Research Article

Quantized ternary pattern and singular value decomposition for the efficient mining of sequences in SRSI images

R. Angelin Preethi[1] · G. Anandharaj[2]

Abstract

The growth and development of particular region over time can be witnessed by remote sensing images. Although such raw images have less possibility to derive the insights, Serial Remote Sensing Images (SRSI) has the large potential to discover the patterns. The evolution of spatial patterns in various areas including urban development, expansion of vegetation cover and agriculture is the evidence for the utilization of SRSI accumulation. The application of conventional sequential pattern-mining algorithms on the SRSI images results in high computational complexity. This issue can be resolved by grouping the pixels and mining sequence patterns. A one-pass framework is introduced to compress and hide the data in the marked stream without any loss. In this paper, we proposed a Quantized ternary pattern based pixel grouping and Singular Value Decomposition—Run Length Coding based pattern mining. The algorithms are experimented using a dataset, namely, the Cropland data layer dataset. The proposed algorithm is efficient in terms of mining time and sequence pattern generation.

Keywords Serial remote sensing images · Quantized ternary pattern · Singular value decomposition · Run length coding · Geo-spatial image processing

1 Introduction

The growth of satellite and space technology has led to the revolution in science and technology.

The satellite images obtained from the space have several insights regarding climate change, town planning, change in urban development pattern, traffic flow pattern, migration of wild, environmental change, vegetation index fragmentation and so on [1]. The geospatial images contain spatial and temporal patterns, which require highly efficient mining techniques to derive insights. Among the satellite images, the Serial Remote Sensing Images (SRSI) has more potential to obtain effective results for earth observation. The SRSI images are captured from a particular area at different time periods that help to identify the habitat changes [2, 3]. The events that occur at certain time intervals such as seasonal changes, deforestation of forests, growth of building in metropolitans, and change in vegetation patterns during different seasons can be easily explored using SRSI. The images are explored based on the mining of frequent patterns that has a higher count than the threshold value. The change of events can be identified using the pixels [4].

The SRSI has the sequential patterns of spatial data using which more knowledge can be mined. In general, a sequential pattern mining algorithm is used to explore the frequent items from conventional databases. In the case of geo spatial data, the data has complex characteristics such as geospatial features, large volume, which hinders the traditional sequential pattern mining algorithm from

✉ R. Angelin Preethi, papupree@gmail.com; G. Anandharaj, younganand@gmail.com | ¹Department of Computer Science, Kamban College of Arts &Science College, Tiruvannamalai, India. ²Department of Computer Science, Adhiparasakthi College of Arts and Scince, Kalavai, India.

exploration. The major complexities of geo spatial datasets are that it has multifarious relationships between the objects and autocorrelation of spatial data is critical [5, 6]. To overcome these complexities several researchers have proposed different techniques used conventional sequential pattern mining algorithms by converting the geo spatial database into a transaction database. The conversion can be done by considering the pixels as an object and the value of pixels as events in the transaction [7, 8]. Even though, the conversion requires a large amount of time to get converted into vertical tables as it has a huge volume of data. The challenges can be resolved by rather analyzing the data at the area level than at the pixel level. The grouping of pixels tends to reduce the time complexity and computational complexity. In this paper, the Quantized ternary pattern based pixel grouping and Singular Value Decomposition (SVD)—Run Length Coding (RLC) based pattern mining is proposed. The pixel grouping algorithms groups the pixels block by block thereby reducing the computation overhead and reduces the time consumption. SVD based pattern mining algorithm is used to increase the robustness of the proposed algorithm and it resists against image processing attacks which improves the visual quality [34, 35]. It also satisfies the threshold support values and helps in generating highly accurate sequence patterns [9].

The remaining sections of the paper are ordered as follows: Section II reviews the applications and challenges in various spatial pattern mining algorithms. Section III has a detailed explanation of the proposed Quantized ternary pattern based pixel grouping and Singular Value Decomposition (SVD)—Run Length Coding (RLC) based pattern mining algorithms. The results of the proposed algorithm according to mining time and support values of pattern sequences and performance analysis of the proposed algorithm are discussed in Section IV along with the dataset utilized for experimentation. Section V concludes the paper by highlighting the features of the proposed work.

2 Literature review

In this section, existing works based on spatio temporal pattern mining algorithms for different applications and their issues and corresponding solutions derived by different researches were discussed.

In a fuel cell, the damages were identified using the sequence of events that occur during an acoustic emission using the Cluster Sequence Mining (CSM) algorithm. The correlated cluster pairs were extracted using the CSM algorithm from data that consist of a sequence of events. The spatial and temporal proximity patterns of the events were considered to pair the clusters. The intervals between the sequences were distributed in a periodic manner and the events were ordered in a proper sequence by the Co-occurring Cluster Mining (CCM) algorithm [10].

The Bayesian inference is used to infer the robustness based on the probability of time intervals. The events were correlated with the help of the Dynamic programming concept in accordance with their correspondence. The challenges in the CCM were that it requires both spatial and temporal proximity for evaluation but the order of the event occurrence is not considered. The probability density functions were assumed based on the inferences made by the Bayesian Inference that they follow a certain specific order. The accuracy of the inference was enhanced using the Dynamic Time Warping (DTW). The challenges were compromised by the CSM algorithm in the identification of damages in fuel cells [11].

The autocorrelation indices for local spatial data were applied to explore the patterns in fragmenting the vegetation land area. The vegetation land in the urban areas is cultivated with different crops and it results in high classification error when it is classified based on the data received from the land cover dataset in a discrete format [12, 13]. This issue was addressed using Local Indicators of Spatial Association (LISA) which helps in clustering the patterns of spatial data in an efficient manner. The authors utilized the information of spatial data for the Zimbabwe cultivation pattern to identify the hot and cold clusters in the corresponding region. The Z scores were utilized to highlight the statistical measure of each parameter in the dataset. The moving window method was adopted to analyze the vegetation pattern. The LIAS helped in increasing the accuracy of classifying the spatial patterns [14, 24]

In the network connection using wireless sensors, the sequential patterns were discovered using an incremental mining algorithm to make the mining process more efficient. A traditional mining algorithm, namely. PrefixSpan was enhanced to reduce the complexity of the conventional algorithm affected due to recursion. The proposed algorithm generated the sequence of frequent patterns and computes the support number using a reticular sequence tree which counts the numbers faster. It was identified that the performance of the enhanced Prefix Span outperforms based on space, time and other sensitivity factors [15].

A Vague Grid Sequence (VGS) based frequent pattern mining algorithm was proposed to overcome the drawbacks in Euclidean space and road networks such as difficulty in identifying appropriate positions. The vertical projection distance was considered to partition the zones into explicit and vague. The grid sequence in the vague zones was obtained by the transformation of its trajectories. The grid boundary problem was resolved in this approach using the vertical projection distance. The trajectory pattern mining algorithm

is an application to identify the migration of wild, to plan traffic, vegetation fragments, and recommending paths and so on [16, 17].

