Article

An Improved Density-Based Time Series Clustering Method Based on Image Resampling: A Case Study of Surface Deformation Pattern Analysis

Yaolin Liu [1,2,3], Xiaomi Wang [1,*], Qiliang Liu [4], Yiyun Chen [1,5,6] and Leilei Liu [7,*]

[1] School of Resource and Environment Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China; yaolinliuwhu@163.com (Y.L.); chenyy@whu.edu.cn (Y.C.)
[2] Key Laboratory of Geographic Information System, Ministry of Education, Wuhan University, 129 Luoyu Road, Wuhan 430079, China
[3] Collaborative Innovation Center of Geospatial information technology, Wuhan University, 129 Luoyu Road, Wuhan 430079, China
[4] Department of Geo-informatics, Central South University, Changsha 410012, China; qiliang.liu@csu.edu.cn
[5] State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China
[6] Suzhou Institute of Wuhan University, Suzhou 215123, Jiangsu, China
[7] Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong 999077, China
* Correspondence: xiaomiw@yeah.net (X.W.); csulll@foxmail.com (L.L.)

Abstract: Time series clustering algorithms have been widely used to mine the clustering distribution characteristics of real phenomena. However, these algorithms have several limitations. First, they depend heavily on prior knowledge. Second, the algorithms do not simultaneously consider the similarity of spatial locations, spatial-temporal attribute values, and spatial-temporal attribute trends (trends in terms of the change direction and ranges in addition and deletion over time), which are all important similarity measurements. Finally, the calculation cost based on these methods for clustering analysis is becoming increasingly computationally demanding, because the data volume of the image time series data is increasing. In view of these shortcomings, an improved density-based time series clustering method based on image resampling (DBTSC-IR) has been proposed in this paper. The proposed DBTSC-IR has two major parts. In the first part, an optimal resampling scale of the image time series data is first determined to reduce the data volume by using a new scale optimization function. In the second part, the traditional density-based time series clustering algorithm is improved by introducing a density indicator to control the clustering sequences by considering the spatial locations, spatial-temporal attribute values, and spatial-temporal attribute trends. The final clustering analysis is then performed directly on the resampled image time series data by using the improved algorithm. Finally, the effectiveness of the proposed DBTSC-IR is illustrated by experiments on both the simulated datasets and in real applications. The proposed method can effectively and adaptively recognize the spatial patterns with arbitrary shapes of image time series data with consideration of the effects of noise.

Keywords: time series clustering; time series resampling; density-based clustering; spatial data mining; surface deformation patterns

1. Introduction

Time series data are pervasive in various areas that range from science, business, finance, economics, health care, and engineering; such data include time series deformation images monitored by the Synthetic Aperture Radar Interferometry (INSAR) technique. Image data is the most extensive type of data source in geography. Hence, in this study, image time series data is the analyzed data source. The location of a pixel with its spatial-temporal attributes in the image time series data is defined as a sequence. These sequences are generally characterized by obvious spatial heterogeneity. Thus, mining the spatial clustering characteristics of the image time series data is extremely important in exploring the potential distribution mechanism that underlies the data.

In the past few decades, many time series clustering algorithms have been developed. These algorithms can be roughly grouped into three classes, as follows: partition-based time series clustering algorithms [1–5], hierarchical time series clustering algorithms [6–9], and density-based time series clustering algorithms [10,11]. The K-means time series clustering algorithm [4] and fuzzy c-means time series clustering algorithm [5] are the typical partition-based time series clustering algorithms; many other partition-based time series clustering algorithms [1,3] are derived from these two algorithms to improve efficiency and accuracy. Compared with the partition-based time series clustering algorithms, hierarchical time series clustering algorithms [6–9] do not need the initial centers of the clusters, which are difficult to set and significantly affect the eventual results. However, both the partition-based time series clustering algorithms and hierarchical time series clustering algorithms do not consider the neighboring relationships, and the clusters obtained by these algorithms disperse in the spatial domain with no clear visualization. To recover these shortcomings, density-based time series clustering algorithms [10,11] are proposed to consider the neighborhood, which cluster neighboring objects with similar time series attributes. However, several parameters must be set by users in density-based time series clustering algorithms. During the clustering mining procedure, a priori knowledge is always lacking and the proper parameters are difficult to set. Recently, several other studies have been conducted on time series data, such as the temporal self-organizing feature maps time series clustering algorithm [12], hidden Markov-based time series clustering algorithm [13], and time series co-clustering algorithms [14,15]. However, these studies all heavily depend on the parameters of clustering or another subjective influence, in which adaptively obtaining satisfactory results is difficult. In summary, existing time series clustering algorithms still have several common shortcomings when dealing with clustering analysis in complicated real applications. First, these algorithms generally consider either spatial-temporal attribute values [16] or spatial-temporal attribute trends [17] to measure the similarity between sequences. The similarity measurement of spatial-temporal attribute values indicates that for two sequences, a smaller difference between the attributes of two sequences in each image indicate higher similarity of the two sequences. The spatial-temporal attribute trends indicate the overall change direction and ranges in addition and deletion over time. Increased similarity of the change direction and ranges between sequences indicates higher similarity between the two sequences. In real applications, the sequences should be regarded as similar in the non-spatial domain when both the spatial-temporal attribute values and spatial-temporal attribute trends between sequences are similar, because sequences with similar attribute trends and different attribute values or similar attribute values and different attribute trends exist in real applications. For example, areas with monsoon climates of medium latitudes are rainy during summer and dry during winter. The rainfall trends in these areas are similar. However, the rainfall capacity depends on the location. Hence, rainfall trends and rainfall capacity should be considered simultaneously to mine areas with similar rainfall. Second, apart from the spatial-temporal attributes, the spatial locations are also important data attributes. In geography, data generally exhibit significant spatial heterogeneity, which commonly results in non-overlapped clusters. To obtain clusters that are non-overlapping to each other with arbitrary geometrical shapes, spatial locations should be considered to construct the spatial proximity relationships between sequences during the clustering procedure. Third, the results of existing clustering algorithms are largely affected by predefined parameters and depend on prior

knowledge, which is generally unavailable in real applications. Finally, image time series data with unequal time intervals are always the true phenomena, which have been ignored by most existing time series clustering algorithms.

To overcome such shortcomings, proposing an adaptive time series clustering method that can consider spatial locations, spatial-temporal attribute values, and spatial-temporal attribute trends simultaneously without the need to set parameters by default or prior knowledge is necessary. Hence, an improved density-based time series clustering method based on image resampling (DBTSC-IR) is proposed in this paper. First, the image resampling method, which may effectively decrease the dataset volume [18] and maintain an acceptable level of information for applications when data is resampled at a representative scale [19], has been adopted to decrease the spatial dimensionality of data to reduce the time cost of further clustering analysis. In this procedure, the image time series data are resampled to a suitable scale by using a scale optimization function. Then, an improved density-based image time series clustering algorithm (DBTSC), with the help of a density-based spatial clustering algorithm (DBSC) [20], is proposed to be conducted on the resampled time series data. The DBTSC algorithm attempts to adaptively obtain clusters forming non-overlapping compact regions with similar spatial-temporal attribute values and trends under the interference of noise. In addition, to consider the dataset with unequal time intervals, the similarity measurements for the attributes are redefined by adding the strategy of weightings.

The remainder of the paper is organized as follows. The similarity measurements between the sequences are first described in Section 2. In Section 3, the resampling method of the image time series data is described in detail. In Section 4, the DBTSC algorithm is fully performed. In Section 5, the evaluation methods for the clustering results are adopted to evaluate the accuracy of the clustering algorithms. In Section 6, validation of DBTSC-IR is conducted based on simulated datasets and real applications. Section 7 summarizes the main findings and emphasizes the directions for future work.

