

Article

A Spatio-Temporal Local Association Query Algorithm for Multi-Source Remote Sensing Big Data

Lilu Zhu ^{1,*}, Xiaolu Su ², Yanfeng Hu ^{2,3}, Xianqing Tai ⁴ and Kun Fu ⁴

¹ School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China

² Institute of Electronics, Chinese Academy of Sciences, Suzhou 215123, China; suxl300371@aircas.ac.cn (X.S.); huyf@aircas.ac.cn (Y.H.)

³ Key Laboratory of Intelligent Aerospace Big Data Application Technology, Suzhou 215123, China

⁴ Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; taixq@aircas.ac.cn (X.T.); fukun@aircas.ac.cn (K.F.)

* Correspondence: zhull@aircas.ac.cn

Abstract: It is extremely important to extract valuable information and achieve efficient integration of remote sensing data. The multi-source and heterogeneous nature of remote sensing data leads to the increasing complexity of these relationships, and means that the processing mode based on data ontology cannot meet requirements any more. On the other hand, the multi-dimensional features of remote sensing data bring more difficulties in data query and analysis, especially for datasets with a lot of noise. Therefore, data quality has become the bottleneck of data value discovery, and a single batch query is not enough to support the optimal combination of global data resources. In this paper, we propose a spatio-temporal local association query algorithm for remote sensing data (STLAQ). Firstly, we design a spatio-temporal data model and a bottom-up spatio-temporal correlation network. Then, we use the method of partition-based clustering and the method of spectral clustering to measure the correlation between spatio-temporal correlation networks. Finally, we construct a spatio-temporal index to provide joint query capabilities. We carry out local association query efficiency experiments to verify the feasibility of STLAQ on multi-scale datasets. The results show that the STLAQ weakens the barriers between remote sensing data, and improves their application value effectively.

Keywords: multi-source remote sensing big data; self-correlation network; cross-correlation network; multi-dimensional index



Citation: Zhu, L.; Su, X.; Hu, Y.; Tai, X.; Fu, K. A Spatio-Temporal Local Association Query Algorithm for Multi-Source Remote Sensing Big Data. *Remote Sens.* **2021**, *13*, 2333. <https://doi.org/10.3390/rs13122333>

Academic Editors: Lalit Kumar and Onesimo Mutanga

Received: 28 April 2021

Accepted: 9 June 2021

Published: 14 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Research Background

With the continuous development of Earth Observation Network technology, various types of and huge amounts of remote sensing data are generated by a large number of sensors in real time. The magnitude of remote sensing data has increased from GBs to TBs and PBs, and it will continue to increase in the future [1]. These remote sensing data are collections of geographic information related to the location, based on a unified space-time reference [1–3]. Remote sensing big data is the fusion of big data and remote sensing data, and is active in time and space. It is big data based on a unified space-time reference, which takes the Earth as the object [1–3]. Remote sensing big data mainly includes space-time reference data, geodetic survey data, gravity and magnetic data, remote sensing image data, and location-related spatial media data. In addition, it has been widely used in many fields such as national defense, agriculture, water conservancy, land planning, smart cities, disaster warning, geological surveys, emergency monitoring and so on [3–5]. Although these massive remote sensing data come from different sources and have different structures, they are often potentially related to each other due to their own spatio-temporal

characteristics, data characteristics and others. In addition, they always have direct or indirect mapping relationships with geographical entities existing in the real world. Therefore, in the purpose of fully mining the application value of multi-source remote sensing big data and providing comprehensive and diversified information support, we should pay more attention to building a unified organization and management for multi-source remote sensing big data to realize data fusion.

Remote sensing data has many characteristics, such as large volume, multi-source heterogeneity, complex relationships, a wide distribution, and being multi-scale, multi-temporal, and multi-topic [6]. Data generated by various industries and departments are usually difficult to organize in a unified manner. Thus, it is easy to cause isolated islands of data [7,8]. The discrepancies in the organizational structure of remote sensing data from various sources and categories may make data fusion and correlation analysis challenging. It is difficult to guarantee the integrity of the search results based on the attributes and semantics of the data, that is, the search results only contain part of the data directly related to the query conditions. The lack of correlation not only brings about the high complexity of data analysis, but also means that the cost of data application continues to accumulate. Therefore, there are still many challenges in remote sensing big data retrieval and management, and it is difficult to provide a full-time and comprehensive remote sensing data support.

1.2. Contributions

The main contributions of our study are as follows:

1. In view of the general problems in remote sensing big data management and retrieval, we design a general and efficient spatio-temporal local association query algorithm for remote sensing data, the STLAQ, on the basis of mining and constructing the relationships of remote sensing data from different perspectives.
2. The STLAQ solves the association problem of high-dimensional remote sensing data, relying on a data model and correlation network, especially when global features degenerate into local features on a unified multi-scale organization model.
3. The STLAQ weakens the barriers between remote sensing data and expands their sharing mode, and has strong versatility and practical value in the field of remote sensing data application.

1.3. Paper Organization

In Section 1, we introduce the main problems faced by remote sensing big data in applications. In Section 2, we introduce the research situation of remote sensing data index and remote sensing data correlation. Then, in Section 3, we present the main structure and concepts of the STLAQ, introducing the STLAQ from two aspects; multi-source remote sensing data association and local association query based on correlation network. In Section 4, we construct the STCN on several typical remote sensing datasets and carry out association query experiments to prove the usability and universality of the STLAQ. Finally, we conclude this study and look forward to the follow-up work in Section 5.

2. Related Work

With the rapid growth of the volume and the increasingly rich sources and types, remote sensing data continues to be applied in various fields. In the applications of remote sensing big data, there are challenges in storage, management, analysis and visualization. It is extremely important to prove an efficient query algorithm for massive multi-source remote sensing data. Traditional remote sensing data query algorithms only provide retrieval schemes for data coming from a single-source, lacking a unified retrieval mechanism for multi-source and multi-type remote sensing data. Designing a remote sensing data index with high-performance and correlation method for multi-source remote sensing

data, to solve the barriers caused by the heterogeneity of multi-source remote sensing data, is one of the focuses of the research field of remote sensing big data.

Research on spatio-temporal index is usually for the improvement and expansion of traditional spatial indexes in dimensions, such as time dimension. The earliest research can be traced back to the 1970s [9–11]. The Quadtree index is proposed in [12], dividing nodes into four regions in a two-dimensional space to solve the problem of combined query of data. In [13], a multi-dimensional binary tree (KD tree) in which each node is a k-dimensional point is introduced, and it could be used to store k-dimensional data; R tree is designed in [14], a hierarchical data structure based on B+ trees, used for the dynamic organization of a set of d-dimensional geometric objects. SDMR tree, a deformed R-tree index structure for multi-scale expression of spatial data was proposed in [15]. In [16], the Hilbert R tree is proposed, which uses Hilbert curve, which has better spatial proximity than other space filling curves [17], in order to perform one-dimensional sorting on k-dimensional space data to improve node storage utilization. In [18], a distributed remote sensing database with indexing techniques, GeoMesa, is presented. It uses space filling curves to map multi-dimensional data to the single lexicographic list managed by the underlying distributed database. In [19], the authors propose a design and implementation of a distributed spatial query index and query algorithm consisting of a hybrid index composed of a Quadtree, R-Tree and Hash structure. In [20], a unified index framework is proposed, which provided efficient data distribution, fault tolerance and multi-dimensional data query processing, and proposed an algorithm for range query and K-nearest neighbor query. In [21], a spatial index is designed, linked with LSM B-tree and LH R-tree, which effectively improved the operation efficiency of the entire index for update and retrieval. In [22], the authors combine grid coding, dimensional attributes of interest and HBase row keys to design a customizable multi-dimensional index structure, ST+. In [23], a remote sensing data indexing method based on Hilbert code is presented, solving the problem of fast access to remote sensing data in the virtual battlefield environment.

Data association is to mine the correlation and dependence between data, and to obtain the internal connection between multi-source geographic data [24]. Geographical data association can realize the overall holographic relationship expression of natural and human geographic elements in the region [25]. Chen Luo's team has carried out a series of studies. In [26], the authors propose a generalized spatial data correlation model for spatial data, place names and general entities, and verifies the effectiveness of the model. In [27], a multi-information correlation database is established for several types of aerospace information, such as remote sensing image information, radiation source information, geographic information, etc. In [28], the authors introduce graph theories and techniques into entity relationship research, and propose a graph-based correlation analysis technique. In [29], a data association model and a geocoding method based on space-time unity are proposed, to realize the association organization and analysis of multi-source heterogeneous data based on a unified space-time framework. However, these studies also have shortcomings. GeoStar proposes a solution of a remote sensing data platform, which mainly focuses on the connection between geographic elements and social thematic data, and establishes a unified data backplane. In these studies, only spatial, visual, and semantic data correlations are considered, with a lack of dynamic relationships in such time and events dimensions. Moreover, these correlation models are all performed at a single scale, ignoring correlation across different scales.

All of the above studies have launched the research on remote sensing data association methods from different angles. In any case, these studies only pay attention to the design and optimization of the model structure, ignoring the influence of the remote sensing data's own features on the quality of the association query results. There may be the disadvantage of incomplete association query results due to the lack of data features at a single scale. Therefore, there are still gaps to be filled in the multi-scale association query combining data features.

3. Materials and Methods

3.1. Local Association Query Algorithm for Remote Sensing Data

In order to realize the unified organization and management of multi-source remote sensing data, eliminate the structural differences between different types of remote sensing data from the application point of view, and analyze the internal connections in the multi-dimensional feature space of remote sensing data and establish correlations, this paper proposes a local association algorithm for querying remote sensing data: STLAQ. Under the unified space-time reference, we carry out feature extraction and digital modeling on remote sensing data, and then establish the relationships of remote sensing data in multi-dimensional feature space across different scales, to provide a cross-scale association retrieval capability. The algorithm mainly includes three components, the spatio-temporal data model (STDM), the spatio-temporal correlation network (STCN) and the spatio-temporal association index, as shown in Figure 1.

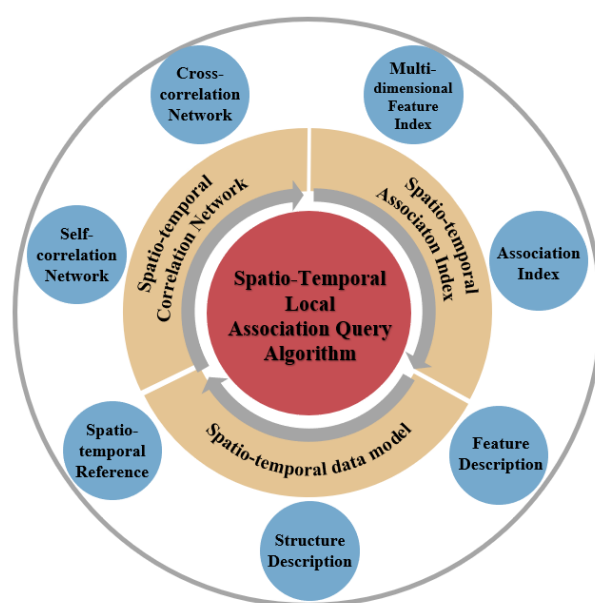


Figure 1. Local Association Query Algorithm for Remote Sensing Data.

