height information.

VI. Conclusion

This paper proposes a new generic framework for content-based HrosSI retrieval, and investigates the significance of height information for HrosSI retrieval. HrosSI retrieval is executed by similarity matching between the features extracted from stereo products that are stereo-extracted DSMs and orthoimages. Experimental results show that the proposed techniques are efficient for HrosSI retrieval. Stereo products generated from various HrosSIs must be at the same GSD and radiometric resolution to obtain the most stable HrosSI retrieval. Height information alone is inefficient and unstable for HrosSI retrieval; however, a combination of height information and planar information is efficient and stable for HrosSI retrieval.

Our research is significant in that it extends the work on HrosSI retrieval. The proposed framework is efficient and suitable for spaceborne stereo images but might also be suitable for airborne stereo images as well. This new framework is specifically designed for HrosSIs retrieval because it takes advantage of the special characteristics of HrosSIs. More importantly, our research also reveals the significance of height information for HrosSI retrieval.

Some drawbacks of the proposed framework need to be addressed. The issues related to the influence of stereo products on HrosSI retrieval need to be investigated including: the accuracy of stereo products; HrosSI and stereo product characteristics; and stereo product preprocessing (such as filtering and resampling). Fractal dimension values and other features obtained from stereo products or HrosSIs with multispectral bands are extracted in real time and combined for HrosSI retrieval. The HrosSI retrieval system should include more stereo images from more remote sensors and involve user participation.

Acknowledgment

Special thanks are given to the data providers for the provision of the stereo datasets: Euromap for the Cartosat-1 data and Digital Globe for the Worldview-1 data. The authors would like to thank the anonymous reviewers for their very constructive comments.

References

[1] Y. Zhang, M. Zheng, J. Xiong, Y. Lu, and X. Xiong, "On-orbit geometric calibration of ZY-3 three-line array imagery with multistrip data sets," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 224–234, Jan. 2014.

[2] F. Pacifici and Q. Du, "Foreword to the special issue on optical multiangular data exploitation and outcome of the 2011 GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 3–7, Feb. 2012.

[3] D. Giribabu, S. Srinivasa Rao, and Y. V. N. Krishna Murthy, "Improving cartosat-1 DEM accuracy using synthetic stereo pair and triplet," *ISPRS J. Photogramm. Remote Sens.*, vol. 77, pp. 31–43, Mar. 2013.

[4] M. A. Aguilar, M. d. M. Saldana, and F. J. Aguilar, "Generation and quality assessment of stereo-extracted DSM from GeoEye-1 and WorldView-2 imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1259–1271, Feb. 2014.

[5] D. K. San and M. Turker, "Automatic building extraction from high resolution stereo satellite images," presented at *Proc. Conf. Inf. Extract. SAR Opt. Data Emphasize Dev. Countries*, Istanbul, Turkey, 2007 [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.117.5774

[6] D. Poli and I. Caravaggi, "3D modeling of large urban areas with stereo VHR satellite imagery: Lessons learned," *Nat. Hazards*, vol. 68, pp. 53–78, 2013.

[7] M. Koga and A. Iwasaki, "Improving the measurement accuracy of three-dimensional topography changes using optical satellite stereo image data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 8, pp. 2918–2923, Aug. 2011.

[8] J. Tian, P. Reinartz, P. d'Angelo, and M. Ehlers, "Region-based automatic building and forest change detection on Cartosat-1 stereo imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 79, pp. 226–239, May 2013.

[9] J. Tian, S. Cui, and P. Reinartz, "Building change detection based on satellite stereo imagery and digital surface models," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 406–417, Jan. 2014.

[10] J. E. Nichol, A. Shaker, and M.-S. Wong, "Application of high-resolution stereo satellite images to detailed landslide hazard assessment," *Geomorphology*, vol. 76, pp. 68–75, Jun. 2006.

[11] P. Gamba, "Human settlements: A global challenge for EO data processing and interpretation," *Proc. IEEE*, vol. 101, no. 3, pp. 570–581, Mar. 2013.

[12] D. Poli and T. Toutin, "Review of developments in geometric modelling for high resolution satellite pushbroom sensors," *Photogramm. Rec.*, vol. 27, pp. 58–73, Mar. 2012.

[13] Y. Feng, J. Ren, and J. Jiang, "Generic framework for content-based stereo image/video retrieval," *Electron. Lett.*, vol. 47, pp. 97–98, Jan. 2011.