2. Similarity Measurements between Sequences

One key component in image time series clustering is the similarity measurements between sequences. The similarity between sequences involves the similarity of locations, the similarity of spatial-temporal attribute values, and the similarity of spatial-temporal attribute trends. Theoretically, any two sequences are regarded as similar if all of the three measurements between them are similar. According to Liu et al. [21], the similarity measurements for spatial locations and attributes should be considered interdependently. In the spatial domain, Euclidean distance is generally adopted to measure the similarity degree between sequences. In the non-spatial domain, the similarity of spatial-temporal attribute values is generally defined as the mean value of the difference of the attribute values of every time interval. The similarity of spatial-temporal attribute trends can be measured by correlation coefficients such as the Pearson coefficient and Spearman coefficient [22]. These coefficients are the most popular for correlation analysis. Nonetheless, because the Spearman correlation coefficient can generally describe the correlation relationship between linear and nonlinear variables with higher accuracy, especially for uncertain distributions [23,24], it is thus adopted to analyze the similarity of the spatial-temporal attribute trends between the two sequences in this study. However, existing similarity measurements are suitable for spatial-temporal attributes with equal time intervals, but in real applications, an image time series dataset always contains images with unequal time intervals. Hence, the weighted similarity measuring functions are proposed in Equations (1) and (2). In general, the larger the time interval between neighboring images is, the greater the proportion the interval is considered to occupy. Thus, the weight is set as the length proportion of the time interval.

For sequences l_1 and l_2, the similarity of the spatial-temporal attribute values between two sequences is defined as follows:

$$D(l_1, l_2) = \sum_{i=1}^{T} w(t(i)) \times \left| l_1^{t(i)} - l_2^{t(i)} \right| \tag{1}$$

where $\mathrm{w}(t(i)) = {}^{(t(i)-t(i-1))}/_{(t(T)-t(0))}$; $t(0)$ is the base time of the image time series data; $t(i)$ is the time point of the ith image; T is the count of images in the image time series data; $l_1^{t(i)}$ is the variation of spatial-temporal attribute values of l_1 from $t(i-1)$ to $t(i)$.

To combine the characteristics of time series data with unequal time intervals, the Spearman correlation coefficient is modified by adding the strategy of weightings and is denoted as scw.

$$\mathrm{scw}(l_1, \, l_2) = 1 - \frac{6 \sum_{i=1}^{T} \mathrm{w}(t(i)) \left(l_1^{t(i)} - l_2^{t(i)} \right)^2}{T^2 - 1} \tag{2}$$

In order to verify the proposed scw, two sequences with similar attribute trends and un-equal time intervals (in Figure 1) are set as an example. The Spearman coefficient and scw are both used to measure the trend similarity between the sequences in Figure 1. According to Ramirez-Lopez et al. [25], the attribute trend between sequences is considered a positive correlation when its correlation coefficient is larger than 0.5 with a significant level scw_sig less than 0.1. The results of the two correlation coefficients between sequences in Figure 1 are as follows: the Spearman correlation coefficient is 0.449 with a significance level scw_sig = 0.092 and scw is 0.892 with a scw_sig = 0.001. The results indicate that scw is more consistent with the distribution features and can be used to measure the spatial-temporal attribute trend similarity between sequences.
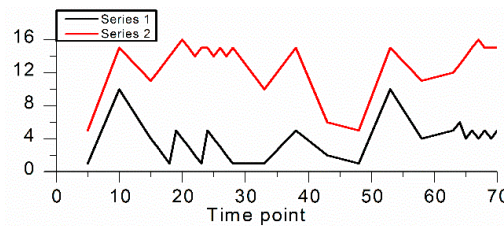

Figure 1. Distribution of two sequences with unequal time intervals.

3. Image Resampling

The large data volume of image time series data induces a great computational burden for clustering analysis. To achieve efficient clustering pattern detection, the dataset can be resampled to a representative minor scale (resolution value). A suitable spatial scale can prevent excessive data redundancy and can effectively express the data information. Existing optimal image scale selection methods [18,19] merely consider the difference between resampled pixels and ignore the difference of values in the resampled pixels. To enhance the veracity of the selection of the optimal resampling scale, a scale optimization function (Equations (3) and (4)) is proposed based on the following principle: an image with a suitable scale has a small standard deviation within resampled intra-pixels and a large coefficient of variation between resampled inter-pixels. To realize the image resampling, an optimizing flow is designed to search the optimal resampling scale by using the proposed scale optimization function (Equations (3) and (4)), as shown in Figure 2 (*u* is the number of selected scales).

$$S(k) = \sum_{i=1}^{T} \frac{\sum_{m=1}^{N} \frac{std\_inner(m)_{t(i)}^{k}}{std\_intra(m)_{t(i)}^{k}}}{N} \tag{3}$$

$$R = \sqrt{k} \times r \tag{4}$$

where *r* is the resolution value of the images in the image time series data; *R* is the resolution value of the resampled images; *N* is the number of pixels in the resampled image; $std\_inner(m)_{t(i)}^{k}$ is the standard deviation of the values of pixels within the resampled *m*th pixel in the resampled *i*th images, which is calculated as:

$$std\_inner(m)_{t(i)}^k = \sqrt{\frac{\sum_{q=1}^{k} \left(p_m^{t(i)}(q) - cell^{t(i)}(m)\right)^2}{k-1}}. \tag{5}$$

where $\sum_{q=1}^{k} p_m^{t(i)}(q)$ are the values of the pixels in the original ith image, which are contained in the mth pixel in the resampled ith image; k is the number of original pixels (in the original ith image) contained in a resampled pixel (in the resampled ith image); $cell^{t(i)}(m)$ represents the mth pixel in the resampled ith image; $std\_intra(m)_{t(i)}^k$ is the coefficient of variation between the mth pixel and its neighboring pixels in the resampled ith image, which is calculated as follows:

$$std\_intra(m)_{t(i)}^k = \sqrt{cell^{t(i)}(m) - mean^{t(i)}(m))^2}. \tag{6}$$

where $mean^{t(i)}(m)$ is the mean value of the spatial-temporal attribute values of the neighboring pixels of the mth pixel in the resampled ith image. The neighboring pixels are set as the eight-connected pixels.

The scale optimization function indicates that if the difference of the non-spatial values between resampled pixels is larger, and the variation of values in the resampled pixels is smaller, the scale is considered better with a smaller value of the scale optimization function. Thus, the output value of the resolution value R is optimal when the value of the scale optimization function is at a minimum. The time cost of the procedure is roughly $O(u \times 2n)$, where n is the number of pixels in the original image.
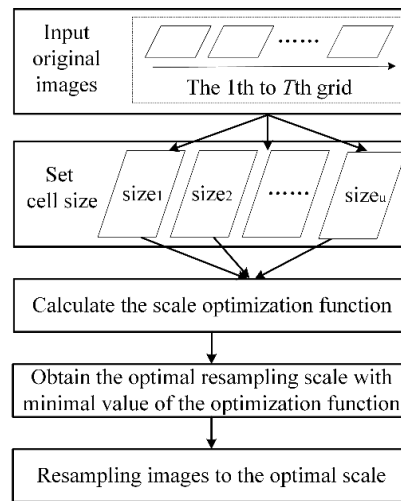

Figure 2. Procedure of resampling the image time series data.