The STDM is a digital model established to facilitate the description and analysis of remote sensing data. It consists of three basic elements: space-time reference (STR), structure description (SD) and feature description (FD). On the basis of the STDM encapsulation, we propose the spatio-temporal correlation network (STCN). It is a modeled description of spatiotemporal data and its association relationship based on the undirected hypergraph model. It includes two parts, spatio-temporal self-correlation network (STSCN) and spatio-temporal cross-associated network (STCCN). The spatio-temporal association index is an index defined according to the structure of the spatio-temporal correlation network, including two parts: a multi-dimensional feature index and a correlation index.

3.1.1. Spatio-Temporal Data Model

The spatio-temporal data model is a digital model proposed to facilitate data description and analyzation, as shown in Figure 2. The spatio-temporal data model can be formally expressed as $STM(STR, SD, FD)$. Based on the spatio-temporal data model, remote sensing data can be encapsulated into two different granularities, remote sensing data unit (STDU) and remote sensing data object (STDO). Remote sensing data can be divided into one or more STDOs, according to definite classification standards. An STDU is the smallest unit in remote sensing data processing, and is an element in an STDO. An STDO can be expressed as a collection of STDUs organized in a hierarchical model.

- **Space-time Reference.** Space-time reference is the basis for describing and measuring remote sensing data, including two parts; time reference and space reference. The origin and measurement scale of the datum need to be clarified in each reference. The space reference system includes several parts, such as plane coordinate system, map projection, and height datum, used in representing the spatial feature of remote sensing data and describing the spatial relationship between them. The time reference system is the basis for time information exchange and calculation. Remote sensing data can be expressed as a moment or a period in the time reference system. In [30], the Gregorian calendar, 24:00, local time or coordinated universal time are designated as the basic time reference system for information exchange.
- **Structure Description.** The structure description of remote sensing data is a description of the underlying remote sensing data organization scheme, including multi-scale model, storage structure and index structure. The multi-scale model is a multi-level-of-detail model, established by way of breaking up different types of data according to uniform rules and dividing the data into different resolution levels. In the multi-scale model, spatio-temporal data units at different resolutions can be correlated based on features, and be stored and indexed uniformly. The storage structure defines the storage form of remote sensing data, to organize data and its characteristics in a certain structure. In order to solve the efficiency problem of massive remote sensing data in joint retrieval, it is necessary to build indexes on the basis of storage structure. The index structure describes the composition of the spatio-temporal index, including retrieval conditions, retrieval values, and keys of the remote sensing data in storage structure.
- **Feature Description.** Feature description is the low-level features exacted from remote sensing data. The relationship between remote sensing data is essentially to calculate the similarity between the various features and then to aggregate the results. Therefore, the extraction of data features and the calculation of the correlation degree are critical parts of the construction of data correlation. This paper divides the features of remote sensing data into three categories: time feature, spatial feature, and data feature. In this case, time feature can be expressed as the production time or the life cycle of remote sensing data; spatial features include two parts, spatial position (such as latitude and longitude) and spatial form (such as points, lines, surfaces). Spatial and time features of the data can be obtained from the data itself, as well as through data analysis. Data features refer to the low-level features extracted from remote sensing data. For example, color features, texture features and shape features of remote sensing image data can be extracted, such as gray-level mean, gray-level standard deviation, direction gradient, principal component features, etc. Different remote sensing data often differ in the types and values of data features.

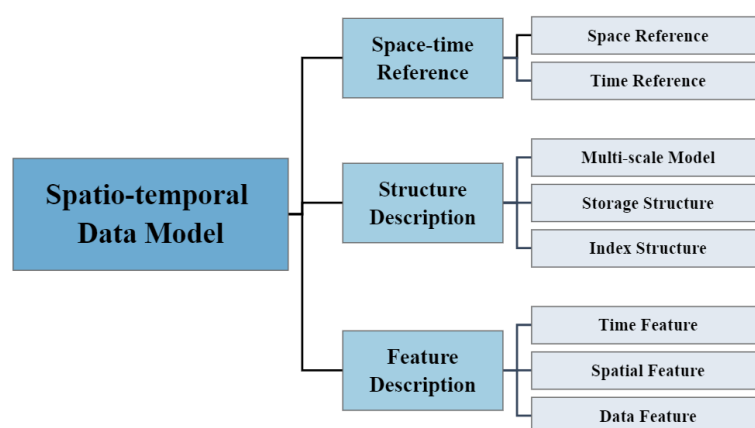


Figure 2. Spatio-temporal data model.

An example of encapsulating digital elevation model data into a spatio-temporal data model is given in Table 1. The detailed explanations are as follows.

1. In the part of space-time reference, the spatial reference is WGS84/EGM96, and the time reference is UTC/GMT+08:00 zone.
2. In the part of structure description, the multi-scale model of the sample data is a tile pyramid model. The key of the storage table is composed of data location, data type, and data acquisition time. There are 2 column families in the storage table, where data is used to store tile data, meta is used to store the meta information. The index structure includes a multi-dimensional feature index and an association index.
3. In the part of feature description, the time feature is the data acquisition time expressed in the format of “yyyy-mm-dd hh:mm:ss”. The spatial feature is the geographic region of the sample data, and the data feature is the low-level features, such as the maximum, minimum, mean, standard deviation and other statistical features.

Table 1. An example of encapsulating elevation data into a spatiotemporal data model.

Spatio-temporal Data Model	
Space-time Reference	
Space reference	WGS84/EGM96
Time reference	UTC+8:00
Structure Description	
Multi-scale Model	tile pyramid model
Storage Structure	key: hilbertcode_type_time data column: tile storage; meta column: property storage, feature storage
Index Structure	multi-dimensional feature index association index
Feature Description	
Time Feature	2011:03:15 19:00:04
Spatial Feature	[22.9998611, 23.9998611, 24.0001389, 25.0001389]
Data Feature	Minimum = 229.000, Maximum = 656.000, Mean = 412.612, StdDev = 32.007, ...

3.1.2. Spatio-Temporal Correlation Network

There is a certain degree of correlation between different remote sensing data, which can be described with the spatio-temporal correlation network (STCN), a model based on the undirected hypergraph model. In the STCN, vertices are used to represent spatio-temporal data objects, hyper-edges are used to represent complex relationships among them, and the weights of hyper-edges are used to define the strength of the correlation between spatio-temporal data objects. The strength of edges between spatio-temporal data objects can be expressed by the correlation degree of remote sensing data, which we will describe in Section 3.2.2.

There are two kinds of remote sensing data correlation network; spatio-temporal self-correlation network and spatio-temporal cross-correlation network. In this case, the STSCN describes the correlation among STDUs within an STDO, and can be regarded as an indivisible hyperedge in the STCN, and the STCCN describes the correlation between different STDOs, that is, the aggregation of multiple STSCN. The structure of STCN is shown in Figure 3. A remote sensing data object may belong to multiple networks, and the variables marked on the connection line represent the degree of correlation between STDOs and STCNs. For example, a lake group with a high degree of spatio-temporal correlation can be represented by constructing a STCCN. The lakes in the lake group can be represented by the STSCNs corresponding to them in the STCCN.

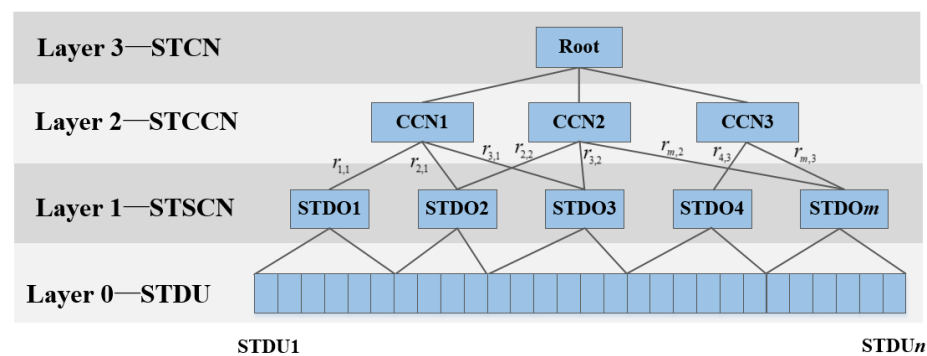


Figure 3. Spatio-temporal correlation network hierarchy.

3.1.3. Preliminaries

- Space Pyramid [31,32]. The spatial pyramid technology is that where the pyramid is a multi-resolution hierarchical model. The essence of its construction process is to block and layer the image to form multiple resolution levels.
- Hypergraph Model [33]. The graph model is a typical model that can effectively represent entities and their relationships. A graph can be expressed as $G = (V, E)$, its vertex set is $V(G) = \{v_1, v_2, \dots, v_n\}$, and the edge set is $E = \{e_1, e_2, \dots, e_m\}$. In the graph, each edge is a point pair (v, w) , where $v, w \in V$. There are two kinds of graphs: directed graph and undirected graph [33]. If the (v, w) is ordered, G is a directed graph, otherwise G is an undirected graph. Hypergraph is a generalization of ordinary graphs and can be used to express complex relationships. For a hypergraph $H = (V, E)$, the set of hyperedges is $E = \{e_1, e_2, \dots, e_m\}$, and each hyperedge e is a subset of the set of vertices V , which can connect two or more vertices and meet the regulation $e_1 \cup e_2 \cup \dots \cup e_m = V$. A weighted hypergraph is to assign a weight as $w(e)$ to each hyperedge, indicating the degree to which the vertices within belong to the hyperedge e . Adjacency matrix and incidence matrix are always used to describe structure of graph G . The adjacency matrix of graph G is a matrix of size $n \times n$, denoted as $A(G)$. If there is an edge between the vertex v_i and v_j , the value of $a_{i,j}$ is 1, otherwise is 0. The incidence matrix of graph G is a matrix of size $n \times m$, denoted as $R(G)$. If the vertex v_i is an endpoint of the edge e_j , the value of $r_{i,j}$ is 1, otherwise it is 0.
- Spectral Cluster [34]. Spectral clustering turns the clustering problem into the graph partitioning problem. To solve the graph cut objective function, the properties of the Rayleigh quotient are usually utilized to map the original data points into a lower dimensional eigen-space by calculating the eigenvectors of Laplacian matrix and then conducting the clustering in the new space [34]. In spectral clustering, all data can be treated as vertices connected by edges. The edge weight value between two points farther apart is lower, and the edge weight value between two points closer together is higher. The purposes of clustering are to make the weight of the edges between the different subgraphs as low as possible, and to make the weight of the edges in the subgraphs as high as possible. The method of spectral clustering used in this paper is normalized cut. According to the implementation of spectral clustering given in [35,36], the application of spectral clustering in STCN construction can be found in Section 3.3.2.

3.2. Multi-Source Remote Sensing Data Association

As shown in Figure 4, there are two main parts in the process of multi-source remote sensing data association, spatio-temporal data model encapsulation and spatio-temporal correlation network construction, which consists of spatio-temporal self-correlation network construction and spatio-temporal cross-correlation network construction. A step-by-step feedback mechanism based on association query results and manual marking is

used to adjust the values of parameters in the whole process. Then, we will introduce them in detail.