The spatio temporal patterns of the seismic dataset were used to explore the frequent items from the transactional database. A group of spatial data was collected and the continuous values are converted to discrete values to mine patterns in a more effective way. The exact patterns were matched using an algorithm namely adapted sequential pattern detection algorithm. The data sequences obtained are ensured using the min-support constraints [18].

In the pattern mining process, the negative and positive sequential patterns are present, in which the negative patterns play a critical role when compared to the negative patterns. The negative sequence patterns in the datasets are generally huge and hence it is difficult to mine the patterns from it [19, 20]. The exploration of negative sequence patterns was made easy by proposing the set theory based negative sequence patterns. The negative patterns were recognized using the set theory approach and it is converted to a positive containment problem. It reduces the time for rescanning the database helps in the utilization of positive sequence patterns. Hence, the available positive patterns were used to mine the negative patterns from a large dataset [21].

The networks namely, residual U-Net and Combined Differential Image network were integrated to detect the urban patters using Synthetic Aperture Radar (SAR) sensors. The conventional log ratio method results in poor compression ratio hence the difference in the images were computed using the weighted difference method, which is used to generate the weight function. The log ratio difference and singular decomposition were combined using neighbor constraints. Finally, residual U-Net was applied to detect the changes [22, 23].

A frequent pattern list is introduced to eliminate the invalid candidate for frequent patterns. The valid candidates are approved thereby reducing the processing time. In comparison with the 9DSPA miner approach, the frequent pattern list reduces the processing time [25]. Various denoising image enhancement and restoration techniques can be used in remote sensing images [26].The spatial overlay operation is integrated with Run Length Encoding for the efficient clustering of pixels. The prefix span algorithm eliminates the scanning of unnecessary pixels [27]. The spatial resolution of remote sensing images can be improved by super resolution technique [28].

3 Proposed methodology

In this section, the proposed Quantized ternary pattern-based identical spatial sequential prediction and Singular Value Decomposition (SVD)—Run Length Encoding (RLC) based clustering to analyze the vegetation fragmentation is explained in detail. The SRSI images obtained from the remote sensing satellites have high-intensity pixels that consist of information about the same area at different points of time. The pixel in an image, which is the elementary unit, has the value based on the color value in that particular area. Such pixels are considered as a sequence of events captured at various time intervals and can be termed as a spatial sequence. The proposed system focused to reduce the mining time by considering the non-repeated pixels alone by introducing the quantized ternary pattern. It avoids the encoding of repeating pixels which is a burden to the mining time. Hybridization of RLC with SVD localizes most of the energy content of the matrix into few singular values and also provided much less operation time.

The sequence of spatial data in the image is represented as follows

$$m \times n \times \sum_{j=0}^{N} P_n^j \qquad (1)$$

where N is the total number of images with 'm' height and 'n'width.

The images in the data set is represented in the sequential order as $imageset = \{I_1, I_2, I_3 \dots I_n\}$ from which the sequences that occur frequently are identified. The automatic threshold T_h is used to recognize the frequent items in the sequence, by which the image value greater than the threshold is considered as the frequent pattern. To apply a sequential pattern mining algorithm on the spatial dataset has to be converted to a table format that has pixel values. But, the conversion process involves both time and space complexities. The flow of the proposed system is shown in Fig. 1. The challenges have been solved by proposing

- Quantized ternary pattern based pixel grouping
- Singular Value Decomposition (SVD)—Run Length Coding based pattern mining

3.1 Quantized ternary pattern based Pixel Grouping

The pixel those have similar spatial sequences is grouped along with its neighbors using the quantized ternary patterns. The SRSI images are converted to images that have 7×7 dimension as follows
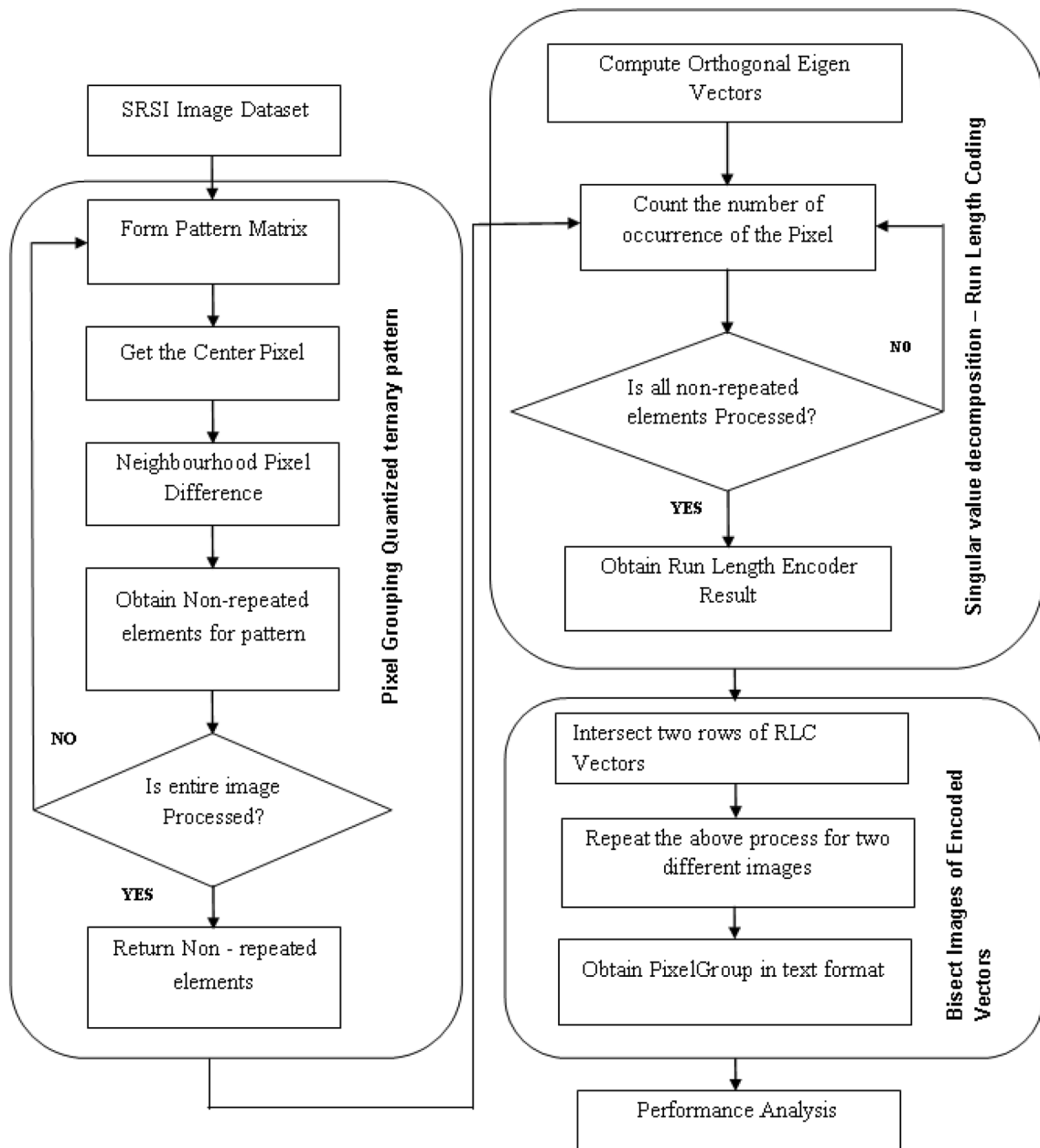
Fig. 1 Flow diagram for the proposed system

$$B_p = \lim_{x \to 1 \ to \ m-7} \ \lim_{y \to 1 \ to \ n-7} \left[I(x : x + 7, y : y + 7) \right] \qquad (2)$$