4. Image Time Series Clustering

To detect time series clusters with similar spatial locations, spatial-temporal attribute values, and spatial-temporal attribute trends, DBTSC is proposed with the help of the DBSC algorithm [20]. The DBSC algorithm has been proven to be effective at mining clusters with similar attributes when spatial heterogeneity is considered. Two main strategies contribute to the feasibility of the DBSC algorithm, as follows. (1) The attributes are considered in the spatial domain and non-spatial domain separately. Given that the spatial locations and attributes are dependent features of objects, the separate consideration of spatial locations and attributes is necessary; (2) The density indicator helps the algorithm to obtain the unique optimal clusters with similar locations and attributes. Hence, to effectively detect time series clusters, the DBTSC algorithm introduces the abovementioned two strategies and integrates the spatial-temporal attributes. The DBTSC algorithm mainly consists of two parts. In Part 1, in the spatial domain, sequences with proximity relationships are considered similar

in the spatial domain. According to Heng et al. [26], the eight-connected sequences can be considered as proximity sequences for the image time series data. In Part 2, in the non-spatial domain, and based on the proximity relationships, clusters with neighboring objects, similar spatial-temporal attribute values, and spatial-temporal attribute trends are detected by using an improved density based time series clustering method, which is realized by integrating the strategy of the density indicator [20] and the proposed similarity measurements in Section 2. The above analysis indicates that the DBTSC algorithm can be summarized as the following five steps. The entire procedure of DBTSC is shown in Figure 3.
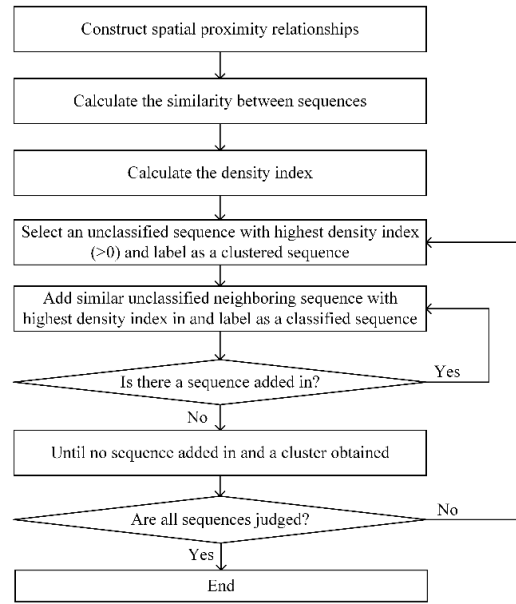

Figure 3. The procedure of the improved density-based image time series clustering algorithm (DBTSC) algorithm.

Step 1   The spatial proximity relationships between sequences are constructed. As is well known, the pixels in the images are regularly distributed in the spatial domain, and the neighboring eight-connected sequences of l_1 are considered the neighbors $ND(l_1)$ of l_1.

Step 2   The similarity degree D of the spatial-temporal attribute values and similarity degree *smw* of the spatial-temporal trends between neighboring sequences are calculated. In addition, the default value of the spatial-temporal attribute threshold TS of the density-based clustering method can be determined during the procedure by the rule of three standard deviations [20]. TS is used to judge whether the sequences have similar spatial-temporal attribute values, and TS is used in Step 3.

Step 3   The density indicator is computed. The computation procedure can be divided into two sub-steps as follows:

(1)   For every sequence, the spatially directly reachable sequences are calculated, defined as follows. Taking sequences l_1 and l_2 as an example, l_2 is spatially directly reachable from l_1 if the following constraints are satisfied:

$$\begin{cases} \mathrm{D}(l_1, l_2) < TS \\ \mathrm{scw}(l_1, l_2) > 0.5 \\ \mathrm{scw\_sig}(l_1, l_2) < 0.1 \end{cases} \tag{7}$$

(2)　　The density indicator DI of the sequences is calculated. For sequence l_1, the density indicator is calculated as

$$DI(l(m_1)) = N_{sdr}(l_1) + N_{sdr}(l_1)/n\_ND(l_1) \qquad (8)$$

where $N_{sdr}(l_1)$ is the number of sequences that are spatially directly reachable from l_1. $n\_ND(l_1)$ is the total number of neighbors of l_1.

Step 4　Time series clustering is implemented. This step can be summarized as the following four operations:

(1)　　An unclassified sequence l_i is selected with the highest indicator value (larger than zero); this is defined as a temp cluster CLU. Meanwhile, the selected sequence l_i is labeled as a classified sequence.

(2)　　An unclassified sequence l_j is added. If the sequence l_j meets the following three conditions, it is added to CLU and is labeled a classified sequence.

Condition 1: l_j is spatially directly reachable from any sequence in CLU.

Condition 2: $D(l_j, Avg(CLU)) < TS$.

Condition 3: $DI(l_j) \geq DI(l_a) \, l_j \in ND(l_o) \, \& \, l_o \in CLU \, \& l_a \notin CLU$.

(3)　　Operation (2) is repeated, and the cluster CLU is then obtained and Operation (4) is conducted until no sequence can be added to CLU.

(4)　　Operation (1) is repeated, and the procedure is stopped when all sequences have been determined. The sequence, which does not belong to any cluster, is recognized as noise.

The time complexity of steps 1, 2, and 3 are $O(N)$, $O(2T \times N)$, and $O(Nlog(N))$, respectively. The time complexity of step 4 is linear in N. Thus, the total time complexity of the DBTSC algorithm is $O(Nlog(N))$.

5. Evaluation of the Clustering Results

Measurement indexes including Rand, recall, and precision are generally derived to assess the cluster detection approaches [27,28]. According to Manning et al [27] and Grubesic et al [28], the Rand index assesses the ability of a particular cluster detection approach to find the known clusters and noises; the recall index evaluates the ability of the clustering algorithm to identify positive detection success; and the precision index captures the subtleties of the clustering algorithm.

For any two clustering results $r1$ and $r2$ of the same dataset, Rand, recall, and precision are denoted as $\{ Rand(r1), recall(r1), precision(r1) \}$ and $\{ Rand(r2), recall(r2), precision(r2)\}$, respectively. If the indexes meet one of the following criteria, then the clustering result $r1$ is regarded as the better result.

Criterion 1: $recall(r1) > recall(r2)$ and $precision(r1) > precision(r2)$.

Criterion 2: $recall(r1) > recall(r2)$, $precision(r1) < precision(r2)$, and $Rand(r1) > Rand(r2)$.

6. Results and Discussion

In order to verify the effectiveness and accuracy of DBTSC-IR, four experiments on simulated datasets and real applications are conducted. In the first experiment, a simulated dataset is set to verify the accuracy of the DBTSC algorithm in comparison with typical time series clustering algorithms, as shown in Section 6.1. In the second experiment, to validate the feasibility of the proposed similarity measurements that have been described in Section 2, the clustering results based on the proposed similarity measurements are compared with those obtained based on the typical similarity measurements. The second experiment is demonstrated in detail in Section 6.2. In the

third experiment, several simulated datasets are designed to evaluate the performance of DBTSC-IR, described in Section 6.3. Finally, the DBTSC-IR algorithm is used for pattern analysis on surface deformation data. Several interesting patterns that cannot be effectively detected by other classical time series clustering algorithms have been found. The detailed description of the real application is shown in Section 6.4.

6.1. Validation of the DBTSC Algorithm

A simulated dataset *SD* is designed to evaluate the performance of the proposed time series clustering algorithm DBTSC. The characteristics of the simulated dataset are as follows:

(1)   The time series dataset with equal time intervals holds eleven images, as shown in Figure 4.
(2)   Nine predefined clusters are labeled S_1 to S_9 (in Figure 5a). In every image, the spatial-temporal attribute values of the same cluster are randomly distributed to one range, and the mean of the spatial-temporal attribute values of each predefined cluster in every image is labeled in Figure 5b.
(3)   To simulate the actual situations, four types of noise are set and distributed in the gray-colored areas in Figure 5a. Type 1 comprises the randomly distributed noise, such as the noise in the gray-colored band N_1 with spatial-temporal attribute values randomly distributed between 1 and 26. Type 2 comprises the gradiently distributed noises. For example, the spatial-temporal attribute values of noise in the gray-colored band N_2 (Figure 5a) gradually change from 1 to 26. Type 3 comprises the noises of spatial-temporal attribute values, such as p_1 (Figure 5a). The characteristics of this type of noise are as follows: (1) having similar spatial-temporal attribute trends with neighboring pixels; and (2) having significantly different spatial-temporal attribute values with neighboring pixels. Type 4 comprises the noises of spatial-temporal attribute trends, such as p_2 (Figure 5a). This type of noise has significantly different spatial-temporal attribute trends and similar spatial-temporal attribute values with neighboring pixels.
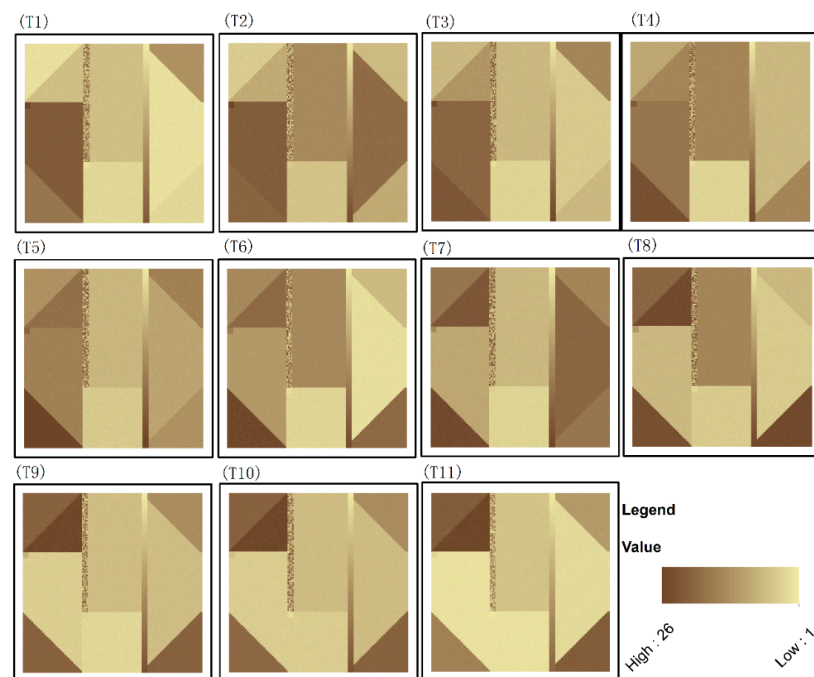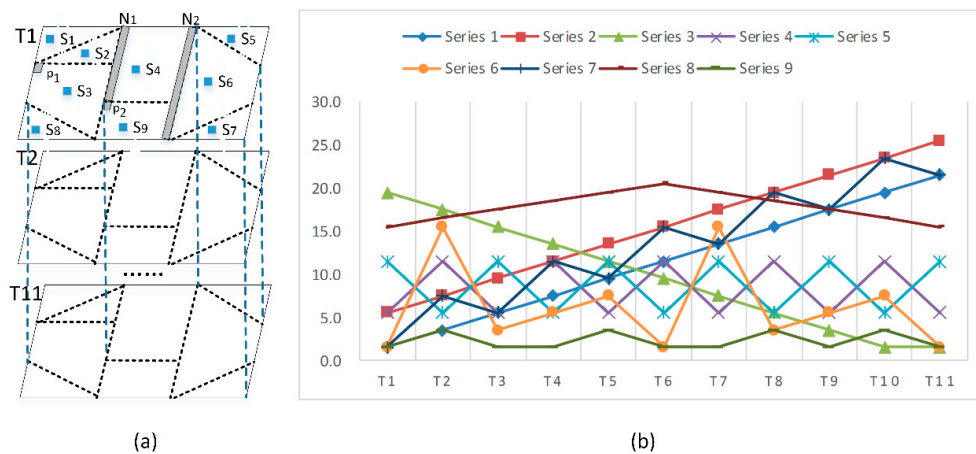

Figure 4. Simulated dataset *SD* with 11 images.

(a)

(b)

Figure 5. The simulated dataset SD: (**a**) the spatial distribution of the predefined clusters and noises; (**b**) the spatial-temporal attribute values of the clusters.

For comparison, three classical time series clustering algorithms, i.e., the k-means time series clustering algorithm [1], the fuzzy c-means time series clustering algorithm [3], and the density-based time series clustering algorithm [11], are also applied to the dataset *SD*. For the k-means time series clustering algorithm and fuzzy c-means time series clustering algorithm, the parameter k (the predefined number of clusters) is set as nine. The density-based time series clustering algorithm requires an input attribute threshold parameter, the most appropriate value for which is obtained by performing a parametric study.

The results of DBTSC and the above-mentioned classical time series clustering algorithms are shown in Figure 6. The figure indicates that the fuzzy c-means time series clustering and k-means time series clustering algorithms cannot recognize the clusters with arbitrary shapes. The pixels in one cluster may be distributed in a cluttered manner in the spatial domain. By contrast, apart from the DBTSC algorithm, the other algorithms are not robust to noise. For example, the fuzzy c-means and k-means time series clustering algorithms cannot recognize the various types of noise, and the density-based time series clustering algorithm wrongly detects the noise of types 2 and 4. Furthermore, the accuracy values of the results (in Figure 7) show that the DBTSC algorithm can more accurately detect the time series clusters than do the classical time series clustering algorithms.
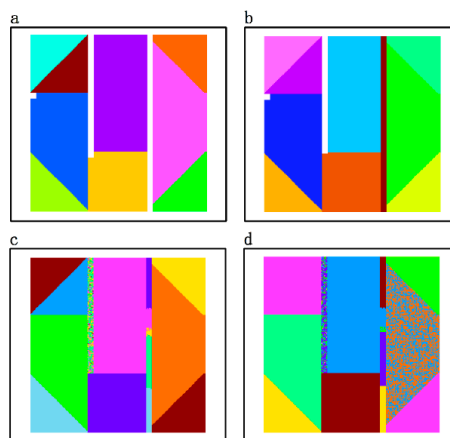

Figure 6. Clustering results of DBTSC and classical time series clustering algorithms on *SD*: (**a**) Result of the DBTSC algorithm; (**b**) Result of the density-based time series clustering algorithm; (**c**) Result of the fuzzy c-means time series clustering algorithm; (**d**) Result of the k-means time series clustering algorithm.
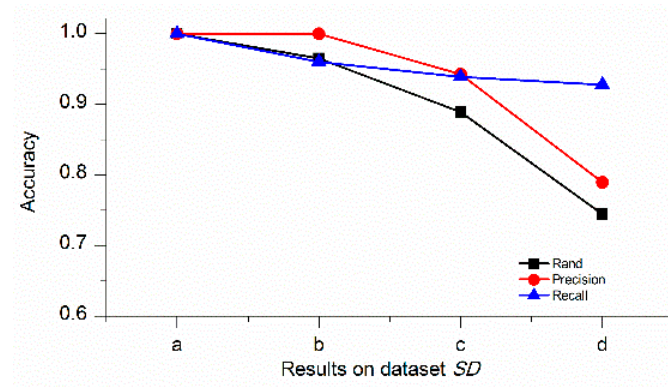
Figure 7. Accuracy values of the clustering results in Figure 6.

The abovementioned experiments demonstrate the advantages of the DBTSC algorithm. The clustering results of the comparative experiments show that the DBTSC algorithm can recognize the predefined clusters and noises with high accuracy. Furthermore, the DBTSC algorithm adapts to datasets with randomly distributed attributes, arbitrary geometrical shapes, and noises.

6.2. Comparison of the DBTSC Algorithm with Typical Similarity Measurements and the Proposed Similarity Measurements

In order to verify the effectiveness of the proposed spatial-temporal attribute similarity measurements proposed in Section 2, DBTSC with typical spatial-temporal attribute similarity measurements and the proposed spatial-temporal attribute similarity measurements are both conducted on the simulated dataset SD (in Figure 4) for comparison. The DBTSC with typical spatial-temporal attribute similarity measurements consists of two situations. One considers the spatial-temporal attribute values, and the mean attribute difference (the mean value of the difference of the spatial-temporal attribute values of every time interval) is applied to measure the similarity degree between sequences. To realize the DBTSC algorithm with the mean attribute difference, the method of calculating the spatial directly reachable sequences (in part 1 of step (3) in Section 4) in DBTSC should be changed as follows: for sequences $l(m_1)$ and $l(m_2)$, if $D(l(m_1), l(m_2)) < \text{TS}$, $l(m_2)$ is spatially directly reachable from $l(m_1)$. Another considers the spatial-temporal attribute trends. As a matter of course, the Spearman coefficient is used to measure the similarity degree between sequences without considering the spatial-temporal attribute values, which can be realized by setting the value of TS (in DBTSC algorithm) to positive infinity.