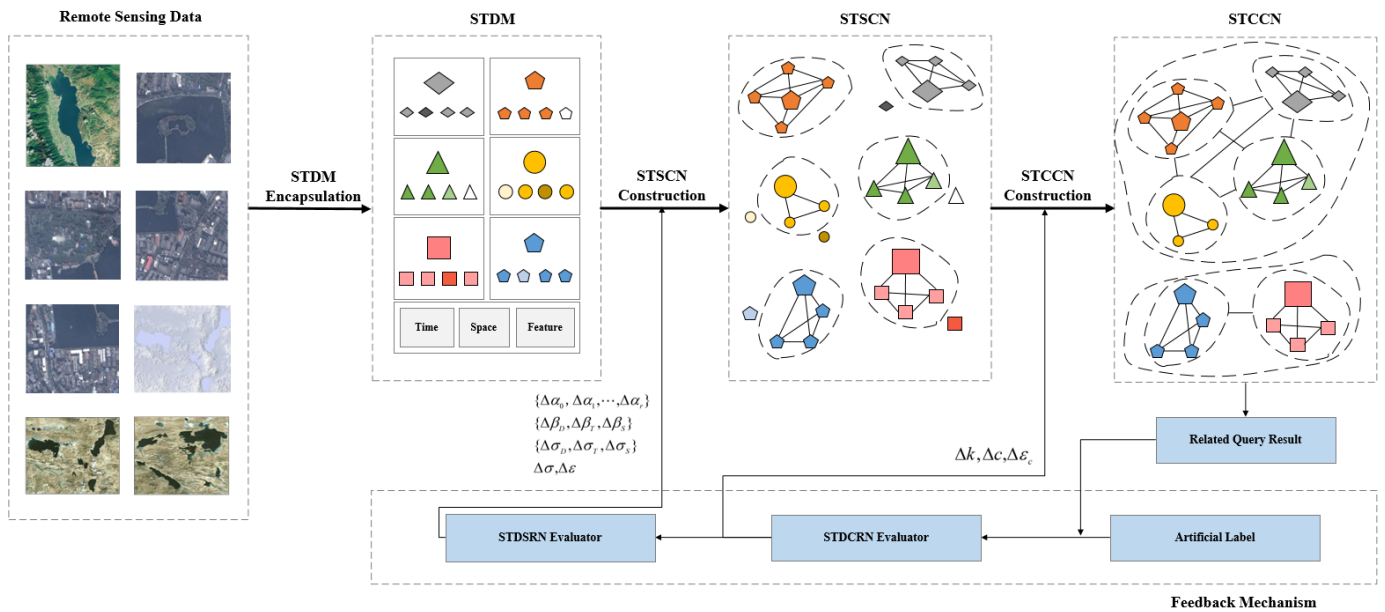


Figure 4. Flow chart of multi-source remote sensing data association method.

3.2.1. Spatio-Temporal Data Model Encapsulation

To complete the encapsulation of the spatio-temporal data model, data analysis, feature extraction, and multi-scale division suitable for remote sensing data of various types and formats are required.

In order to perform remote sensing data feature extraction, firstly, we extract color features, shape features and other visual features, such as grayscale histogram, grayscale mean, directional gradient histogram, and image principal component features [37–40], to establish the data feature description. Secondly, we read the time stamp information from the metadata of the data, to establish the time feature description. Finally, we read the coordinate system information, origin coordinates and resolutions to calculate the geographic extent of the data and to establish a spatial feature description.

As for remote sensing image data, we can establish a remote sensing data grid model based on the spatial pyramid [41,42] for multi-scale division, which is the basis for later associating feature on spatio-temporal data units. Then, we split remote sensing data into data units, build storage structure and index structure on them, and perform local correlation of remote sensing data features on data units of different scales.

As shown in Figure 5, we take lake remote sensing image data as an example, and to explain, the construction of a remote sensing data grid, based on spatial pyramid, is performed.

1. Associate spatio-temporal information on the basis of the pyramidal grid, and establish the correspondence between remote sensing data and grid models according to the spatial resolution and geographic scope of the remote sensing data.
2. In the global remote sensing data grid, the size of the grid cell is fixed (such as 256×256), and the spatial resolution of grid cells at different levels can be obtained via

$$gridRes = \frac{lon_{max} - lon_{min}}{2^{level} \times gridSize} \quad (1)$$

where lon_{max} and lon_{min} represent the maximum and minimum value of the longitude range of the current coordinate system, respectively, $level$ represents the level of current grid, $gridRes$ represents the resolution, and $gridSize$ represents the size of current grid.

3. Copy the remote sensing data to the relevant position in the grid based on the correspondence to obtain a grid generated from the remote sensing data.
4. Use resampling algorithm (such as bilinear interpolation algorithm) to down-sample remote sensing data step-by-step, to obtain grid data at lower resolution levels, obtaining a complete remote sensing data pyramid.
5. Multi-source remote sensing data is sampled and processed on a unified remote sensing data pyramid hierarchy to obtain global spatio-temporal data units with multiple levels of detail.

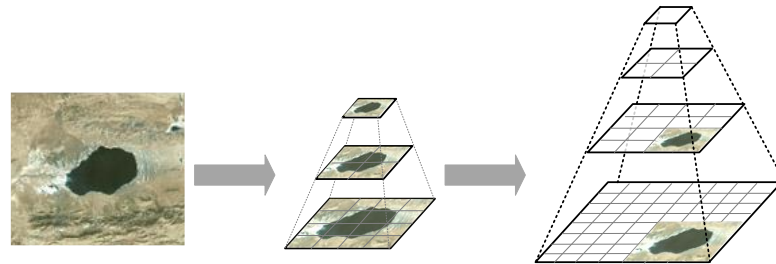


Figure 5. The construction of remote sensing data grid model.

3.2.2. Spatio-Temporal Correlation Network Construction

According to the hierarchical model of the spatio-temporal correlation network shown in Figure 3, the construction of network consists of two steps: the construction of spatio-temporal self-correlation network and of cross-correlation network. In STCN, an STDO is usually represented as a set of STDUs divided at different scales. Each unit in the collection inherits some of the features of the remote sensing data and may be related to other datasets. In order to obtain the relationships among STDOs, it is necessary to aggregate the relationships among STDUs within the objects. Moreover, on the basis of calculating the similarity of the local features of the STDU, STCN can change the conventional single method of establishing remote sensing data correlation based on the same scale, and establish cross-scale correlation.

The relationships among three STDOs, U_I , U_J and U_K , are shown in Figure 6. In Figure 6, spherical models of different sizes are used to represent STDUs at different scales. The correlation between spatio-temporal data units, U_{I1} and U_{J1} , can be calculated by Equation (6), which is an aggregation result of spatial similarity, temporal similarity, and data similarity.

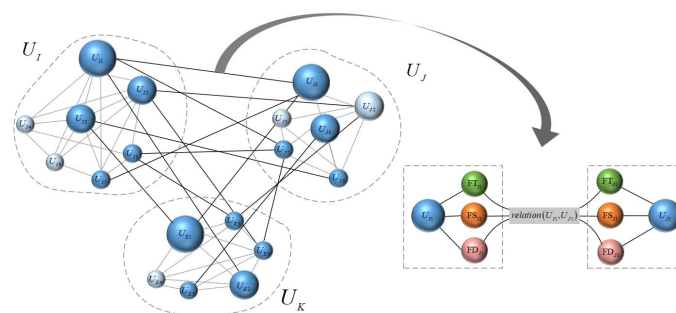


Figure 6. The correlation among STDOs.

We use the adjacency tensor to represent the correlation between different STDOs in the multi-dimensional feature space, as shown in Figure 7. In Figure 7a, the adjacency tensor of the remote sensing data cross-correlation network simplified in matrix form is given. The structure of the adjacency tensor between the spatio-temporal data objects U_I and U_J is given in Figure 7b, identified with the red rectangles. In Figure 7c, the mapping relationship between the similarity of the remote sensing data unit in time, spatial and data feature, and the data relevance is given in β_T , β_S and β_D , respectively, representing

the contribution weight of time feature, spatial feature and data feature in the calculation of relevance, using different colors to indicate the difference in the correlation degree of spatio-temporal data units under different feature similarities.

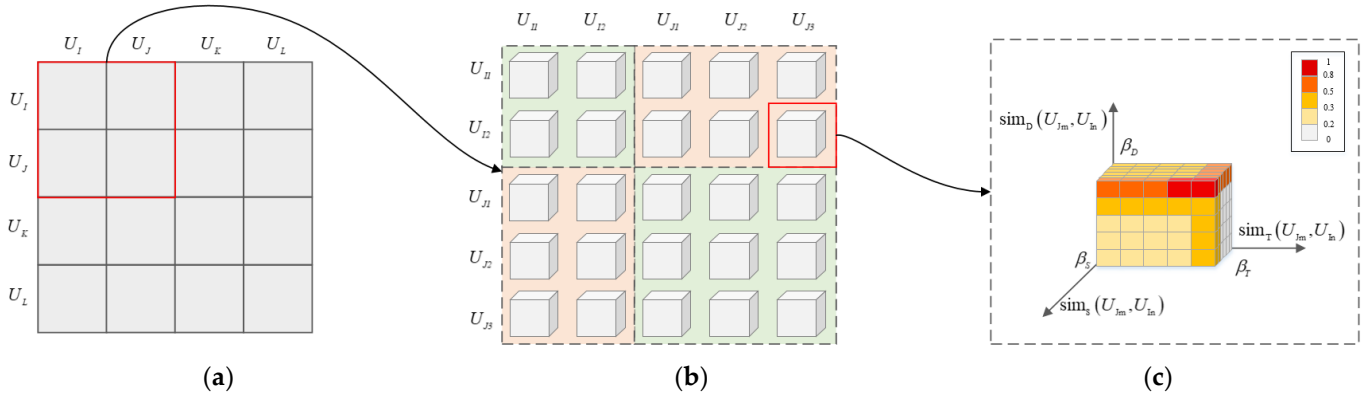


Figure 7. Adjacency tensor of STDUs. (a) Adjacency matrix, (b) structure of adjacency tensor, and (c) correlation degree between STDUs.

The Construction of the Self-Correlation Network

Firstly, in the process of self-correlation network construction, we choose some features of remote sensing data as a set of vectors. Then, we used the optimized FCM algorithm [43,44] to analyze the relationship between data containing the same feature from the grid dataset, to obtain a set of remote sensing datasets containing different geographic elements. The similarity can be obtained according to the distance of the grid data in the feature space, as the correlation degree.

The similarity between two different spatio-temporal data units in characteristics can reveal their potential correlation. Therefore, the correlation degree of remote sensing data can be measured by the similarity of features. Considering that remote sensing data has multiple features of different dimensions, it is necessary to calculate the similarity in different feature spaces to construct the joint similarity. Therefore, in the self-correlation network, metrics involved in the data relationship include the similarity of data, time, and space features, the joint similarity of multi-dimensional features, and correlation degree.

- Similarity in Data Features

Considering the difference in sources and types of remote sensing data, data features can usually be extracted, and can be regarded as different high-dimensional feature spaces. In order to describe the relationship of remote sensing data in data features, a general representation method is needed to calculate the similarity degree between remote sensing data of the same type based on data features, which can usually be expressed by the distance in the data feature space of remote sensing data.

With reference to the definition of Gaussian kernel function [35] and Euclidean distance [34], the similarity in data features between remote sensing data unit u and v is defined as:

$$\begin{aligned} \text{sim}_D(u, v) &= \exp\left(-\frac{\|F_i(u) - F_i(v)\|^2}{2\sigma_D^2}\right) \\ &= \exp\left(-\frac{\sqrt{\sum_{i=1}^r \alpha_i (F_i(u) - F_i(v))^2}}{2\sigma_D^2}\right) \\ &\text{s.t. } \sum_{i=1}^r \alpha_i = 1, 0 \leq \alpha_i \leq 1 \end{aligned} \quad (2)$$

where r represents the dimension of common data features of remote sensing data unit u and v , $F_i(u)$ and $F_i(v)$ represents the value of the i -th dimension data feature of u and v respectively, $\|F_i(u) - F_i(v)\|^2$ is the Euclidean distance of u and v in the r -dimensional

feature space, and α_i represents the weight of the i -th dimension data feature in the similarity judgment.