[14] M. Quartulli and I. G. Olaizola, "A review of EO image information mining," *ISPRS J. Photogramm. Remote Sens.*, vol. 75, pp. 11–28, Jan. 2013.

[15] M. Datcu *et al.*, "Information mining in remote sensing image archives: System concepts," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 12, pp. 2923–2936, Dec. 2003.

[16] S. S. Durbha and R. L. King, "Semantics-enabled framework for knowledge discovery from earth observation data archives," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 11, pp. 2563–2572, Nov. 2005.

[17] Y. Li and T. R. Bretschneider, "Semantic-sensitive satellite image retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 853–860, Apr. 2007.

[18] S. Chi-Ren *et al.*, "GeoIRIS: Geospatial information retrieval and indexing system–content mining, semantics modeling, and complex queries," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 839–852, Apr. 2007.

[19] E. Aptoula, "Remote sensing image retrieval with global morphological texture descriptors," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 3023–3034, May 2014.

[20] D. Espinoza-Molina and M. Datcu, "Earth-observation image retrieval based on content, semantics, and metadata," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 11, pp. 5145–5159, Nov. 2013.

[21] M. Wang and T. Song, "Remote sensing image retrieval by scene semantic matching," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2874–2886, May 2013.

[22] Y. Yi and S. Newsam, "Geographic image retrieval using local invariant features," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 818–832, Feb. 2013.

[23] A. Samal, S. Bhatia, P. Vadlamani, and D. Marx, "Searching satellite imagery with integrated measures," *Pattern Recognit.*, vol. 42, pp. 2502–2513, Nov. 2009.

[24] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Comput. Surv.*, vol. 40, pp. 1–60, Apr. 2008.

[25] E. Bedoya, "Semantic assessment of similarity between raster elevation datasets," *Rev. Fac. Ing. Univ. Antioquia*, vol. 59, pp. 37–46, Jun. 2011.

[26] T. Toutin, "Review article: Geometric processing of remote sensing images: Models, algorithms and methods," *Int. J. Remote Sens.*, vol. 25, pp. 1893–1924, May 2004.

[27] B. Sirmacek, H. Taubenboeck, P. Reinartz, and M. Ehlers, "Performance evaluation for 3-D city model generation of six different DSMs from air- and spaceborne sensors," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 59–70, Feb. 2012.

[28] T. Toutin, "Comparison of stereo-extracted DTM from different high-resolution sensors: SPOT-5, EROS-a, IKONOS-II, and QuickBird," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2121–2129, Oct. 2004.

[29] T. Toutin, "State-of-the-art of geometric correction of remote sensing data: A data fusion perspective," *Int. J. Image Data Fusion*, vol. 2, pp. 3–35, 2011.

[30] M. A. Aguilar, F. Agüera, F. J. Aguilar, and F. Carvajal, "Geometric accuracy assessment of the orthorectification process from very high resolution satellite imagery for common agricultural policy purposes," *Int. J. Remote Sens.*, vol. 29, pp. 7181–7197, Dec. 2008.

[31] M. A. Aguilar, M. d. M. Saldaña, and F. J. Aguilar, "Assessing geometric accuracy of the orthorectification process from GeoEye-1 and WorldView-2 panchromatic images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 21, pp. 427–435, Apr. 2013.

[32] L. Zhang and A. Gruen, "Multi-image matching for DSM generation from IKONOS imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 60, pp. 195–211, May 2006.

[33] T. Toutin, C. V. Schmitt, and H. Wang, "Impact of no GCP on elevation extraction from WorldView stereo data," *ISPRS J. Photogramm. Remote Sens.*, vol. 72, pp. 73–79, Aug. 2012.

[34] P. Reinartz, R. Müller, P. Schwind, S. Suri, and R. Bamler, "Orthorectification of VHR optical satellite data exploiting the geometric accuracy of TerraSAR-X data," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, pp. 124–132, Jan. 2011.

[35] W. Sun, G. Xu, P. Gong, and S. Liang, "Fractal analysis of remotely sensed images: A review of methods and applications," *Int. J. Remote Sens.*, vol. 27, pp. 4963–4990, Nov. 2006.

[36] H. Taud and J.-F. Parrot, "Measurement of DEM roughness using the local fractal dimension," *Géomorphol.: Relief Processus Environ.*, vol. 11, pp. 327–338, 2005.

[37] S. Peleg, J. Naor, R. Hartley, and D. Avnir, "Multiple resolution texture analysis and classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 4, pp. 518–523, Jul. 1984.