Here, 'x' and 'y' represents the row and column index for the pattern,'m' and 'n' denotes the number of rows and number of columns respectively. B_p is the pattern matrix which is the pattern formed from the image. The pixel value of the center point of the pattern matrix is obtained in which

the value of the 4th row and 4th column is considered as the center pixel. The neighborhood pixels are subtracted from the center pixel to predict the non-repeating values in the image. The process of identifying the non-repeated pixel values is continued for the entire image using an automatic threshold approach. For each block in the image, a unique value $U_j = unique\left(d_j\right)$ is obtained as the non-repeating pixel.

Algorithm 1: Quantized ternary pattern

Input: input image I

Output: returned non repeated pixels, P_i

Step1: Convert the input image into the pattern with 7*7 dimension,

Step 2: Get the center pixel value of the formed pattern matrix,

$$I_C = B_p(4,4)$$

Step 3: For each neighborhood pixel, calculate the difference between the estimated center pixel and neighborhood pixels,

$$for\ j = 1\ to\ N_p$$

$$d_j = |B_p(j) - I_C|$$

$$end\ for$$

Step 4: Predict the non-repeated value for the entire image by computing the unique elements

$$U_j = unique(d_j)$$

Step 5: Obtain the non-repeated pixel value for a particular block

$$P_i = U_j + I_C$$

Step 5: Repeat step 1 to step 5 until the entire image processing.

3.2 Singular value decomposition (SVD)—Run length coding (RLC) based pattern mining

The RLC is the image compression technique that processes the image pixel by pixel and hence increases the time complexity. The SVD—RLC is applied to reduce the time complexity of encoding the pixels, which obtains a single value for processing by decomposing the neighbor pixel values. The images in Fig. 2a and b show the input SRSI image and the Singular Value Decomposed image obtained from the datasets respectively. The input image that has unique pixels is processed using SVD-RLC based clustering that results in encoded vectors. The SVD is applied for each and every row and columns in an image. An orthogonal Eigenvector of the image is computed and multiplied with the image. Further, the RLC process is carried out for the SVD decomposed image. The intensity of each pixel starting from 0 to ith position is obtained for all the unique values. The frequency of occurrences of the pixel with the same intensity is identified and the number of occurrences is counted. The process is repeated for each row of the image to get the encoded image vector.

Algorithm 2: Singular value decomposition – Run Length Coding vector

Input : input image I , unique pixels P_i

Output: encoded vectors Enc_{vec}

Step1: Compute the size of the image in terms of rows and columns $[m\ n] = size(I)$

Step 2: Compute the orthogonal Eigenvector $IA = V * I$

Step 3: Perform RLC, for each unique pixel values, get the intensity values

$$for\ p = 0\ to\ P_i$$

$$I_{int} = P_i^p$$

end for

Step 4: Check the number of occurrence of the corresponding pixel by initializing the count to 1 and traversing through the image pixels in terms of rows and columns.

$$cnt = 1$$

$$for\ i = 0\ to\ m-1$$

$$for\ j = 0\ to\ n-1$$

$$if\ (I_{int} == getPixelColour(IA_{i,j}))$$

$$cnt = cnt + 1$$

$$end\ if$$

$$Enc_{vec}.add(line)$$

$$end\ for$$

$$end\ for$$

Step 4: Save the resultant of the occurrence of the corresponding pixel for each row, repeat the above process for the next row.

Step 5: Return Enc_{vec} for the entire image.

3.3 Bisect images for encoded vectors

The image bisecting process gets the encoded vectors as input and returns the grouped pixels in a text file. The length of the encoded vector is initialized to intersect two rows of two different images. The number of pixels in both the images is initialized and represented as num^1 and num^2 respectively. Compare both the images pixel by pixel and perform the intersection operation. The sequence from the images is added and a new row with items along with their sequences is formed, which return a new row of pixels by which the current row is replaced. The encoded vectors are added to the new sequence to return the grouped pixels using Group prefix span algorithm

---

Algorithm 3: Bisect images for encoded vectors

Input: Encoded vectors Enc_{vec}

Output: text file contains $pixelGroup$

Initialize the length of encoded vectors and intersect two rows of two different images

$k = length[Enc_{vec}]$

$currentRow = Enc_{vec}[0].line$

$for\ a = 0 : k - 1$

$\quad newRow_i = intersectRows(currentRow, Enc_{vec}[a].line)$

$num^1 = row^1.items$

$num^2 = row^2.items$

$for\ g = 0 : num^1 - 1$

$item^1 = row^1.items[g]$

$while\ (h < num^2)$

$newRow.add = \begin{cases} makeNewRow(item^1.sequence + item^2.sequence, num^1) & if\ num^1 \le num^2 \\ makeNewRow(item^1.sequence + item^2.sequence, num^2) & elseif\ num^2 \le num^1 \end{cases}$

$\quad\quad increment\ h{+}{+}$

$end\ while$

$end\ for$

Return $newRow$

$currentRow_i = newRow_i$

$end\ for$

Return $pixelGroup = newRow_i$

---

4 Results and discussion

In this section, the proposed SVD pixel grouping based clustering algorithm is validated regarding time consumption and mining time (Table 1). The Cropland data layer dataset is utilized to measure the performance of the algorithm. The data set is obtained using the link (https://nassgeodata.gmu.edu/ Crop Scape/) [29]. The Table 2 shows the dataset details. The dataset consists of geospatial raster images that are captured by a satellite with a reasonable resolution. The images cover the continental area of the United States, namely, Iowa State. It covers over 150,000 square kilometers, in which each pixel in the image covers about 900 square meters. The different crops cultivated in the region are identified using the colors on the map. The datasets are segregated as four datasets, namely D1, D2, and D3 and D4 based on the period of time in which it is captured. The comparison of existing methodologies is given in Table 3 and performance analysis of the proposed method is explained. Figures 3 and 4 shows a comparison of mining time for existing algorithms namely, prefix span, group prefix span, GFS pattern and group prefix span AC with the proposed SVD pixel grouping algorithm. The values in Tables 4 and 5 are plotted as graphs in Figs. 3 and 4 as a representation of graphical visualization. It is depicted from Fig. 3 that the SVD pixel grouping algorithm outperformed the Prefix span and group prefix span algorithm. The mining time of the prefix span is high due to its inability to project the database in a vertical manner. The support values of the sequence patterns generated by different sequential pattern mining algorithms are compared in Fig. 5. The plot in Fig. 5 is an interpretation of values in Table 6, which shows that the