- Similarity in Time Features

Remote sensing data always has time features when it is generated. Considering that the change in the time dimension is continuous, time features of a data can be quantified at different time resolutions according to the need. The similarity in time features of remote sensing data can be represented with difference of the quantified time feature, which is:

$$\text{sim}_T(u, v) = \exp\left(-\frac{\|T(u) - T(v)\|^2}{2\sigma_T^2}\right) = \exp\left(-\frac{|T(u) - T(v)|}{2\sigma_T^2}\right) \quad (3)$$

where $T(u)$ and $T(v)$ represent the quantified time characteristic values of u and v respectively, such as the data production time expressed in milliseconds.

- Similarity in Spatial Features

Remote sensing data usually has a description of spatial features, which represent the spatial location and scope of remote sensing data. By calculating the spatial proximity of the remote sensing data, the similarity of the spatial features of the data can be obtained, and the correlation on the spatial features can be established as well.

Taking into account the continuity of remote sensing data in the spatial dimension, in order to facilitate the compression coding and distance measurement of the spatial features of the data, the space filling curve is used to fill the remote sensing data grid, converting a remote sensing data grid to a point on the space filling curve. Establish the mapping relationship between spatial features and points on the space filling curve in measuring the spatial similarity of spatio-temporal data units, then encode the points and use the size of the matched coding bits and the corresponding spatial resolution to obtain.

Define spatial similarity between spatio-temporal data units u and v as:

$$\text{sim}_S(u, v) = \exp\left(-\frac{\|S(u) - S(v)\|^2}{2\sigma_S^2}\right) = \exp\left(-\frac{|S(u) - S(v)|}{2\sigma_S^2}\right) \quad (4)$$

where $S(u)$ and $S(v)$ represent the spatial feature coding of u and v respectively.

- Joint similarity

In order to express correlation degree of spatio-temporal data units, treat the similarity of remote sensing data in data features, temporal features, and spatial features as the attributes of the remote sensing data correlation in three-dimensional space. Therefore, the joint similarity can be expressed as:

$$\begin{aligned} \text{sim}(u, v) &= \exp\left(\frac{\|\text{sim}_k(u, v)\|^2}{2\sigma^2}\right) \\ &= \exp\left(\frac{\sum_{k \in \{D, S, T\}} \beta_k (\text{sim}_k(u, v))^2}{2\sigma^2}\right) \\ \text{s.t. } \sum_{k \in \{D, S, T\}} \beta_k &= 1, 0 \leq \beta_k \leq 1, k \in \{D, S, T\} \end{aligned} \quad (5)$$

where β_T , β_S and β_D represent the weight of the similarity of the time, space and data features of remote sensing data in the relationship of the remote sensing data, respectively. In the actual application process, β_T , β_S and β_D can be dynamically adjusted according to the requirements of data relational retrieval, so as to construct the correlation in a single feature and in combinations of features in different dimensions.

- Correlation degree

We use the similarity of spatio-temporal data units in multiple dimensions to express the correlation degree between remote sensing data. Meanwhile, in order to reduce the cost in storage and computing, the threshold ε is specified when calculating the correlation degree. When it is less than ε , the correlation degree of two remote sensing data is considered

to be 0, that is, there is no correlation between them; when it is greater than or equal to ε , the joint similarity of the spatio-temporal data model can be used to express their correlation degree. The determination of the value of ε is achieved through a feedback adjustment mechanism. Firstly, we set an initial value for ε and calculate the correlation degree and build STCN. Then we performed associative queries based on STCN, and compare the results of the associated queries with the manually marked results to obtain the direction and offset of ε in further adjustment. Next, we use the adjusted ε to re-execute the above process until the result of the associated query is similar to the result of the manual mark to the greatest extent.

In summary, we define the correlation degree of the remote sensing data unit as:

$$relation(u, v) = \begin{cases} sim(u, v), & sim(u, v) \geq \varepsilon \\ 0, & sim(u, v) < \varepsilon \end{cases} \quad (6)$$

where

$$sim(u, v) = \exp\left(\frac{\beta_D \left(\exp\left(-\sqrt{\frac{\sum_{i=1}^r a_i (F_i(u) - F_i(v))^2}{2\sigma_D^2}}\right) \right)^2 + \beta_T \left(\exp\left(-\frac{|T(u) - T(v)|}{2\sigma_T^2}\right) \right)^2 + \beta_S \left(\exp\left(-\frac{|S(u) - S(v)|}{2\sigma_S^2}\right) \right)^2}{2\sigma^2}\right) \quad (7)$$

The corresponding pseudo-code can be found in Algorithm 1.

Algorithm 1. The algorithm implementation of the multi-source remote sensing data correlation method.

STSCN Construction Algorithm

Input: $STDU$

Output: $STSCN(U, E_U, A_U, R_U)$,

1. $U = \{STDU\}$;
 2. $E_U = \emptyset, A_U = \emptyset, R_U = \emptyset$;
 3. Use optimized FCM algorithm for clustering to get E' ;
 4. for $e'_l \in E'$
 5. $e_l = \emptyset$
 6. for $u_i, u_j \in e'_l$
 7. Calculate $sim_D(u_i, u_j), sim_T(u_i, u_j), sim_S(u_i, u_j)$;
 8. Calculate $sim(u_i, u_j), relation(u_i, u_j)$;
 9. Calculate $a_{i,j}, r_{i,l}$;
 10. if $r_{i,l} > \varepsilon$
 11. $A_U[i][j] = A_U[j][i] = a_{i,j}$;
 12. $R_U[i][l] = r_{i,l}$;
 13. $e_l = e_l \cup \{v_i\}$;
 14. $E = E \cup \{e_l\}$;
 15. end if
 16. end for
 17. end for
 18. return $STSCN(U, E_U, A_U, R_U)$.
-

The Construction of the Cross-Correlation Network

Based on the self-correlation network, we analyze the correlation degree of the remote sensing data within different self-correlation networks to establish the correlation and obtain a wider correlation network of remote sensing data, in order to provide full-time, full-type association query capabilities for remote sensing data.

On the basis of spectral clustering, the construction process of STCN can be seen as the correlation graph cut to obtain different subgraphs as different correlation networks.

In the process of graph cut, the main goal is to make the correlation degrees between spatio-temporal data objects in the same subgraph as high as possible, and the correlation degrees between spatio-temporal data objects in different subgraphs are as low as possible.

Firstly, we take spatio-temporal data objects as nodes to generate a spatio-temporal cross-correlation network based on spectral clustering. Remote sensing data is a subgraph of the remote sensing data object graph. In order to divide the remote sensing data object with the highest degree of correlation into the same correlation network, the remote sensing data object correlation graph needs to be cut to generate k subgraphs that are not connected to each other. The set of each subgraph is defined as $G_1, G_2, G_3, \dots, G_k$, where $G_1 \cup G_2 \cup G_3 \cup \dots \cup G_k = G$, $G_i \cap G_j = \emptyset$. Referring to the graph cutting process in the literature [45], the generation process of the definition is as follows.

1. Establish similarity matrix, adjacency matrix and degree matrix for the correlation among STDOS. The adjacency tensor of STCCN can be degenerated into an adjacency matrix $A = (a_{i,j})$, let

$$a_{i,j} = \begin{cases} \text{relation}(U_i, U_j), & \text{if } U_i \sim U_j \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

As for U_i in STCCN, the degree d_i is defined as the sum of the weights of all edges connected to it, that is

$$d_i = \sum_{j=1}^{|U|} a_{i,j} \quad (9)$$

The Laplacian matrix of STCCN can be calculated from the degree matrix and the adjacency matrix as $L = D - W$.

2. Calculate and standardize the Laplacian matrix of the correlation graph of STDOS;
3. Calculate the smallest eigenvalue and its corresponding eigenvector on the standardized matrix;
4. Standardize obtained eigenvectors by rows to form a characteristic matrix of size $n \times k$;
5. Regard matrix M as n k -dimensional samples, perform FCM clustering, and divide the complete set of spatio-temporal data objects into c clusters, $C = \{c_1, c_2, \dots, c_c\}$, corresponding to c STCCNs.

In the above process, a set of STDOS can be divided into c STCCNs, and each STDO belongs to only one STCCN. However, in practical applications, a piece of STDOS may belong to different STCCNs, with different correlation degrees. So, it is convenient to control the data returned meeting different association query conditions. Therefore, it is necessary to expand the STCCN obtained in the above clustering process. In summary, the expansion process of STCCN is as follows. The corresponding pseudo-code can be found in Algorithm 2.

1. Take a remote sensing data cross-correlation network c_j from $C(c_1, c_2, \dots, c_c)$;
2. Take a remote sensing data object o_i from c_j , and calculate the correlation degree of o_i to the element c_l in \bar{c}_j as

$$\text{relation}(o_i, c_l) = \frac{1}{|c_l|} \sum_{o_m \in c_l} \text{relation}(o_i, o_m) \quad (10)$$

where \bar{c}_j is the complement of c_j , that is, the union of other cross-correlation networks in C except c_j ;

3. If $\text{relation}(o_i, c_l)$ obtained in step 2 is greater than the threshold ε_c , it is considered that o_i belongs to c_l , let $c_l = c_l \cup \{o_i\}$;
4. Execute step 2 to step 4 cyclically, until all the STDOS in c_j have verified the correlation of each subnet;

5. Execute step 1 to step 4 cyclically, until all the spatio-temporal cross-correlation networks in C have completed the completion operation.

Algorithm 2. The algorithm implementation of the multi-source remote sensing data correlation method.

STCCN Construction Algorithm

Input: Spatio-temporal data objects O

Output: $STCCN(O, E_O, A_O, R_O)$

```

1.  $E_O = \emptyset, A_O = \emptyset, R_O = \emptyset$ 
2. for  $o_i, o_j \in O$ 
3.    $r_{i,j} = relation(o_i, o_j);$ 
4.    $R_o[i][j] = R_o[j][i] = r_{i,j};$ 
5.    $d_i = \sum_{o_j \in O} r_{i,j};$ 
6.    $D[i][i] = d_i;$ 
7. end for
8. Calculate  $L = D - R;$ 
9. Standardize  $L;$ 
10. Get the smallest  $k$  eigenvalues and eigenvectors of  $L;$ 
11. Standardize the eigenvectors by rows to obtain an eigenmatrix  $M$  of size  $|O| \times k;$ 
12. Consider  $M$  as  $k$ -dimensional samples, and use FCM for clustering to obtain the divided cluster set  $C;$ 
13. for  $o_i \in O, c_l \in C$ 
14.   if  $o_i \notin c_l$ 
15.     Calculate  $relation(o_i, c_l);$ 
16.     IF  $relation(o_i, c_l) > \varepsilon_C$ 
17.        $c_l = o_i \cup c_l;$ 
18.   end if
19. end if
20. end for
21. for  $c_l \in C$ 
22.   for  $o_i, o_j \in c_l$ 
23.     Calculate  $a_{i,j}, r_{i,l};$ 
24.      $A_o[i][j] = a_{i,j};$ 
25.      $R_o[i][l] = r_{i,l};$ 
26.   end for
27. end for
28. return,  $STCCN(O, E_O, A_O, R_O)$ 

```

In the above two sections, multiple adjustable parameters are involved. In process of self-correlation network construction, there are adjustable parameters such as the contribution weights β_T , β_S , and β_D in the calculation of the correlation degree, and the widths σ_T , σ_S , and σ_D of the Gaussian kernel function, and the judgment of the correlation degree of spatio-temporal data objects threshold ε , etc. In process of cross-correlation network construction, adjustable parameters include the number of feature vectors k during spectral clustering, the number of correlation networks c , and the correlation threshold ε_c , etc.