The DBTSC with the proposed similarity measurements, with the Spearman coefficient and with the mean attribute difference for comparison purposes are applied to SD, and the results are shown in Figures 8 and 9. The result in Figure 8b shows that the DBTSC with mean attribute difference is unable to identify the noises of type 4, such as noises p_2. The result in Figure 8c shows that when the Spearman coefficient is used to measure the similarity degree between the sequences, the clusters and noises with similar spatial-temporal attribute trends are wrongly detected as the same cluster, although they have significantly different spatial-temporal attribute values. For example, clusters S_1 and S_2 are falsely detected as the same cluster, and noises p_1 are wrongly recognized as a part of the cluster S_3. In summary, the results of the comparison experiments clearly indicate that the DBTSC with the proposed similarity measurements can detect clusters with similar spatial locations, spatial-temporal attribute values, and spatial-temporal attribute trends with the highest accuracy.
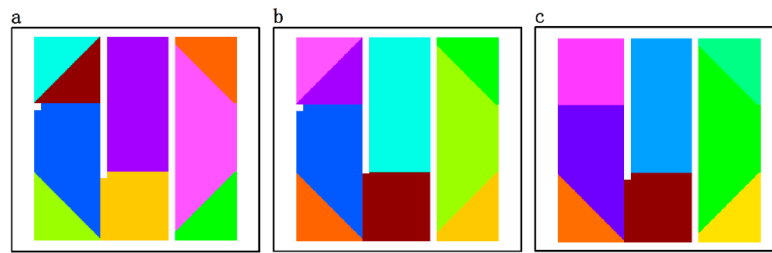
Figure 8. Clustering results of DBTSC on *SD* with different similarity measurements: (**a**) Result of DBTSC with the proposed similarity measurements; (**b**) Result of DBTSC with the mean attribute difference as the similarity measurement; (**c**) Result of DBTSC with the Spearman coefficient as the similarity measurement.

6.3. Validation of the DBTSC-IR Method

Several simulated datasets are designed to evaluate the performance of the DBTSC-IR. The datasets holding eleven images are shown in Figure 9(a1–d1), and the clustering results by using the DBTSC algorithm are shown in Figure 9(a2_1–d2_1). The results of DBTSC-IR are shown in Figure 9(a2_2–d2_2). The accuracy values of the results in Figure 9(a2_3–d2_3) show that the DBTSC algorithm can accurately recognize the predefined clusters with high accuracy. Furthermore, the accuracy values of the results indicate that the DBTSC-IR can still accurately obtain clusters when the computation cost is significantly reduced. In conclusion, the DBTSC-IR is useful in recognizing clusters with high effectiveness and high accuracy.
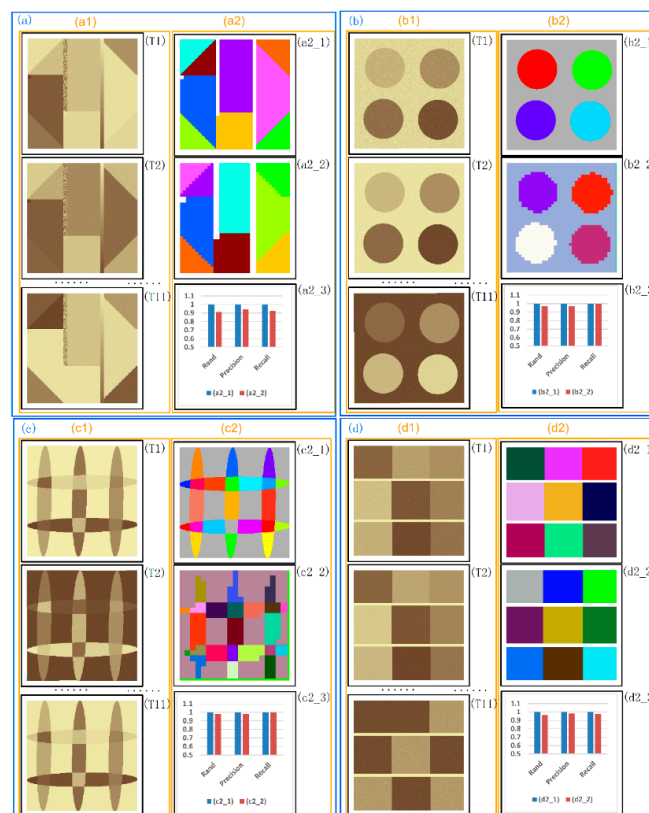

Figure 9. Results of the time series images resampling and clustering techniques. (**a1–d1**) Simulated dataset with 11 images; (**a2_1–d2_1**) Results of the DBTSC algorithm; (**a2_2–d2_2**) Results of resampling and clustering by the DBTSC algorithm; (**a2_3–d2_3**) Accuracy values of the results.

6.4. Application on Detecting Surface Deformation Patterns

Surface deformation, which is caused by nature and humanity, is an environmental and geological phenomenon worldwide. The uneven deformation over a large area of the earth's surface may potentially damage infrastructure and buildings. Ningbo City is located in the coastal area of East China with a specific geological condition and geographical position. Its surface deformation is ubiquitous and elicits significant attention. Hence, studying the characteristics of the surface deformation in Ningbo City is necessary.

The application aims to analyze the deformation patterns of Ningbo City that may reveal interesting regional deformation characteristics. According to existing deformation research [29–31], the overexploitation of groundwater and urban construction are regarded as the most important factors that affect the surface deformation. However, the exploitation of groundwater in Ningbo City in recent years can be ignored because of schemes that have prohibited such activity since 2005 [32]. Thus, urban construction is the key factor affecting surface deformation in Ningbo City. Recently, studies on urban construction mostly focus on analyzing the influence of construction protocols on surface deformation at building zones [29,33]. Few studies have focused on the pattern recognition of surface deformation over large areas and long periods, which can reveal the deformation characteristics of large regional construction to a certain extent. Furthermore, these deformation characteristics are important for further assessing surface deformation mechanisms, as well as for providing references for supervising and forecasting surface deformation. Hence, the time series clustering methods, which are efficient tools for exploring the distribution features of phenomena, can be utilized to mine the surface deformation patterns. The proposed DBTSC-IR with high efficiency and accuracy is chosen in this section.

Image time series data monitored by the INSAR technique, which have been processed with a high degree of precision following the studies [34,35], are provided by the Ningbo Bureau of Surveying and Mapping in China. The dataset includes 27 deformation images obtained from 17 September 2011 to 7 August 2015. To better analyze the characteristics of surface deformation patterns of Ningbo city, the land cover data in 2014 and several years of remote sensing images provided by the Ningbo Bureau of Surveying are also collected. The analyzed area regarding the surface deformation is shown in Figure 10. Deformation values between two neighboring time points are shown in Figure 11 and the end time point of the time interval is labeled above the image. The resolution of the data is 20 m, and the size of each image is 2800 × 2800.
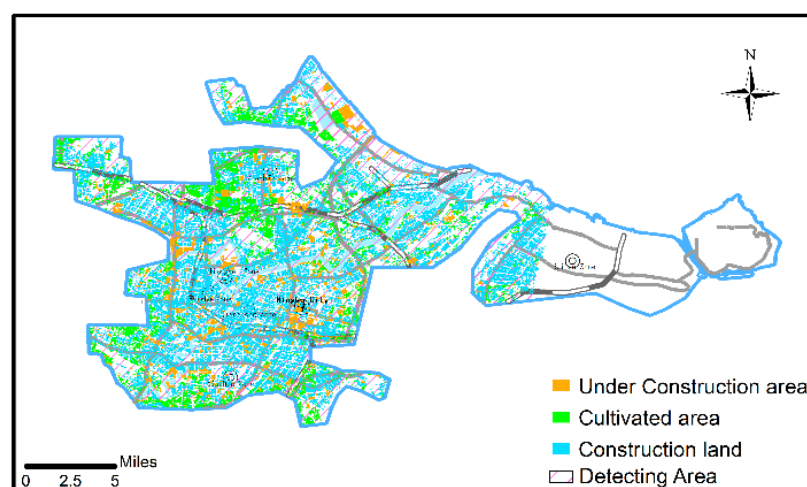

Figure 10. The Detection area of surface deformation in the main urban area of Ningbo city.