---

Table 1 Parameter settings

Parameter	Parameter description	Value range
δ_{max}	Advance reserve space	220 to 255 (pixel variation)
Tol	Tolerance constant	0.000227
Err	Error estimated parameter at each loop	0 to 0.5
V	orthogonal eigenvectors	-1 to $+1$
S	Singular values	-569 to $+569$
'x' and 'y'	Row and column index of the pattern	0 to 1024 (depends upon image size)
'm' and 'n'	Number of rows and columns of the pattern	1024 and 1024 (depends upon image size)
cnt	Chosen intensity count	0 to 255 (maximum limit)
D	The difference for all neighborhood pixels	0 to $+5$ (depends upon pattern)
B_p	Pattern matrix	3 * 3 dimension
Pi	Vector of unique values of pattern pixels	4 to 8 dimension and range depends upon intensity region
EncVec	Encoded vector	Vector of the count of occurrence (50 to 100)
num	Number of items of rows	1024 (depends upon image height)
newRow	New row appended	Depends upon the number of rows

Table 2 Dataset details

Dataset ID	Region	Pixels count	Time duration	Volume of data
Dataset-D1	Small portion in butler county	21* 132	From 2010 to 2016	1.7 MB
Dataset-D2	Region of butler county	1323 * 1348	From 2010 to 2016	20 MB
Dataset-D3	ASD 1910 region	5971 * 3534	From 2010 to 2016	141 MB
Dataset-D4	Region of state of IOWA	17,795 * 11,671	From 2010 to 2016	594 MB

Table 3 comparison with existing methodologies

Reference	Methodology used	Advantage	Disadvantage
[8]	Dictionarybased Compression	The benefit of having access to all of the message	Store a large part of the message in memory for processing
[30]	Huffman code stream	More efficient (uses less bits)	Harder to encode and decode
[31]	Neural image compression (NIC)	Performed more accuratein latent space giga pixel application	Losses high level information when suppressing pixel level noise
[32]	Fractal dimension transformation	Consistent and well quantitative analysis	Distortion occurred when compression
[33]	JPEG steganographic scheme	High ability to face some signal processing or noise	Attacker can simply reorders the color vectors in the palette

proposed SVD group prefix span generated more accurate sequence patterns. The prefix span used to skip more number of rows during scanning; hence it fails to generate the patterns with required support values. In comparison with the prefix span and group prefix span algorithm, the SVD group prefix span generated the support values greater than the threshold value. In general, the SRSI images are utilized to analyze the growth and development of a particular area over time. It is realized from the experiment that the threshold value plays an important role, where more patterns can be gained with low threshold values. These patterns are utilized in the prediction of crop yields.

4.1 Performance analysis

The proposed Quantized ternary pattern-based identical spatial sequential prediction and Singular Value Decomposition (SVD)—Run Length Encoding (RLC) based Pattern mining is compared with some related methodologies and analyzed in terms of memory storage, encoding and decoding information, Compression and Security. Table 3 compares the existing related methods were compression can be processed using SVD without using large memory storage. Encoding and Decoding can be easily done by sequential operation. We retain the high level of information when non repeated pixel based approach is performed. The structure and texture of information are

not lost while compression by SVD-RLC algorithm. Here pattern by pattern pixel grouping is performed so attackers cannot reorder the color of the vector in the palate. The idea begins this is to propose an efficient algorithm that can compress Remote sensing image set by SVD Group prefix span which can be applied in big datasets in terms of time and to overcome the challenges in SRSI images.