[38] C. H. Grohmann, M. J. Smith, and C. Riccomini, "Multiscale analysis of topographic surface roughness in the midland valley, Scotland," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 4, pp. 1200–1213, Apr. 2011.

[39] S. De Jong and P. Burrough, "A fractal approach to the classification of Mediterranean vegetation types in remotely sensed images," *Photogramm. Eng. Remote Sens.*, vol. 61, pp. 1041–1053, 1995.

[40] W. Ju and N. S. N. Lam, "An improved algorithm for computing local fractal dimension using the triangular prism method," *Comput. Geosci.*, vol. 35, pp. 1224–1233, Jun. 2009.

[41] E. Remias, G. Sheikholeslami, and Z. Aidong, "Block-oriented image decomposition and retrieval in image database systems," in *Proc. Int. Workshop Multimedia Database Manage. Syst.*, 1996, pp. 85–92.

[42] S. W. Myint, "Fractal approaches in texture analysis and classification of remotely sensed data: Comparisons with spatial autocorrelation techniques and simple descriptive statistics," *Int. J. Remote Sens.*, vol. 24, pp. 1925–1947, 2003.

[43] R. Lopes and N. Betrouni, "Fractal and multifractal analysis: A review," *Med. Image Anal.*, vol. 13, pp. 634–649, 2009.

[44] S. Novianto, Y. Suzuki, and J. Maeda, "Near optimum estimation of local fractal dimension for image segmentation," *Pattern Recognit. Lett.*, vol. 24, pp. 365–374, 2003.

[45] A. P. Pentland, "Fractal-based description of natural scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 661–674, Nov. 1984.

[46] J. S. Payne, L. Heppelwhite, and T. J. Stonham, "Perceptually based metrics for the evaluation of textural image retrieval methods," in *Proc. IEEE Int. Conf. Multimedia Comput. Syst.*, 1999, pp. 793–797.

[47] WG I/4 ISPRS. (2013, Dec. 12). *ISPRS Stereo Benchmark Test* [Online]. Available: http://www2.isprs.org/commissions/comm1/wg4/benchmark-test.html

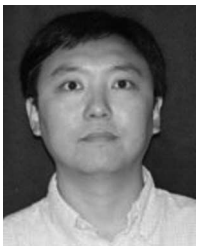
[48] P. Reinartz *et al.*, "Benchmarking and quality analysis of dem generated from high and very high resolution optical stereo satellite data," in *Proc. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XXXVIII, Calgary, Canada, 2010 [Online]. Available: http://www.isprs.org/proceedings/XXXVIII/part1/09/09_03_Paper_33.pdf

[49] P. d'Angelo and P. Reinartz, "Semiglobal matching results on the ISPRS stereo matching benchmark," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XXXVIII-4/W19, pp. 79–84, 2011.

[50] ENVI. (2014, May 29). *ENVI Tutorials* [Online]. Available: http://www.exelisvis.com/Learn/Resources/Tutorials.aspx

Feifei Peng is currently working toward the Ph.D. degree at the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China.

His research interests include remote sensing image processing and information extraction.



Le Wang received the B.E. degree from Wuhan Technical University of Surveying and Mapping, Wuhan, China, in 1996, the M.S. degree from Peking University, Beijing, China, in 1999, and the Ph.D. degree in environmental science, policy, and management from the University of California, Berkeley, in 2003.

He is an Associate Professor with the Department of Geography, The State University of New York at Buffalo, Buffalo, New York, NY, USA. He is an Editor of the *International Journal of Remote Sensing* and the Vice Chair of the Remote Sensing Speciality Group at the American Association of Geographers. His research interests include remote sensing image classification, urban population estimation, invasive species spread modeling, and coastal mangrove forest characterization.

Jianya Gong received the Ph.D. degree in photogrammetry and remote sensing from Wuhan Technical University of Surveying and Mapping, Wuhan, China, in 1992.

He is a Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China. He is an Academician of the Chinese Academy of Sciences. He is the President of Commission VI of the International Society for Photogrammetry and Remote Sensing. His research interests include remote sensing image processing, spatial data infrastructure, geospatial data sharing and interoperability.

Huayi Wu received the B.Sc. and M.Sc. degrees in mathematical statistics from Fudan University and Wuhan University, China, in 1988 and 1991, respectively and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, China, in 1999.

He is a Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China. He is a Distinguished Professor of Cheung Kong Scholars Program. His research interests include geospatial information sharing and interoperability, remote sensing image processing, and high-performance geospatial computing.