The settings of these parameters play an extremely important role in the construction of the spatio-temporal correlation network. It has a great impact on the quality of the association query results as well. In order to obtain the optimal correlation network, a feedback mechanism is introduced. We use the artificial marking results and the query results based on the correlation network as input to calculate the evaluation index to measure the quality of the association query. Then, we calculate the offset of the above parameters, and realize the step-by-step feedback from the spatio-temporal cross-correlation network to the self-correlation network. Finally, we obtain the value of each parameter meeting the optimal result of the correlation network. In Section 3, we will take threshold ε as an example, to show how different parameter values will affect the results of the associated network construction.

3.2.3. Spatio-Temporal Correlation Network Update

Taking into account the features of remote sensing data such as huge amount and fast update speed, it is necessary to dynamically update the correlation network to meet the demand for fast correlation retrieval. However, comparing these remote sensing data objects with all the data stored to establish correlation network after feature extraction will lead to a great consumption of computing resources and significant decline in the efficiency of data storage and access, which is contrary to the original intention of establishing the spatio-temporal correlation. Therefore, it is especially important to realize an effective dynamic construction method of remote sensing data correlation.

If there are correlations between o_i and o_j , and between o_j and o_k , then o_i and o_k may be related to each other. Therefore, we can take advantage of the transitivity of spatial-temporal correlation to reduce the amount of calculation in the spatio-temporal correlation network update. The spatio-temporal correlation network update process is as followed:

1. Encapsulate the features of the remote sensing data object o into query conditions to query eligible candidate remote sensing datasets O' ;
2. Query the correlation network set C' , corresponding to all spatio-temporal data objects in O' ;
3. Take an element c' from C' and calculate the correlation degree between o and c' ;
4. Determine whether there is relationship based on the correlation threshold. If there is a correlation, update the structure of the correlation network accordingly, that is, update the adjacency matrix and the correlation matrix stored;
5. Repeat steps 3 and 4 until all elements are traversed.

The pseudo-code is given in Algorithm 3.

Algorithm 3. The algorithm implementation of the multi-source remote sensing data correlation method.

STCN Update Algorithm

Input: remote sensing data object o

Output: $STCCN(O, E_O, A_O, R_O)$

1. Transform features of o into query conditions;
 2. Traverse data table to get a set of STDOs as O' ;
 3. for $o'_i \in O'$
 4. Get a set of STCCNs which o'_i belongs to as C'_i ;
 5. $C' = C' \cup C'_i$
 6. end for
 7. for $c'_l \in C'$
 8. Calculate $relation(o_i, c_l)$;
 9. IF $relation(o, c'_l) > \varepsilon_C$
 10. $c_l = o \cup c'_l$;
 11. $C = C \cup c_l$
 12. end for
 13. Update A_O and R_O ;
 14. return $STSCN(U, E_U, A_U, R_U), STCCN(O, E_O, A_O, R_O)$
-

3.2.4. Algorithm Implementation

The algorithm implementation of the multi-source remote sensing data association method is given in Algorithm 4.

Algorithm 4. The algorithm implementation of the multi-source remote sensing data association method.

Multi-Source Remote Sensing Data Association Method Algorithm

Input: remote sensing data

Output: $STCCN(O, E_O, A_O, R_O)$

```

1.  $U = \emptyset, O = \emptyset$ ;
2. switch (OperateType)
3. {
4. case CONSTRUCTION:
5.   Extract data, spatial and time features to construct  $FD$ ;
6.   Analyze the data structure to construct  $SD$ ;
7.   Encapsulate  $STDU(PD', SD', FD')$ , where  $SD' \subseteq SD, FD' \subseteq FD$ ;
8.   Get  $STSCN(U, E_U, A_U, R_U)$  through STSCN Construction Algorithm
9.   for  $e_l \in E_U$ 
10.     $o = \emptyset$ ;
11.    for  $u_i, u_j \in e'_l$ 
12.      $o = o \cup \{u_i\}$ ;
13.    end for
14.     $O = O \cup \{o_l\}$ 
15.  end for
16.   $E_O = \emptyset, A_O = \emptyset, R_O = \emptyset$ ;
17.  Get  $STCCN(O, E_O, A_O, R_O)$  through STSCN Construction Algorithm;
18. break
19. case UPDATE:
20.   Extract data, spatial and time features to construct  $FD$ ;
21.   Analyze the data structure to construct  $SD$ ;
22.   Encapsulate  $STDU(PD', SD', FD')$ , where  $SD' \subseteq SD, FD' \subseteq FD$ ;
23. Get  $STSCN(U, E_U, A_U, R_U)$  through STSCN Construction Algorithm;
24. for  $e_l \in E_U$ 
25.    $o = \emptyset$ ;
26.   for  $u_i, u_j \in e'_l$ 
27.     $o = o \cup \{u_i\}$ ;
28.   Update  $STCCN(O, E_O, A_O, R_O)$  through STSCN Update Algorithm
29.   end for
30.    $O = O \cup \{o_l\}$ 
31. end for
32. break
33. }
34. return  $STCCN(O, E_O, A_O, R_O)$ 

```

3.3. Local Association Query Based on Correlation Network

3.3.1. Spatio-Temporal Index

Spatio-temporal correlation network is a network with hierarchical, localized, multi-scale and other characteristics, defined with the hypergraph model. It can be transformed into a spatial data index structure through compression coding to optimize spatial query capabilities, and perform efficient remote sensing data retrieval. The spatio-temporal index proposed is an index established to meet multi-level, multi-scale, and multi-category joint query requirements. It solves the problem of large-scale remote sensing data from the aspects of feature correlation and spatio-temporal correlation. The spatio-temporal index consists of the multi-dimensional feature index and the association index, described as follows.

- Multi-dimensional feature index

The multi-dimensional feature index table is a secondary index table established on the basis of data storage table, including several parts such as Rowkey, Column and version, as shown in Figure 8. Rowkey is used to search the ID of remote sensing data by their

feature. There is only one Column, which is used to store the ID of the remote sensing data. version is used to identify different index versions by their corresponding creation time.

	Rowkey	Column	Version
Space Dimension	SPAC_geocode1_key1	key1	TimeStamp1
	SPAC_geocode2_key2	key2	TimeStamp2
Time Dimension	TIME_yyyyMMdd1_key1	key1	TimeStamp1
	TIME_yyyyMMdd2_key2	key2	TimeStamp2
	TIME_yyyyMMdd2_key3	key3	TimeStamp3
Feature Dimension	FEAT_name1_value1_key1	key1	TimeStamp1
	FEAT_name1_value2_key2	key2	TimeStamp2

Figure 8. Multi-dimensional feature index on remote sensing data.

The Rowkey in the multi-dimensional feature index table consists of the retrieval type, retrieval value, and the key of the unit in the data table, integrating the spatial, time, and data features of the remote sensing data into the index structure to ensure that similar data in different feature spaces can be stored adjacently. More detailed descriptions are as follows.

- “SPAC_geocode1_key1” means that the value of the key corresponding to the grid whose spatial feature coding value is “geocode1” is key1;
- “TIME_yyyyMMdd1_key1” means that the value of the key corresponding to the grid whose time attribute value is “yyyyMMdd1” is key1;
- “FEAT_name1_value1_key1” means that the value of the data feature name1 is the grid data of value1, and the value of the key stored in the data table is key1;
- “SPAC”, “TIME” and “FEAT” are used to distinguish spatial, time and data features.

The multi-dimensional feature index table contains column family Key and column Key, which are used to store the key of the grid data in the data table. The version of the grid data in the index table is the data production time.

- Association index

The association index table is an index table established on the basis of spatio-temporal correlation network, as shown in Figure 9.

Rowkey	Column	Version
DO_dataID1_stcnID1	stcnID1	TimeStamp1
DO_dataID1_stcnID2	stcnID2	TimeStamp2
RN_stcnID1_dataID1	dataID1	TimeStamp1
RN_stcnID1_dataID2	dataID2	TimeStamp2
RN_stcnID1_dataID3	dataID3	TimeStamp2

Figure 9. Association index on remote sensing data.

As shown in Figure 9, the Rowkey in the association index consists of several parts, unique identification of correlation network and unique identification of remote sensing data, etc.

- “DO_dataID1_stcnID1” indicates that the remote sensing data dataID1 belongs to the spatio-temporal correlation network stcnID1;
- “CN_stcnID1_dataID1” means that the spatio-temporal correlation network stcnID1 contains the remote sensing data dataID1.

The association index table contains column family named Key, and column named Key, which is used to store the key of the grid data in the data table. The version of the grid data in the index table is the data production time.

3.3.2. Local Association Query

Based on the multi-dimensional feature index and association index, the spatio-temporal correlation network can be used to realize the local association query of remote sensing data, as shown in Figure 10. The detailed steps are as follows.

1. Extract the eligible candidate remote sensing dataset O' from the database according to query conditions;
2. Query that the correlation network set C' consists of all the correlation network spatio-temporal data objects in O' belongs to;
3. Encapsulate the query conditions as virtual spatio-temporal data objects o ;
4. Calculate the correlation degree between o and the element c'_l in set C' , in turn;
5. Put the spatio-temporal data objects in c'_l into the set of spatio-temporal data objects O .
6. Remove the remote sensing data in O whose correlation degrees are less than ε , meaning that they do not meet the requirement, and sort spatio-temporal data objects by correlation degree.

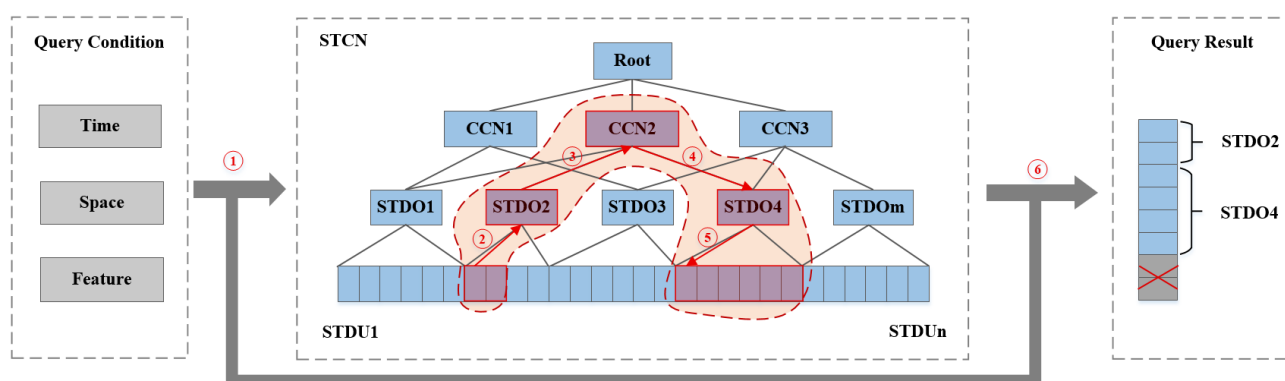


Figure 10. The flow of local association query.

The algorithm realization of local association query process on remote sensing data is given in Algorithm 5.