5 Conclusion

Quantized ternary pattern based frequent spatial sequence and Singular Value Decomposition (SVD)—Run Length Coding (RLC) based pattern mining along with Group PrefixSpan for pixel grouping is proposed to overcome the challenges in SRSI images such as computational overhead due to huge volume of images. The images are compressed by QTP based SVD-RLC Quantized ternary pattern algorithm, which reduces the overhead by processing the images in a block by block manner. The Group prefix sequence pattern outperformed the prefix grouping by scanning the rows without any skips. The algorithms are validated using Cropland data layer dataset and proved that it results in less mining time and generates appropriate sequence patterns above the threshold support values. The proposed method is highly efficient, as it compresses remote sensing images set by the concept of the SVD
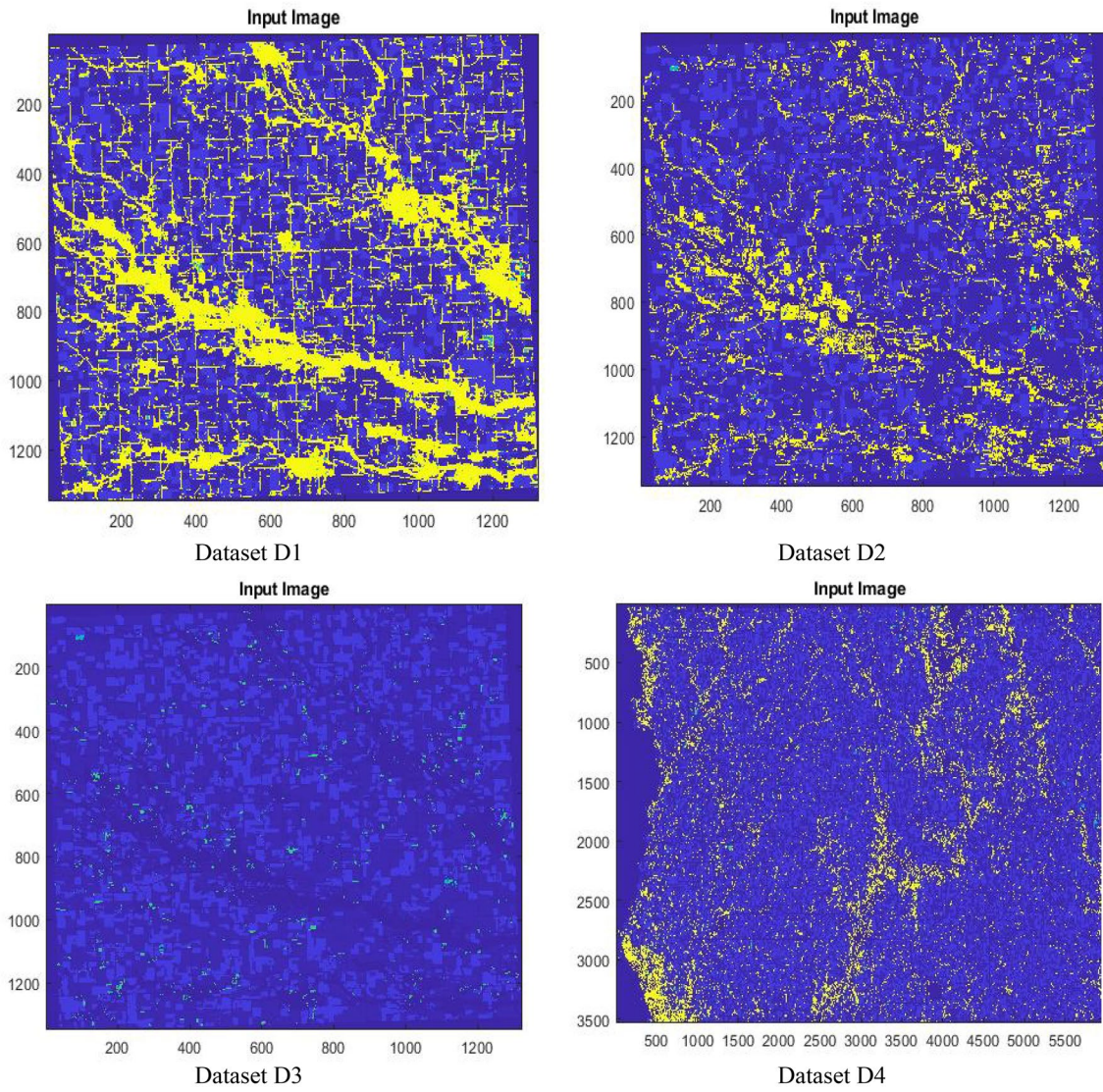
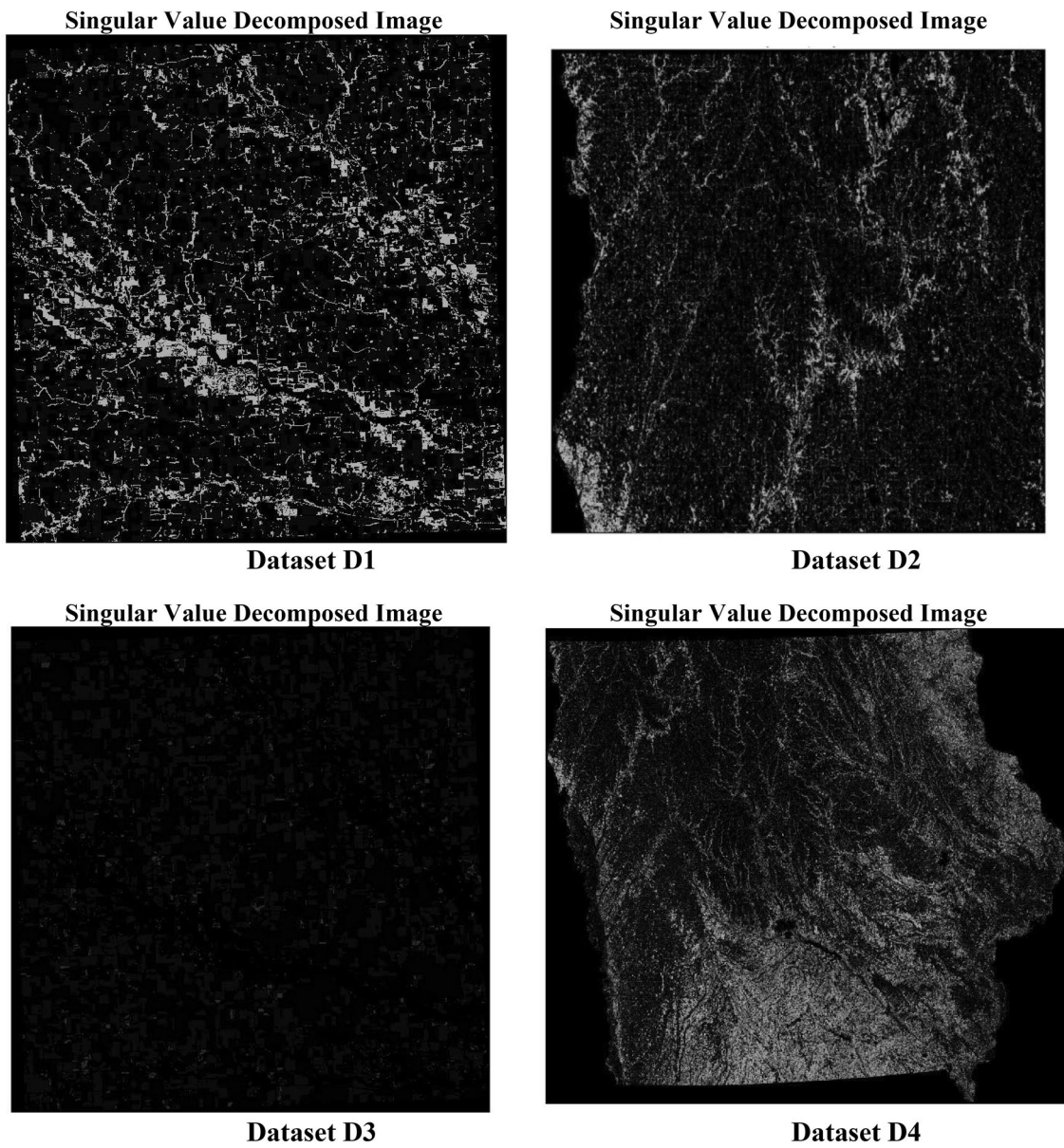(a) Input images of Dataset D1, D2, D3 and D4

Fig. 2 **a** Input images of Dataset D1, D2, D3 and D4 **b** Singular Value Decomposed images of Datasets D1, D2, D3 and D4

Group Prefix Span even in the case of big datasets. The method can generate all sequential patterns at reasonable times.

In the future, the mining time and compression ratio can be enhanced using the fraction based run length encoding technique. In addition, the computational overhead in processing the SRSI images can be reduced using compression based classification.
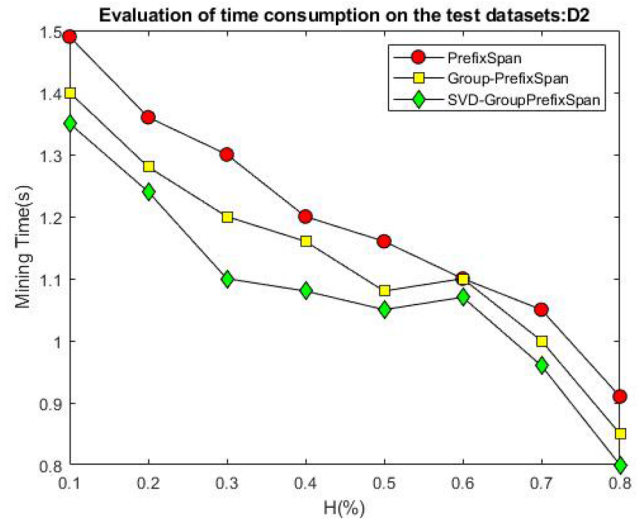
Singular Value Decomposed Image