To analyze the surface deformation patterns of Ningbo city, two main steps are taken. In the first step, given that the data volume of the surface deformation dataset is large, the dataset is resampled

to enhance the computation efficiency of the clustering process, following the method in Section 3. The results of resampling will be shown in Section 6.4.1. In step 2, the resampled data is clustered to obtain the deformation patterns, which will be analyzed in Section 6.4.2.
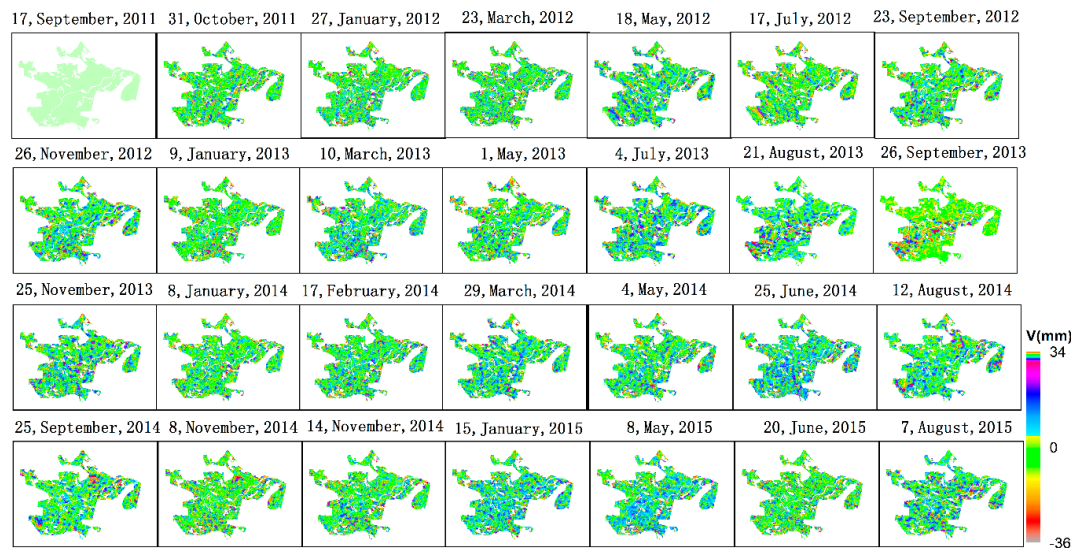

Figure 11. Image time series data of the surface deformation detection in Ningbo City.

6.4.1. Resampling of Surface Deformation Data

To obtain the suitable spatial resampling scale of the image time series data of the surface deformation, the resampling method in Section 3 is used. The analyzed result is shown in Figure 12. The right vertical axis represents the corresponding scale optimization function (Equation (1)) values of several spatial resolutions (horizontal axis) and the left vertical axis is the time cost of the time series clustering procedure. The decrease of the spatial resolution greatly accelerates the computational efficiency, and the optimal resampling scale is 80 m. Hence, the surface deformation images are resampled to 80 m.
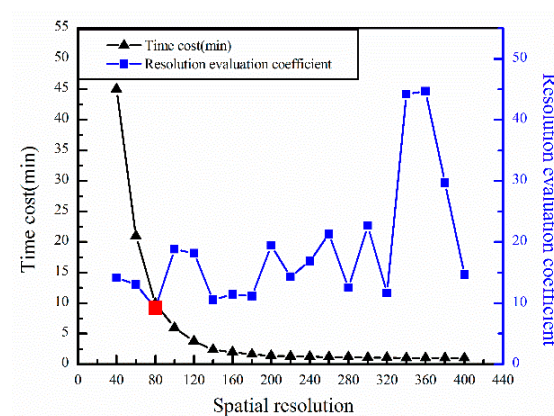

Figure 12. Statistics of resampling of the surface deformation data.

6.4.2. Implementing Pattern Recognition by the DBTSC Algorithm

In order to effectively obtain the spatial deformation characteristics, the DBTSC algorithm is conducted on the resampled images. The result is shown in Figure 13. Several interesting patterns can be found in the clustering result. The result is analyzed in the following paragraph combining

the statistical values of the accumulative deformation of the clusters in Figure 14. The accumulative deformation of a cluster is the mean value of the deformation of the pixels from 17 September 2011 to the time point in the horizontal axis.
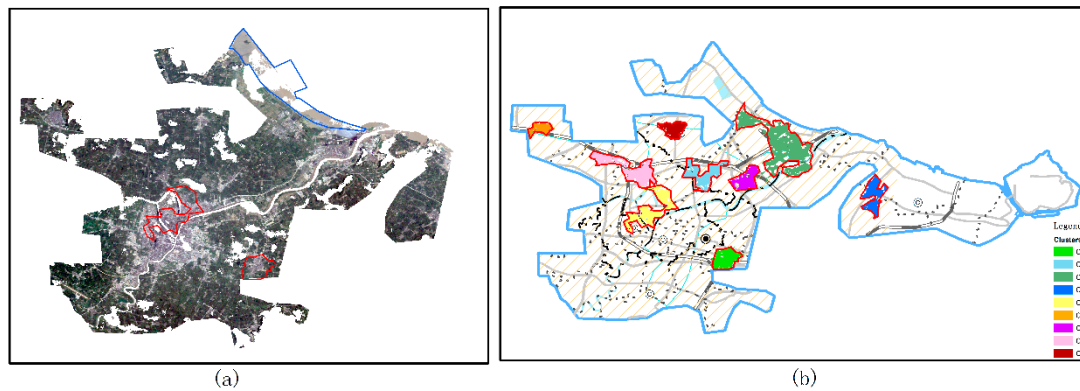

Figure 13. (**a**) Landsat image of 1995 in the deformation detecting area; (**b**) Clustering result of surface deformation using the DBTSC-IR.
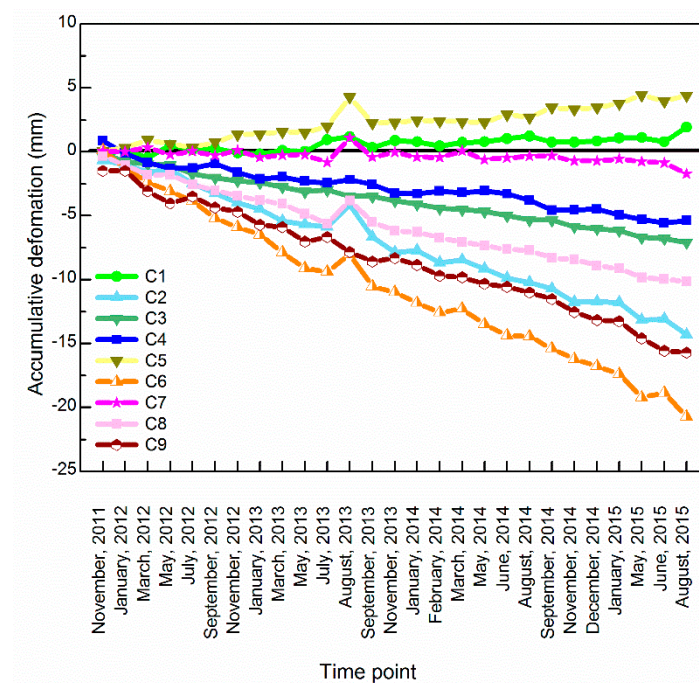

Figure 14. Accumulative deformation of clusters in Figure 13b.