Algorithm 5. Local association query for remote sensing data.**Local Association Query for Remote Sensing Data****Input:** Query conditions, threshold ε **Output:** O

```

1.  $O = \emptyset, C = \emptyset, C' = \emptyset$ ;
2. Traverse data table to get a set of STDs as  $O'$ ;
3. for  $o'_i \in O'$ 
4.   Get a set of STCCNs which  $o'_i$  belonging to as  $C'_i$ ;
5.    $C' = C' \cup C'_i$ 
6. end for
7. Convert query conditions to a set of virtual STDs as  $O$ ;
8. for  $c'_l \in C'$ 
9.   Calculate  $relation(o, c'_l)$ ;
10.  if  $relation(o, c'_l) > \varepsilon$ 
11.    for  $o'_i \in c'_l$ 
12.       $c_l = \emptyset$ 
13.      if  $relation(o, o'_i) > \varepsilon$ 
14.         $O = O \cup \{o'_i\}$ ;
15.         $C = C \cup c_l$ ;
16.      end if
17.    end for
18.  end for
19. Sort the elements in  $O$  by correlation degree;
20. return  $O$ ;
```

4. Experiments**4.1. Experimental Materials**

We take remote sensing data products from the following sources as test data to effectiveness and versatility of STLAQ's applications on multi-source remote sensing data.

- Remote sensing satellite image data

The remote sensing satellite image data used in the experiments come from Microsoft's Bing map service. Data in the regions of the Hawaiian Islands and Qinghai-Tibet Plateau lakes are selected as the test dataset, in PNG format. There are 256×256 pixels in each image, of which the spatial resolution is 10 m.

- Three-dimensional terrain data

Three-dimensional terrain data come from the Google Earth platform. Data in the region of Qinghai-Tibet Plateau lakes are selected as the test dataset, in PNG format. In each image, there are 256×256 pixels to present the spatial coverage, of which the spatial resolution is 1.3 m.

- Street view image data

The street view data come from the Baidu panoramic map platform, consisting of parts of the Qinghai-Tibet Plateau, in PNG format.

- Lake distribution data of the Qinghai-Tibet Plateau

The distribution data of lakes on the Qinghai-Tibet Plateau come from the National Qinghai-Tibet Plateau Science Data Center, such as China lake dataset (1960–2020), in Shape File format.

4.2. The STCN Construction Results

We encapsulate remote sensing satellite image data into STDUs, marked with rectangles in Figure 11. Then, the STCN model proposed in this paper takes STDUs at different resolution scales as the smallest unit of remote sensing data on the basis of multi-scale

division, and establishes the relationship among STDOS based on the STDU, so as to realize different remote sensing data correlations across different resolution scales.

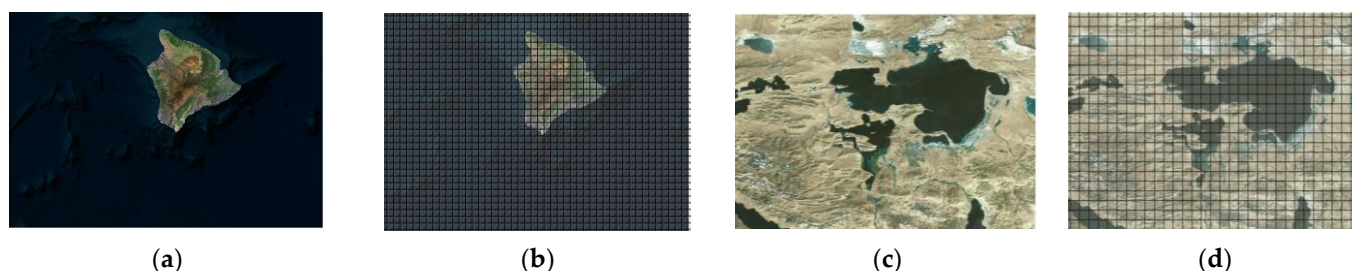


Figure 11. Encapsulation result of STDOS. (a) Original dataset; (b) STDOS; (c) original dataset; (d) STDOS.

The construction results of self-associated networks and cross-associated networks are given in Figures 12 and 13, respectively. In Figure 12b,d, we use rectangles to mark data units. In particular, in Figure 12d, white rectangles and red rectangles are used to represent spatio-temporal data units of two different resolution levels, level 12 and level 13, respectively. It can be seen that in association query of remote sensing data, correlations can be established within remote sensing data of the same scale and of different scales. Therefore, the STCN model proposed in this paper is effective in cross-scale association query, which can be used to establish the relationship between remote sensing data at different resolutions. Thus, the STCN can be used to solve the problems caused by the different shapes of feature elements, and to overcome difficulties in expressing data characteristics uniformly.

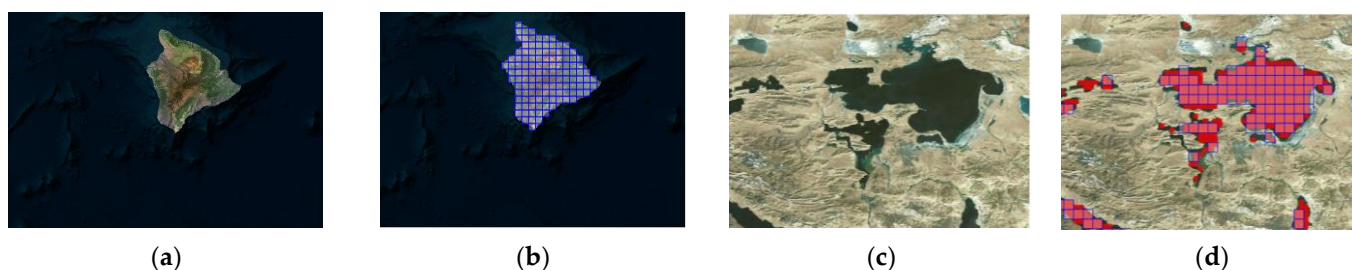


Figure 12. Construction result of STCN. (a) Original dataset; (b) STCN; (c) original dataset; (d) STCN.

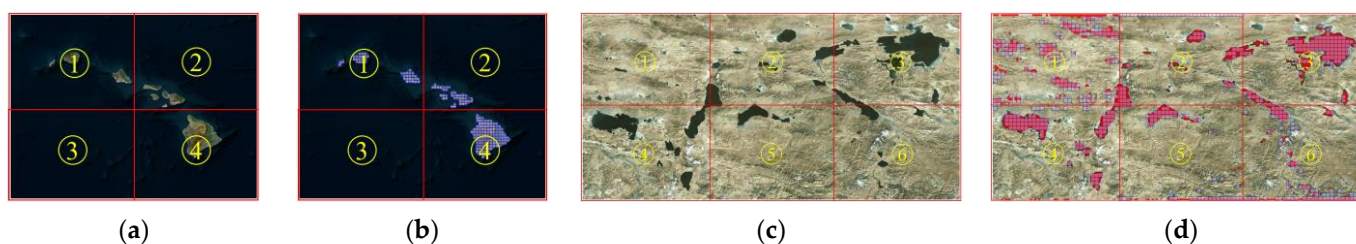


Figure 13. Construction result of STCCN. (a) Original dataset; (b) STCCN; (c) original dataset; (d) STCCN.

The settings of these parameters play an extremely important role in the construction of the spatio-temporal correlation network. We take α , the feature weight of mean, and ε , the threshold of correlation, as examples, to show how they will affect the results of the correlation network construction, as given in Figure 14. Four indicators, recall rate, precision rate, F1 rate and Kappa coefficient, are used to evaluate the clustering results of remote sensing data.

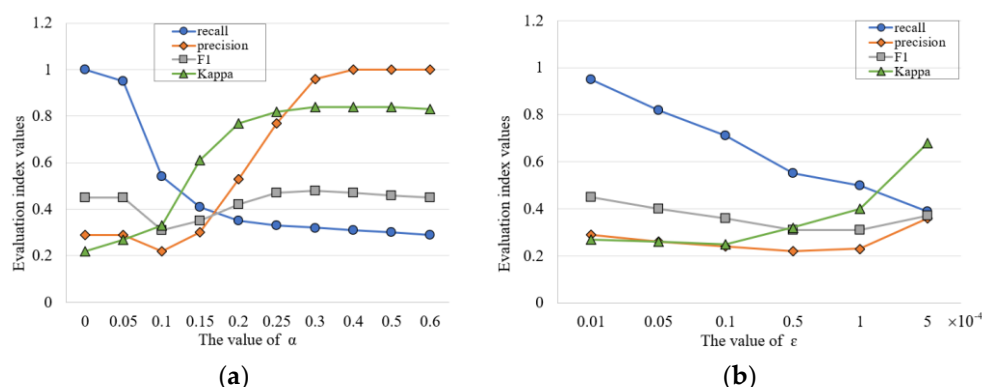


Figure 14. Evaluation index values. (a) Evaluation index values on various α ; (b) evaluation index values on various ϵ .






















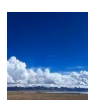












As shown in Figure 14a, a larger value of α always leads to lower recall rate, higher precision, higher F1, and higher Kappa coefficient. In addition, as shown in Figure 14b, if the value of ϵ is too small, it will cause redundant data, and if the value of ϵ is too big, it will cause greater incompleteness.

4.3. Local Association Query Results

4.3.1. Association Query Evaluation

We take remote sensing data products introduced in Section 4.1 as test datasets, and we take “lake images in a certain geographic range” as the query target to perform local association query tests on multiple types of remote sensing datasets, as shown in Table 2. Considering the convenience of reading, the satellite image data is expressed in STCNs, and rectangular boxes are used to identify STDUs in the self-correlation network.

Table 2. Remote sensing data related query results.

Target	Lake Data in a Certain Geographic Area									
Query conditions	Spatial region	(86.43°E, 28.24°N) ~ (92.61°E, 32.64°N)								
	Time region	Unlimited								
	Feature query condition	Grayscale mean, standard deviation, grayscale histogram entropy, energy, direction gradient histogram, etc.								
Query result	Satellite image data									
	3D terrain data									
	Streetview image data									
	Lake distribution vector data									

4.3.2. Association Query Performance

We use Apache JMeter as the performance testing tool to continuously access the data query interface, to test the retrieval efficiency of three different index structures. The index structures used in the test are as follows.

- Perform remote sensing data retrieval based on the STLAQ proposed in this article;
- Perform remote sensing data retrieval based on Geomesa, a distributed architecture for spatio-temporal fusion, proposed in [18]. We built indexes on the remote sensing data in time and event dimensions.

- Perform remote sensing data retrieval based on a customizable multi-dimensional index structure named ST+, proposed in [22].

In order to verify the retrieval efficiency of the STLAQ, we choose the remote sensing image data of layers 0–9 in the world as the test dataset, a total of 690,000. The experiment process is as follows.

1. We perform spatial retrieval, time retrieval, and feature retrieval on tables that store different amounts of data, respectively.
2. Then, we use query windows of different sizes to carry out retrieval in multiple dimensions such as space, time, and feature, on the table with the largest amount of stored data.
3. Finally, we adopt two ways to query the combination of two dimensions, space-time and space-feature, to test the efficiency of STLAQ in multi-dimensional combination query.

The results of local association query on spatial, time and feature dimension are given in Figures 15–17, respectively.