Singular Value Decomposed Image

Dataset D1

Dataset D2

Singular Value Decomposed Image

Singular Value Decomposed Image

Dataset D3

Dataset D4

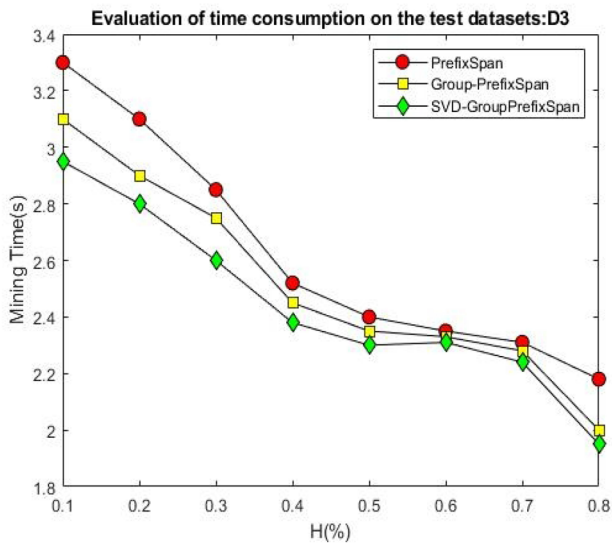(b) Singular Value Decomposed images of Datasets D1, D2, D3 and D4
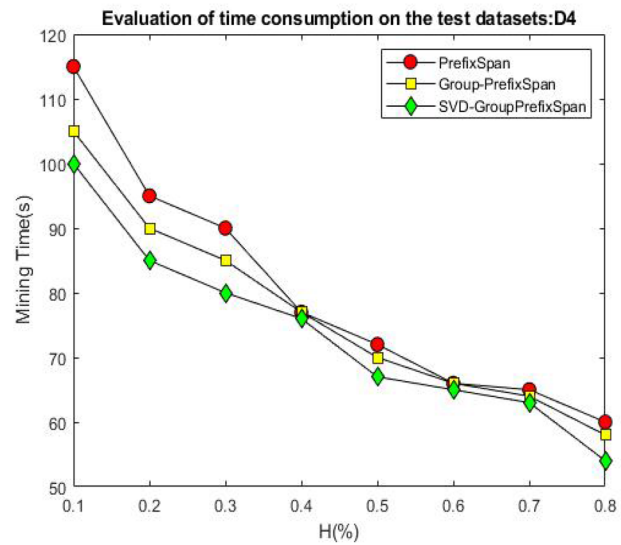
Fig. 2 (continued)
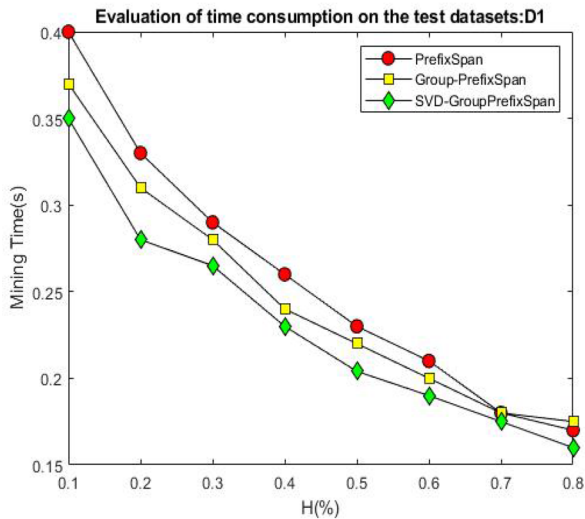
(a) Evaluation of time consumption on test dataset **(5.4% decreased)**

(b) Evaluation of time consumption on test dataset **(3.57% decreased)**

(c) Evaluation of time consumption on test dataset **(4.83% decreased)**

(d) Evaluation of time consumption on test dataset **(3.86% decreased)**

Fig. 3 **a** Evaluation of time consumption on test dataset (3.57% decreased) **b** Evaluation of time consumption on test dataset (5.4% decreased) **c** Evaluat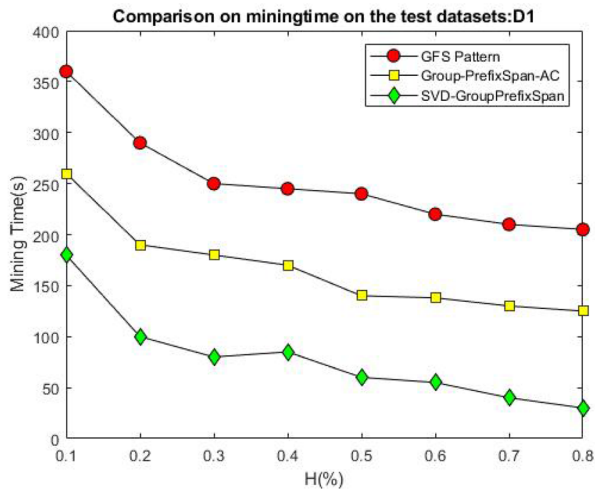ion of time consumption on test dataset (4.83% decreased) **d** Evaluation of time consumption on test dataset (3.86% decreased)

Table 4 Mining time of prefix span, group-prefix span and SVD group prefix span algorithm based on the threshold values

DATASET	Threshold H (%)	Prefixspan (Sec)	Group-prefixspan (Sec)	SVD-group prefixspan (Sec)
Dataset D1	0.1	0.4	0.37	0.35
	0.2	0.33	0.31	0.28
	0.3	0.29	0.28	0.265
	0.4	0.26	0.24	0.23
	0.5	0.23	0.22	0.204
	0.6	0.21	0.2	0.19
	0.7	0.18	0.18	0.175
	0.8	0.17	0.175	0.16
Dataset D2	0.1	1.49	1.40	1.35
	0.2	1.36	1.28	1.24
	0.3	1.30	1.20	1.10
	0.4	1.20	1.16	1.08
	0.5	1.16	1.08	1.05
	0.6	1.10	1.10	1.07
	0.7	1.05	1.00	0.96
	0.8	0.91	0.85	0.80
Dataset D3	0.1	3.30	3.10	2.95
	0.2	3.10	2.90	2.80
	0.3	2.85	2.75	2.60
	0.4	2.52	2.45	2.38
	0.5	2.40	2.35	2.30
	0.6	2.35	2.33	2.31
	0.7	2.31	2.28	2.24
	0.8	2.18	2.00	1.95
Dataset D4	0.1	115	105	100
	0.2	95	90	85
	0.3	90	85	80
	0.4	77	77	76
	0.5	72	70	67
	0.6	66	66	65
	0.7	65	64	63
	0.8	60	58	54

**Table 5** Mining time of GFS patten, group prefix span—AC and SVD grouping algorithm based on the threshold values

| DATASET	Threshold H (%)	GFS Pattern (Sec)	Group-PrefixSpan-AC (Sec)	SVD-Group PrefixSpan (Sec)
Dataset D1	0.1	360	260	180
	0.2	290	190	100
	0.3	250	180	80
	0.4	245	170	85
	0.5	240	140	60
	0.6	220	138	55
	0.7	210	130	40
	0.8	205	125	30
Dataset D2	0.1	340	245	150
	0.2	265	180	110
	0.3	230	160	90
	0.4	225	150	70
	0.5	210	140	60
	0.6	160	130	55
	0.7	155	125	40
	0.8	150	105	30
Dataset D3	0.1	300	175	90
	0.2	245	150	75
	0.3	200	130	70
	0.4	195	125	60
	0.5	190	120	50
	0.6	140	115	40
	0.7	135	112	35
	0.8	130	105	25
Dataset D4	0.1	235	170	140
	0.2	225	145	105
	0.3	180	130	86
	0.4	185	125	75
	0.5	170	120	70
	0.6	130	110	60
	0.7	125	95	45
	0.8	120	93	41