Figure 13 shows that several main clusters are discovered by the DBTSC algorithm. Based on the land cover data in Figure 10, the LANDSAT image of 1995 in Figure 13a, and the accumulative deformation in Figure 14, the clustering result is analyzed as follows: nine interesting clusters were obtained. Clusters 1 and 5 with a positive value of deformation are located in the areas (areas in the red boundaries in Figure 13a) constructed before 1995, which shows that the old building zones may exist in a slightly uplifted fashion in a certain period after more than two centuries. Other clusters detected possess negative values of deformation, which are almost distributed in the construction area without being constructed projects. The abovementioned phenomena show that the construction areas probably experienced on-going ground settlements in the following 20 years. In addition, large areas are recognized as noise under the interference of human activities. For example, the area in the blue

boundary in Figure 13a was a reclamation area that was exploited from 1999 to 2002 and the soft soil results in significantly uneven deformation and is detected as noise. By combining the land cover type in Figure 10, we found that areas under construction in recent years and the cultivated areas are strongly disturbed by human activities and are recognized as noise, thereby meeting real-world conditions. Furthermore, the reason for the occurrence of the different accumulative deformation values of clusters in Figure 14 may be the difference of the years of construction, geological conditions, and the attributes of buildings. Given the lack of the related data, a more detailed analysis will be conducted in the future. In summary, the abovementioned analysis shows that the clustering result of the DBTSC-IR reveals the construction characteristics of the zones, and the result is consistent with the actual situation.

In order to verify the feasibility of the DBTSC-IR, classical time series clustering algorithms such as the k-means time series clustering algorithm, the fuzzy c-means time series clustering algorithm, and the density-based time series clustering algorithm are also performed on the same data for comparison. To enhance the comparison between the classical time series clustering algorithms and the DBTSC-IR, the numbers of clusters are set as nine, i.e., the predefined number of clusters. The results of the classical time series clustering algorithms on the image time series data of the surface deformation are shown in Figure 15 and the variations of clusters in the abovementioned clustering results are shown in Table 1. The Figure and the Table show that the results of the classical time series clustering algorithms are seriously affected by the areas with uneven deformation and the variations of clusters in the results of the classical time series clustering algorithms are significantly larger than those of the DBTSC-IR, thereby indicating that the result of the DBTSC-IR is more reasonable. Furthermore, the clusters of the k-means time series clustering and fuzzy c-means time series clustering algorithms are randomly distributed throughout the spatial domain, thereby preventing the clear visualization and intelligibility of the surface deformation patterns of Ningbo City.
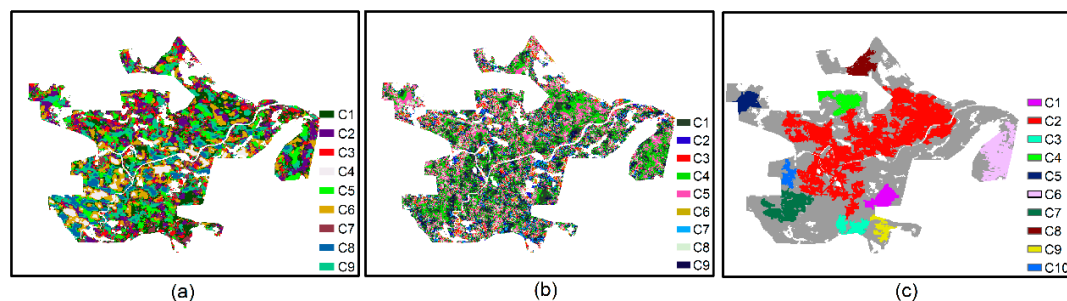

Figure 15. Classical time series clustering results on the surface deformation dataset: (**a**) Result of the k-means time series clustering algorithm; (**b**) Result of the fuzzy c-means time series clustering algorithm; (**c**) Result of the density-based time series clustering algorithm.

Table 1. Statistical information of clusters in Figures 12b and 14.

Standard Deviation	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
DBTSC algorithm	31	22	37	17	15	22	32	33	27	
K-means based algorithm	163	177	198	91	172	178	178	118	185	
Fuzzy c-means based algorithm	184	97	146	193	187	107	135	176	162	
Density-based algorithm	35	121	42	38	34	34	42	37	37	43

According to the analysis above, the following rules can be observed:

(1) The proposed DBTSC-IR algorithm can detect clusters with arbitrary shapes under the interference of uneven deformation areas with higher efficiency and accuracy compared with the classical time series clustering algorithms.

(2)　The results of the DBTSC-IR algorithm can provide a reference for analyzing the patterns of city development. For example, it can separate the old urban district, the newly constructed district, and the zones under construction.

(3)　Most of the constructed areas in 20 years continue to have subsidence.

(4)　Several districts constructed more than two centuries ago are slightly uplifted due to ground rebound.

(5)　The surface deformation in the reclamation area in Ningbo city remains unstable.

7. Conclusions

In this paper, the DBTSC-IR has been proposed to adaptively and effectively detect clusters with similar spatial locations, spatial-temporal attribute values, and spatial-temporal attribute trends. In this method, two major improvements are attained to adaptively detect image time series clusters. First, to reduce the data volume, we propose an adaptive framework in which a new scale optimization function is used to select the most suitable resampling scale, based on which the subsequent analysis efficiency is significantly advanced. Then, the density-based time series clustering algorithm is improved by combining the strategy of the density indicator and the proposed similarity measurements. The improved density-based time series clustering algorithm can adaptively detect non-overlapping clusters with similar spatial locations, spatial-temporal attribute values, and spatial-temporal attribute trends under the interference of noise.

In order to verify the effectiveness and efficiency of the proposed DBTSC-IR algorithm, two comparative experiments on both simulated datasets and practical applications have been conducted, from which the following conclusions can be made. First, the efficient performance of the DBTSC-IR is demonstrated in full by introducing the image resampling method, and the resampling procedure is adaptively conducted by adopting the proposed resampling selection framework with a proposed optimal resampling function; Second, DBTSC-IR can detect non-overlapping clusters with arbitrary shapes, randomly distributed attributes, and noise by combining the strategy of the density indicator and the proposed similarity measurements; Thirdly, the adaptive image time series clustering algorithm DBTSC-IR automatically reveals the distribution characteristics even without sufficient prior knowledge of the dataset. Finally, the DBTSC-IR is theoretically applicable for the time series clustering of geographic data or image datasets. In this paper, the DBTSC-IR was designed for the image dataset. The DBTSC-IR can also be slightly modified for geographic data by constructing a proximity relationship by using Delaunay triangulation or a Voronoi diagram.

Several prospects are available for future studies. For example, the DBTSC-IR is designed only for numerical variables, and it can be further extended to deal with multi-type variables. For another example, the combination of the time series clustering algorithm with association rules to mine the association of clusters and other correlative phenomena should be given more importance in the future.

Author Contributions: Xiaomi Wang and Yaolin Liu conceived and designed the experiments; Xiaomi Wang performed the experiments; All the authors analyzed the data; Xiaomi Wang and Yaolin Liu wrote the paper. All authors contributed to revising the manuscript.

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

References

1.　Guyet, T.; Nicolas, H. Long term analysis of time series of satellite images. *Pattern Recognit. Lett.* **2016**, *70*, 17–23. [CrossRef]

2. Bidari, P.S.; Manshaei, R.; Lohrasebi, T.; Feizi, A.; Malboobi, M.A.; Alirezaie, J. Time series gene expression data clustering and pattern extraction in arabidopsis thaliana phosphatase-encoding genes. In Proceedings of the 2008 8th IEEE International Conference on BioInformatics and BioEngineering (BIBE 2008), Athens, Greece, 8–10 October 2008; pp. 1–6.