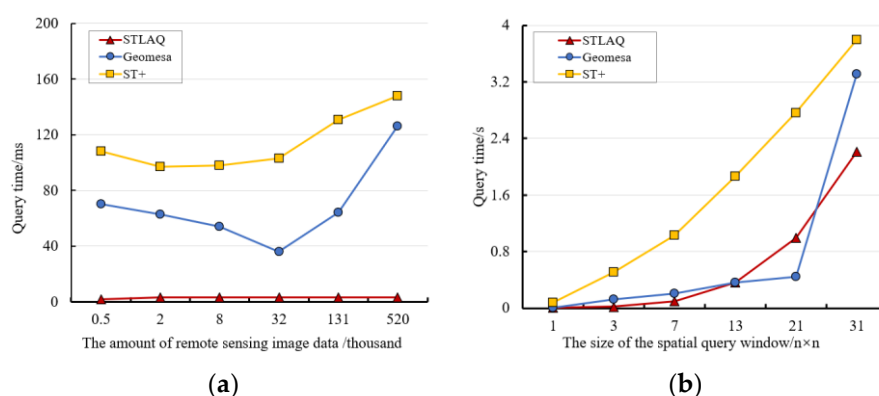


Figure 15. Local association query performance in spatial dimension. (a) Query on different storage data volumes; (b) query with spatial windows of different sizes.

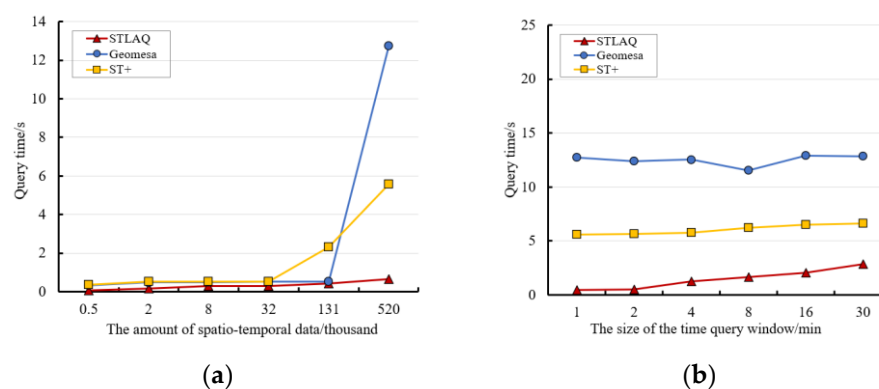


Figure 16. Local association query performance in time dimension. (a) Query on different storage data volumes; (b) query with time windows of different sizes.

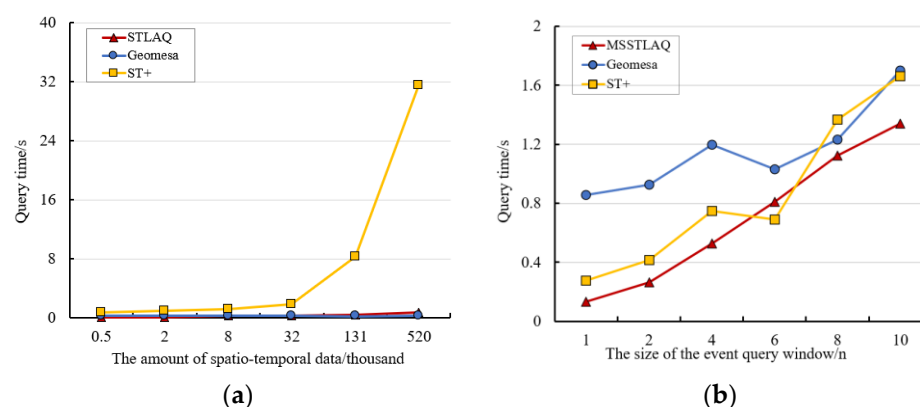


Figure 17. Local association query performance in feature dimension. (a) Query on different storage data volumes; (b) query with feature windows of different sizes.

It can be seen from Figure 15 that the STLAQ have better query performance on spatial association query, compared with Geomesa and ST+. As the amount of remote sensing data grows, the ST+ and Geomesa show a significant decline in query efficiency, while the STLAQ is relatively stable. With the expansion of the spatial query window, the STLAQ shows the slowest decline in efficiency.

In general, the STLAQ performs better in remote sensing data association queries on time dimension, followed by Geomesa, and ST+ performs poorly, according to the result given in Figure 16. The query efficiency of the ST+ index drops sharply as the amount of stored data increases, compared to the other two indexes. As for queries in time windows of different sizes, the ST+ index fluctuates to a certain extent, while the Geomesa index is relatively stable. The STLAQ has the highest query efficiency, which slightly decreases with the expansion of the query window.

As shown in Figure 17, when the feature query window is small, STLAQ has the highest query efficiency. The increase of the window size leads to a significant downward trend in query efficiency of the three indexes. Among them, the Geomesa index has the slowest decline, STLAQ's is second, and the decline of ST+ is the most severe.

The results of association query on space-time and space-feature dimensions are given in Figure 18. In the experiments, the data of layer 9 is queried by default.

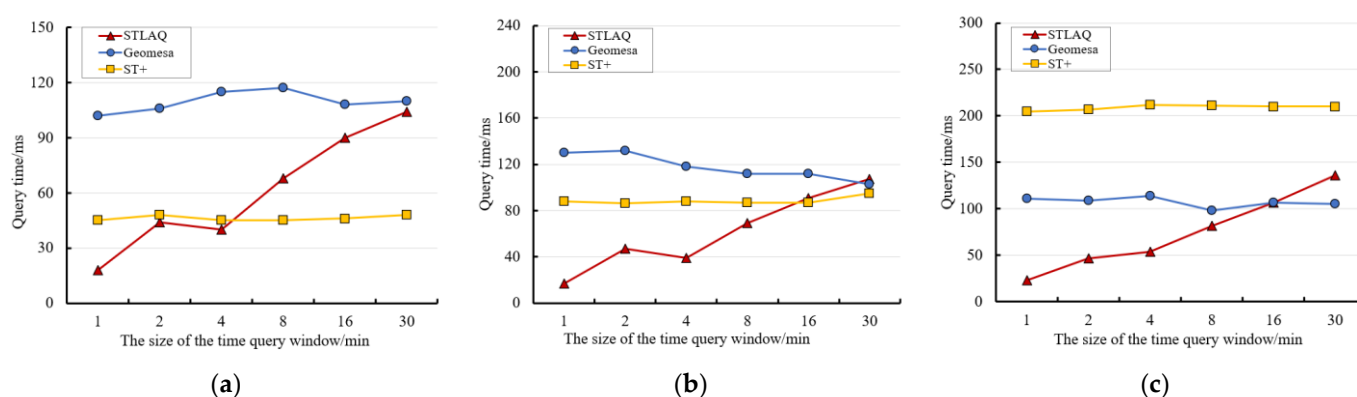


Figure 18. Local association query performance in spatial and time dimension with different window sizes. (a) Query with spatial window size of 7×7 ; (b) query with spatial window size of 13×13 ; (c) query with spatial window size of 21×21 .

According to the results given in Figure 18, STLAQ is more sensitive to the size of the time query window. Under different spatial query ranges, it always shows that as the time query range increases, the query efficiency slightly decreases; while the other two indexes remain relatively stable. As the range of spatial queries increases, the gap between the STLAQ and Geomesa indexes and the ST+ index gradually increases.

It can be seen from Figure 19 that, although the spatial query ranges are different, the query performance of STLAQ and Geomesa always decreases gradually as the time query range increases. ST+ has no advantage in feature query, and the efficiency gap between it and the other two indexes gradually widens with the gradual expansion of the query window.

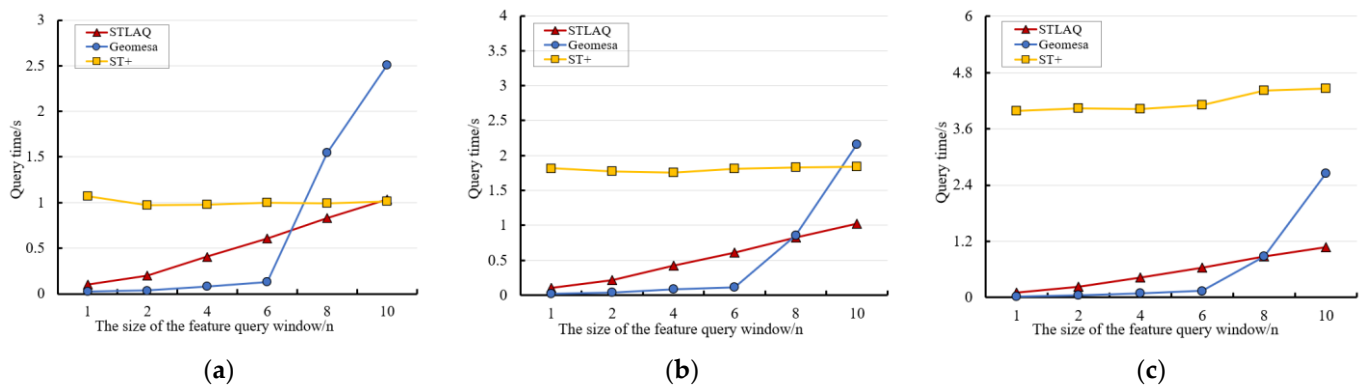


Figure 19. Local association query performance in spatial and feature dimension with different window sizes. (a) Query with spatial window size of 7×7 ; (b) query with spatial window size of 13×13 ; (c) query with spatial window size of 21×21 .

According to the above results, the proposed STLAQ has good efficiency in association queries of massive remote sensing data in spatial, temporal, and feature dimension, especially when querying in a combination of a large spatial range and a small time and feature range. There are two main reasons:

- STLAQ has established indexes on both the features and association, and maps multi-dimensional features to a one-dimensional space. It avoids the drawbacks of a sharp increase in the scanning range during data retrieval as the data dimension increases;
- When constructing indexes on the time and feature dimensions, the sequential coding method ensures the locality of the data. This also brings about the insufficient matching accuracy when the scope of the query on these two dimensions expands, resulting in a decrease in retrieval performance.

In the future, we will optimize from two aspects.

- Firstly, under the premise of ensuring the continuity of time and attributing characteristics, we should optimize the structure of the index, improve the process of feature encoding, and achieve a more accurate match between the query range and the feature index;
- Secondly, we should introduce a parallel retrieval mechanism to improve data at the system level search efficiency.

Generally speaking, the query efficiency is usually contradictory to the resource occupation. The improvement of query efficiency often leads to more consumption of resources such as CPU, memory, disk reads and writes and other resources. In order to verify the performance of STLAQ in terms of resource occupancy, we take spatial association query and time association query as examples, to compare the resource consumption of STLAQ and other indexes. The results are given in Figure 20.

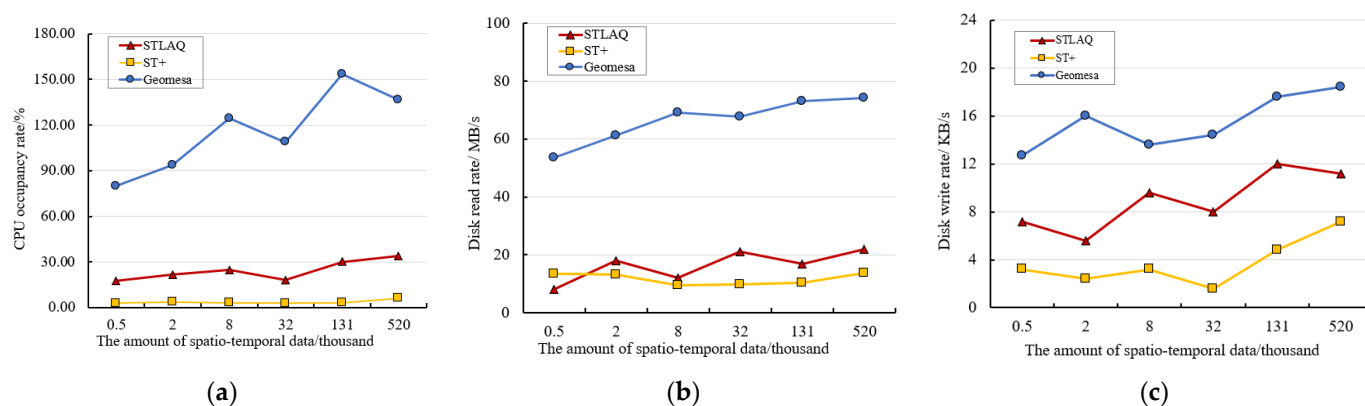


Figure 20. Resource occupancy rate in association query. (a) CPU occupancy rate, (b) disk read rate, (c) disk write rate.