(a) Evaluation of Mining Time on test dataset **(29.41% decreased)**
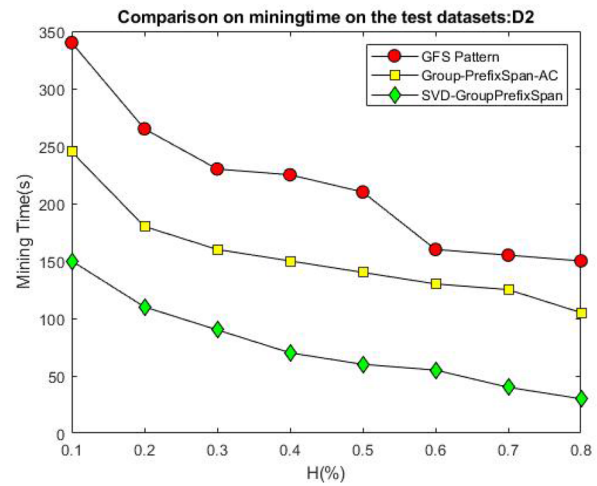
(b) Evaluation of Mining Time on test dataset **(40% decreased)**
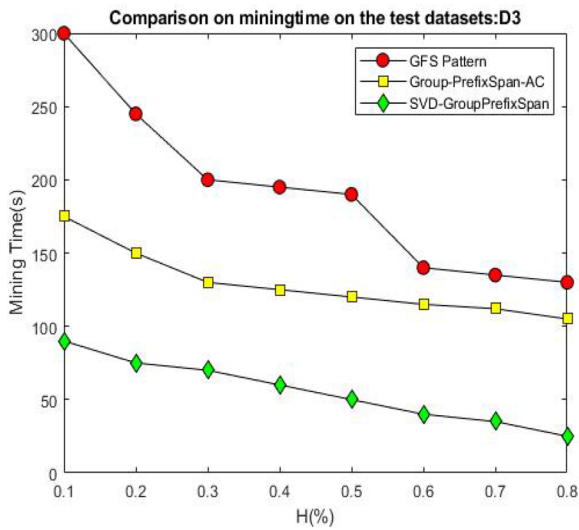
(c) Evaluation of Mining Time on test dataset **(48.57% decreased)**
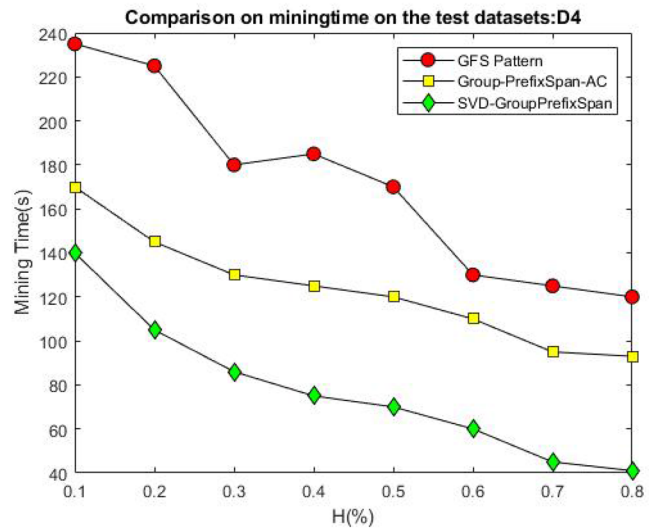
(d) Evaluation of Mining Time of test dataset **(45.67% decreased)**

Fig. 4 **a** Evaluation of mining time on test **b** Evaluation of mining time on test dataset (29.41% decreased) **c** Evaluation of mining time on test (48.57% decreased) **d** Evaluation of mining time of test dataset dataset (45.67% decreased)
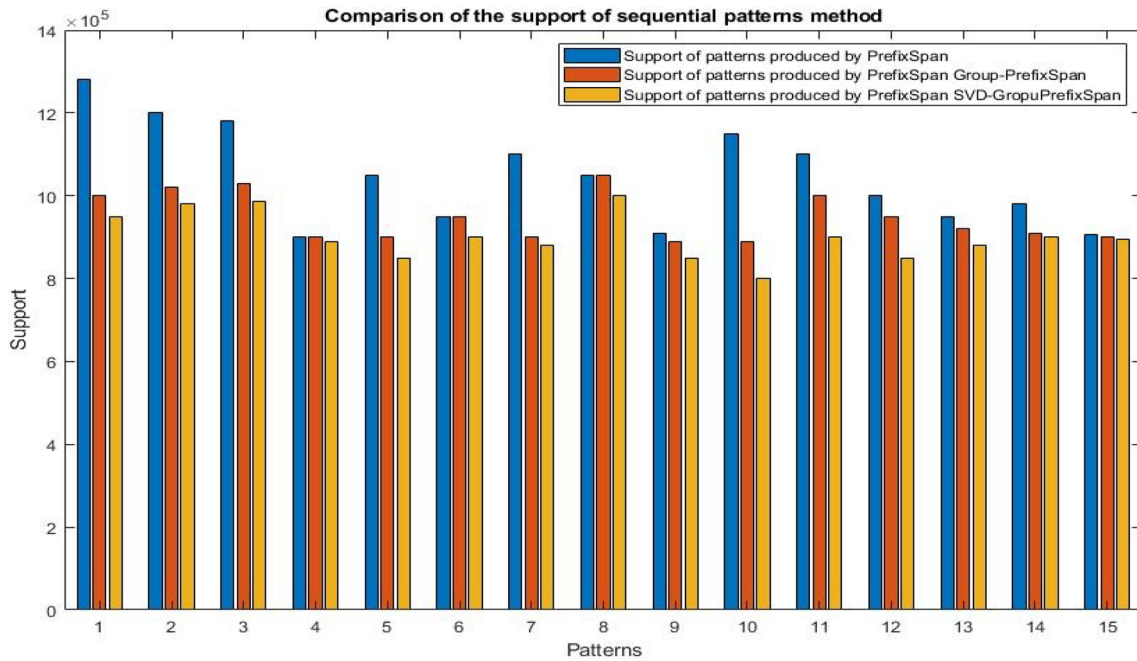
Fig. 5 Comparison of the support values of sequential pattern mining methods