3. Kaur, G.; Dhar, J.; Guha, R.K. Minimal variability owa operator combining anfis and fuzzy c-means for forecasting bse index. *Math. Comput. Simul.* **2016**, *122*, 69–80. [CrossRef]

4. Krishna, K.; Narasimha Murty, M. Genetic k-means algorithm. *IEEE Trans. Syst. Man Cybern. B Cybern.* **1999**, *29*, 433–439. [CrossRef] [PubMed]

5. Möller-Levet, C.S.; Klawonn, F.; Cho, K.H.; Wolkenhauer, O. *Fuzzy Clustering of Short Time-Series and Unevenly Distributed Sampling Points*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 330–340.

6. Yin, J.; Zhou, D.; Xie, Q.Q. A clustering algorithm for time series data. In Proceedings of the International Conference on Parallel and Distributed Computing, Applications and Technologies, Taipei, Taiwan, 4–7 December 2006; pp. 119–122.

7. Jiang, D.; Pei, J.; Zhang, A. Dhc: A density-based hierarchical clustering method for time series gene expression data. In Proceedings of the Third IEEE Symposium on Bioinformatics and Bioengineering, Bethesda, MD, USA, 10–12 March 2003; pp. 393–400.

8. Chis, M.; Grosan, C. Evolutionary hierarchical time series clustering. In Proceedings of the International Conference on Intelligent Systems Design and Applications, Porto, Portugal, 16–18 October 2006; pp. 451–455.

9. Rodrigues, P.P.; Gama, J.; Pedroso, J. Hierarchical clustering of time-series data streams. *IEEE Trans. Knowl. Data Eng.* **2008**, *20*, 615–627. [CrossRef]

10. Uijlings, J.R.R.; Duta, I.C.; Rostamzadeh, N.; Sebe, N. Realtime video classification using dense hof/hog. In Proceedings of the International Conference on Multimedia Retrieval, Glasgow, UK, 1–4 April 2014; pp. 145–152.

11. Chandrakala, S.; Sekhar, C.C. A density based method for multivariate time series clustering in kernel feature space. In Proceedings of the IEEE International Joint Conference on Neural Networks and 2008 IEEE World Congress on Computational Intelligence, Hongkong, China, 1–8 June 2008; pp. 1885–1890.

12. Mörchen, F.; Ultsch, A.; Hoos, O. Extracting interpretable muscle activation patterns with time series knowledge mining. *Int. J. Knowl. Based Intell. Eng. Syst.* **2005**, *9*, 2006. [CrossRef]

13. Zanotto, C.; Giangaspero, M.; Büttner, M.; Braun, A.; Morghen, C.G.; Elli, V.; Panuccio, A.; Radaelli, A. Evaluation of poliovirus vaccines for pestivirus contamination: Non-specific amplification of poliovirus sequences by pan-pestivirus primers. *J. Virol. Methods* **2002**, *102*, 167–172. [CrossRef]

14. Xu, T.; Shang, X.; Yang, M.; Wang, M. Bicluster algorithm on discrete time-series gene expression data. *Appl. Res. Comput.* **2013**, *30*, 3552–3557.

15. Yan, L.; Kong, Z.; Wu, Y.; Zhang, B. Biclustering nonl inearly correlated time series gene expression data. *J. Comput. Res. Dev.* **2008**, *45*, 1865–1873.

16. Warren Liao, T. Clustering of time series data—A survey. *Pattern Recognit.* **2005**, *38*, 1857–1874. [CrossRef]

17. Lhermitte, S.; Verbesselt, J.; Verstraeten, W.W.; Coppin, P. A comparison of time series similarity measures for classification and change detection of ecosystem dynamics. *Remote Sens. Environ.* **2011**, *115*, 3129–3152. [CrossRef]

18. Lottering, R.; Mutanga, O. Optimising the spatial resolution of worldview-2 pan-sharpened imagery for predicting levels of gonipterus scutellatus defoliation in kwazulu-natal, south africa. *ISPRS J. Photogramm. Remote Sens.* **2016**, *112*, 13–22. [CrossRef]

19. Orlhac, F.; Soussan, M.; Chouahnia, K.; Martinod, E.; Buvat, I. 18F-FDG pet-derived textural indices reflect tissue-specific uptake pattern in non-small cell lung cancer. *PLoS ONE* **2015**, *10*, e0145063. [CrossRef] [PubMed]

20. Liu, Q.; Deng, M.; Shi, Y.; Wang, J. A density-based spatial clustering algorithm considering both spatial proximity and attribute similarity. *Comput. Geosci.* **2012**, *46*, 296–309. [CrossRef]

21. Liu, Y.; Wang, X.; Liu, D.; Liu, L. An adaptive dual clustering algorithm based on hierarchical structure: A case study of settlement zoning. *Trans. GIS* **2016**. [CrossRef]

22. Fu, T.-C. A review on time series data mining. *Eng. Appl. Artif. Intell.* **2011**, *24*, 164–181. [CrossRef]

23. Chao, X. A review on correlation coefficients. *J. Guangdong Univ. Technol.* **2012**, *29*, 12–17.

24. Zhang, W. Measuring mixing patterns in complex networks by spearman rank correlation coefficient. *Phys. A Stat. Mech. Appl.* **2016**, *451*, 440–450. [CrossRef]

25. Ramirez-Lopez, L.; Schmidt, K.; Behrens, T.; van Wesemael, B.; Demattê, J.A.M.; Scholten, T. Sampling optimal calibration sets in soil infrared spectroscopy. *Geoderma* **2014**, *226–227*, 140–150. [CrossRef]

26. Heng, X.; Junjie, L.; Guo, J.; Qin, Z.; Shao, L. Approximate query algorithm based on eight-neighbor grid clustering for heterogeneous xml documents. *J. Xi'an Jiaotong Univ.* **2007**, *41*, 907–911.

27. Manning, C.D.; Raghavan, P.; Schütze, H. *Introduction to Information Retrieval*; Cambridge University Press: Cambridge, MA, USA, 2008; pp. 824–825.

28. Grubesic, T.H.; Wei, R.; Murray, A.T. Spatial clustering overview and comparison: Accuracy, sensitivity, and computational expense. *Ann. Assoc. Am. Geogr.* **2014**, *104*, 1134–1156. [CrossRef]

29. Katebi, H.; Rezaei, A.H.; Hajialilue-Bonab, M.; Tarifard, A. Assessment the influence of ground stratification, tunnel and surface buildings specifications on shield tunnel lining loads (by FEM). *Tunn. Undergr. Space Technol.* **2015**, *49*, 67–78. [CrossRef]

30. Luo, C.-Y.; Shen, S.-L.; Han, J.; Ye, G.-L.; Horpibulsuk, S. Hydrogeochemical environment of aquifer groundwater in shanghai and potential hazards to underground infrastructures. *Nat. Hazards* **2015**, *78*, 753–774. [CrossRef]

31. Toivanen, T.L.; Leveinen, J. *Groundwater Level Variation and Deformation in Clays Characteristic to the Helsinki Metropolitan Area*; Springer International Publishing: Basel, Switzerland, 2015; pp. 309–312.

32. Fu, Y. A predictive analysis of groundwater regime and land subsidence in ningbo city. *Resour. Surv. Environ.* **2014**, *35*, 142–146.

33. Chen, B.; Gong, H.; Li, X.; Lei, K.; Ke, Y.; Duan, G.; Zhou, C. Spatial correlation between land subsidence and urbanization in beijing, china. *Nat. Hazards* **2014**, *75*, 2637–2652. [CrossRef]

34. Liao, M.; Wei, L.; Timo, B.; Zhang, L. Application of tomosar in urban deformation surveillance. *Shanghai Land Resour.* **2013**, *34*, 7–11.

35. Liao, M.; Lin, H. *Synthetic Aperture Radar Interferometry: Principle and Signal Processing*; Surveying and Mapping Press: Beijing, China, 2003.