In Figure 20, ST+ has the lowest resource occupancy rate, followed by STLAQ. Geomesa, which performs well in the previous test, consumes a lot of resources such as CPU, disk read and write. There is a slow increase in the value of resource occupancy metrics with data quantity unceasing increase. Therefore, it can be considered that although the STLAQ has a large contribution to query efficiency, it will not carry out more consumption to resources.

To sum up, the STLAQ proposed in this paper can break through the limitation of resolution scale. It establishes a cross-scale correlation between multi-source remote sensing data with spatial, time, and date features. Moreover, it makes the resource consumption of high-precision data query tend to be flat. The STLAQ case expand data sharing mode, and has high practical significance in the application of remote sensing data.

5. Conclusions

In view of the current issues of remote sensing big data management and retrieval, we propose a spatio-temporal local association query algorithm, STLAQ. We design a spatio-temporal data model and a bottom-up spatio-temporal correlation network. We adopt the method of the FCM clustering and the method of spectral clustering, to measure the correlation between spatio-temporal correlation networks. Then, we design the spatio-temporal index and local association query method based on spatio-temporal correlation networks, to provide multi-dimensional association query capabilities for remote sensing big data. We construct the STCN and carry out query experiments on several typical remote sensing datasets. The results show that the STLAQ can effectively establish cross-scale correlations of remote sensing data, and can provide capabilities of joint retrieval in multiple dimensions, such as spatial features, time features and data features. The STLAQ provide a solution to deal with the problems of implicit correlation and fast service under large-scale remote sensing data, especially with high spatial resolution in various fields. It expands the sharing mode of remote sensing data.

Considering that multi-source remote sensing data have broad application prospects, it is necessary to improve the adaptive ability of the algorithm in the later stage. Future studies based on neural network learning will be performed, to speed up the convergence of network weights, reducing the influence of experience. With the support of parallel computing technology, a distributed framework of data index and analysis would be provided to enhance the practicality of the algorithm. On the other hand, the dynamic expansion and maintenance method of the index during spatio-temporal insertion is also one of the main contents of future research.

Author Contributions: Y.H. proposed the concept; L.Z. and X.S. designed the STLAQ, made the first prototype of the STLAQ and wrote the original draft; X.T. and K.F. reviewed and edited the manuscript; Y.H. provided the project. All the authors contributed equally to the revisions of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China, grant number 2017YFC0821954.

Data Availability Statement: Not applicable.

Acknowledgments: Lake distribution data of the Qinghai-Tibet Plateau are provided by National Tibetan Plateau Data Center, 2019. doi:10.11888/Hydro.tpd.c.270302. CSTR:18406.11.Hydro.tpd.c.270302. (accessed from <https://data.tpd.c.ac.cn/> on 17 December 2020).

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

References

- Li, D.R.; Ma, J.; Shao, Z.F. Discussion on spatio-temporal big data and its application. *Satell. Appl.* **2015**, *9*, 7–11.
- Yang, C.; Huang, Q.; Li, Z.; Liu, K.; Hu, F. Big data and cloud computing: Innovation opportunities and challenges. *Int. J. Digit. Earth* **2017**, *10*, 13–53. [CrossRef]
- Wang, J. Spatio-temporal big data and its application in smart cities. *Satell. Appl.* **2017**, *5*, 10–17.
- Li, D. The intelligent processing and service of spatio-temporal big data. *J. Geo-Inf. Sci.* **2019**, *21*, 1825–1831.
- Li, D. Towards geo-spatial information science in big data era. *Acta Geod. Cartogr. Sin.* **2016**, *45*, 379–384.
- Zhang, Y. *Research on the Theory and Key Technology of Global Spatial Information Multi-Grid with China's Geographic Characteristics Considered*; Huazhong University of Science & Technology: Hangzhou, China, 2014.
- Wang, S.; Zhong, Y.; Wang, E. An integrated GIS platform architecture for spatio-temporal big data. *Future Gener. Comput. Syst.* **2019**, *94*, 160–172. [CrossRef]
- Chen, X.; Wu, J.; Yuan, G. Research on the construction of spatio-temporal information cloud platform for big data. *Geomat. Spat. Inf. Technol.* **2020**, *43*, 138–140.
- Hua, Y.; Zhou, C. Description frame of data model of multi-granularity spatio-temporal object for pan-spatial information system. *J. Geo-Inf. Sci.* **2017**, *19*, 1142–1149.
- Huang, X. *Research on Spatio-Temporal raster Data Modeling Based on Grid Mode*; Zhejiang University: Hangzhou, China, 2015.
- Yuan, F. *A New Strategy of Storage & Retrieval for Massive Tile Data of Remote Sensing Images*; University of Electronic Science and Technology of China: Chengdu, China, 2013.
- Bentley, R.A.F.J.L. Quad trees a data structure for retrieval on composite keys. *Acta Inform.* **1974**, *4*, 1–9.
- Robinson, J.T. The K-D-B-tree: A search structure for large multidimensional dynamic indexes. In Proceedings of the 1981 ACM SIGMOD International Conference on Management of Data; ACM: New York, NY, USA, 1981; pp. 10–18.
- Guttman, A. R-Trees: A Dynamic Index Structure for Spatial Searching; ACM: New York, NY, USA, 1984; pp. 47–57.
- Zhao, N. A hybrid structure of spatial multilevel index based on grids and R-tree. *Comput. Technol. Dev.* **2009**, *19*, 91–94.
- Kamel, I.; Falout, S.; Hilbert, C. Hilbert R-tree: An improved R-tree using fractals. In Proceedings of the 20th Very Large Databases, Santiago, Chile, 12–15 September 1994; pp. 500–509.
- Yang, Y. Tile quadtree and filling curve realizing massive terrain dataset management. *Comput. Eng. Appl.* **2016**, *52*, 192–196.
- Hughes, J.N.; Annex, A.; Eichelberger, C.N.; Fox, A.; Hulbert, A.; Ronquest, M. GeoMesa: A distributed architecture for spatio-temporal fusion. In *Geospatial Informatics, Fusion, and Motion Video Analytics V*; International Society for Optics and Photonics: Baltimore, MD, USA, 20 April 2015.
- Zhao, X.Y.; Huang, X.D.; Qiao, J.L. A spatio-temporal index based on skew spatial coding and r-tree. *J. Comput. Res. Dev.* **2019**, *56*, 666–676.
- Xu, J.F.; Tan, Y.L. Optimization of multidimensional index query mechanism based on HBase. *J. Comput. Appl.* **2020**, *40*, 571–577.
- Qian, B.Z. *Research on Linked Spatial Index Based on LSM-Tree*; Zhejiang University: Hangzhou, China, 2020.
- Zhao, Y.H.; Lü, L.; Xu, Q. A multidimensional retrieval strategy for massive spatio-temporal data. *Sci. Surv. Mapp.* **2020**, *45*, 203–208.
- Wu, Y.H.; Cao, X.F. Hilbert code index method for spatiotemporal data of virtual battlefield environment. *Geomat. Inf. Sci. Wuhan Univ.* **2020**, *45*, 1403–1411.
- Tang, N.; Zhu, Z.H.; Li, J.J. Temporal-spatial phase point moving object data indexing: PM-Tree. *Chin. J. Comput.* **2021**, *44*, 579–593.
- Wu, Y.; Chen, L.; Xiong, W.; Zhong, Z.N.; Jing, N. Multi-source geospatial data correlation model for efficient retrieval. *Chin. J. Comput.* **2014**, *9*, 1999–2010.
- Liu, P.F.; Cui, T.J. Research progress in geographic data association. *J. Tianjin Norm. Univ. (Nat. Sci. Ed.)* **2019**, *39*, 10–15.
- Wu, Y. *Research on Key Techniques of Entity Relationship Association Analysis Based on Graph*; National Defense University: Changsha, China, 2014.
- Xu, Y.J.; Tan, C.G. Research on the organization and application of spatio-temporal data. *Surv. Mapp. Bull.* **2017**, *2*, 98–101.
- Li, P.Y.; Pan, H.W.; Li, Q. Top-k query method of medical image based on relational graph model. *Comput. Technol. Dev.* **2009**, *19*, 91–94.
- Shi, Y.; Zhan, M.; Yin, L. Research on associated organization and analysis of target-oriented multi-source heterogeneous data. *Bull. Surv. Mapp.* **2015**, *1*, 102–104.

31. Lü, X.; Cheng, C.; Gong, J.; Guan, L. Review of data storage and management technologies for massive remote sensing data. *Sci. China Technol. Sci.* **2011**, *41*, 1561–1573. [[CrossRef](#)]
32. Zheng, W.U.; Chengming, L.I.; Pengda, W.U.; Jianming, S.H.E.N.; Wei, S.U.N. Integrated storage and management of vector and raster data based on Oracle database. *Acta Geod. Cartogr. Sin.* **2017**, *46*, 639–648.
33. Wootton, C. ISO 8601 Date Format Output. *Dev. Qual. Metadata* **2007**, 419–420.
34. Douglas, B.W.; West, J.L. *Introduction to Graph Theory*; Machinery Industry Press: Beijing, China, 2006.
35. Shao, S.; Lou, W.; Yan, L. Optimization of Algorithm of Similarity Measurement in High-Dimensional Data. *Comput. Technol. Dev.* **2011**, *2*, 7–10.
36. Wang, T. *High-Dimensional Data Clustering Based on Hypergraph Partition*; Lanzhou University: Lanzhou, China, 2016.
37. Jia, H.J.; Ding, S.F.; Shi, Z.Z. Approximate weighted kernel k-means for large-scale spectral clustering. *J. Softw.* **2015**, *26*, 2836–2846.
38. Yan, W. *Research on Image Feature Extraction Method*; Northwestern Polytechnical University: Xian, China, 2007.
39. Wang, C. *Study on Nondestructive Detection Method of Potato Grading Based on Multi-Source Information Fusion*; Huazhong Agricultural University: Wuhan, China, 2014.
40. Qing, Y.; Song, W. Remote sensing image feature extraction and selection and its application in image classification. *Sci. Servicing Mapp.* **2008**, *33*, 176–199.
41. Chen, P. *Research on Principal Component Analysis and Its Application in Feature Extraction*; Shanxi Normal University: Linfen, China, 2014.
42. Cao, M. *Research on Intelligent Recognition and Extraction of Feature Elements Based on Remote Sensing Images*; Changan University: Xian, China, 2015.
43. Xu, D. *Research on the Key Techniques of Multi-Source Remote Sensing Big Data Management under the Cloud Computing Environment*; University of Chinese Academy of Sciences: Beijing, China, 2018.
44. Zhang, M.; Yu, J. Fuzzy partitional clustering algorithms. *J. Softw.* **2004**, *15*, 858–869.
45. Zhou, K. *Theoretical and Applied Research on Fuzzy c-Mean Clustering and Its Cluster Validation*; Hefei University of Technology: Hefei, China, 2014.