Table 6 Support values of sequential pattern methods for each pattern

Pattern	Prefixspan (Km^2)	Prefixspan group-prefixspan (Km^2)	Prefixspan SVD-group prefixspan (Km^2)
P1	1,280,000	1,000,000	950,000
P2	1,200,000	1,020,000	980,000
P3	1,180,000	1,030,000	985,000
P4	900,000	900,000	890,000
P5	1,050,000	900,000	850,000
P6	950,000	950,000	900,000
P7	1,100,000	900,000	880,000
P8	1,050,000	1,050,000	1,000,000
P9	910,000	890,000	850,000
P10	1,150,000	890,000	800,000
P11	1,100,000	1,000,000	900,000
P12	1,000,000	950,000	850,000
P13	950,000	920,000	880,000
P14	980,000	908,760	900,000
P15	905,000	900,000	895,000

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Wylie BK, Pastick NJ, Picotte JJ, Deering CA (2019) Geospatial data mining for digital raster mapping. GISci Remote Sens 56(3):406–429
2. Gan, W., Lin, J.C.W., Fournier-Viger, P., Chao, H.C. and Yu, P.S., 2018. A survey of parallel sequential pattern mining. arXiv preprint. arXiv:1805.10515
3. He Z, Zhang S, Wu J (2019) Significance-based discriminative sequential pattern mining. Expert Syst Appl 122:54–64
4. Xue Y, Li T, Liu Z, Pang C, Li M, Liao Z, Hu X (2018) A new approach for the deep order preserving submatrix problem based on sequential pattern mining. Int J Mach Learn Cybernet 9(2):263–279
5. Cai G, Lee K, Lee I (2018) Itinerary recommender system with semantic trajectory pattern mining from geo-tagged photos. Expert Syst Appl 94:32–40
6. Atluri G, Karpatne A, Kumar V (2018) Spatio-temporal data mining: a survey of problems and methods. ACM Comput Surveys (CSUR) 51(4):83

7. Fukui KI, Okada Y, Satoh K, Numao M (2019) Cluster sequence mining from event sequence data and its application to damage correlation analysis. Knowl-Based Syst 179:136–144

8. Zhang Z, He Q, Tong H, Gou J, Li X (2016) Spatial-temporal traffic flow pattern identification and anomaly detection with dictionary-based compression theory in a large-scale urban network. Transp Res Part C Emerg Technol 71:284–302

9. Kowe P, Mutanga O, Odindi J, Dube T (2019) Exploring the spatial patterns of vegetation fragmentation using local spatial autocorrelation indices. J Appl Remote Sens 13(2):024523

10. Lyu X, Ma H (2019) An efficient incremental mining algorithm for discovering sequential pattern in wireless sensor network environments. Sensors 19(1):29

11. Chen Y, Yuan P, Qiu M, Pi D (2019) An indoor trajectory frequent pattern mining algorithm based on vague grid sequence. Expert Syst Appl 118:614–624

12. Campisano, R., Porto, F., Pacitti, E., Masseglia, F. and Ogasawara, E.S., 2016, October. Spatial Sequential Pattern Mining for Seismic Data. In SBBD, (pp. 241–246).

13. Dong, X., Zheng, Z., Cao, L., Zhao, Y., Zhang, C., Li, J., Wei, W. and Ou, Y., 2011, October. e-NSP: efficient negative sequential pattern mining based on identified positive patterns without database rescanning. In: Proceedings of the 20th ACM international conference on Information and knowledge management, (pp. 825–830). ACM.

14. Fan Y, Ye Y, Chen L (2016) Malicious sequential pattern mining for automatic malware detection. Expert Syst Appl 52:16–25

15. Su L, Gong M, Zhang P, Zhang M, Liu J, Yang H (2017) Deep learning and mapping based ternary change detection for information unbalanced images. Pattern Recogn 66:213–228

16. Li L, Wang C, Zhang H, Zhang B, Wu F (2019) Urban building change detection in SAR images using combined differential image and residual U-net network. Remote Sens 11(9):1091

17. Häberle M, Werner M, Zhu XX (2019) Geo-spatial text-mining from Twitter–a feature space analysis with a view toward building classification in urban regions. Euro J Remote Sens 52:1–10

18. Wang S, Eick CF (2018) A data mining framework for environmental and geo-spatial data analysis. Int J Data Sci Anal 5(2–3):83–98

19. Demattê JAM, Fongaro CT, Rizzo R, Safanelli JL (2018) Geospatial Soil Sensing System (GEOS3): a powerful data mining procedure to retrieve soil spectral reflectance from satellite images. Remote Sens Environ 212:161–175

20. Koranteng MK, Adu-Asare A (2018) Geospatial assessment of vegetation changes around the odublasi quarry in ghana. West African J Appl Ecol 26:73–86

21. Tiwari P, Shukla PK (2019) A review on various features and techniques of crop yield prediction using geo-spatial data. Int J Organ Collect Intell (IJOCI) 9(1):37–50

22. Moomen A, Bertolotto M, Lacroix P, Jensen D (2019) Inadequate adaptation of geospatial information for sustainable mining towards agenda 2030 sustainable development goals. J Clean Prod 238:117954

23. Curran MF, Cox SE, Robinson TJ, Robertson BL, Rogers KJ, Sherman ZA, Adams TA, Strom CF, Stahl PD (2019) Spatially balanced sampling and ground-level imagery for vegetation monitoring on reclaimed well pads. Restor Ecol 27(5):974–980

24. Silveira EMDO, Espírito-Santo FDB, Acerbi-Júnior FW, Galvão LS, Withey KD, Blackburn GA, de Mello JM, Shimabukuro YE, Domingues T, Scolforo JRS (2019) Reducing the effects of vegetation phenology on change detection in tropical seasonal biomes. GIScie Remote Sens 56(5):699–717

25. Chang Y-I et al (2019) Mining image frequent patterns based on a frequent pattern list in image databases. J Supercomput 76:2597–2621

26. Sowmya V et al (2019) recent advances in computer vision. Springer, Cham, pp 401–424

27. Wu X, Zhang X (2019) An efficient pixel clustering-based method for mining spatial sequential patterns from serial remote sensing images. Comput Geosci 124:128–139

28. Zhang N et al (2020) An unsupervised remote sensing single-image super-resolution method based on generative adversarial network. IEEE Access 8:29027–29039

29. https://catalog.data.gov/dataset/cropscape-cropland-data-layer. Accessed 12 Sept 2019

30. Artusi A et al (2019) Overview and evaluation of the JPEG XT HDR image compression standard. J Real-Time Image Process 16(2):413–428

31. Tellez, David, et al. (2019) Neural image compression for gigapixel histopathology image analysis. IEEE Trans Pattern Anal Mach Intell PP(99):1–1

32. Chen Z, Ye Hu, Zhang Y (2019) Effects of compression on remote sensing image classification based on fractal analysis. IEEE Trans Geosci Remote Sens 57(7):4577–4590

33. Li, Fengyong, et al. (2020) Anti-compression JPEG steganography over repetitive compression networks. Signal Proc 170:107454

34. Yousefi Valandar M et al (2019) A blind and robust color images watermarking method based on block transform and secured by modified 3-dimensional Henon map. Soft Computing 24:771–794

35. Jafari Barani M et al (2019) A blind video watermarking algorithm robust to lossy video compression attacks based on generalized Newton complex map and contourlet transform. Multimed Tools Appl 79:2127–2159

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.