International Conference on Substantial Development in the field of Engineering and Management
Mahratta Chamber of Commerce, Industries and Agriculture Tilak Road, Pune (India)
29th December 2019 www.conferenceworld.in ISBN : 978-81-943584-8-0

# Sequential Pattern Discovery using Equivalence Classes (SPADE) algorithm for Clustering Digital Images sequentially

**S Hakkim[1] , Dr.K.Nirmala[2]**

[1]*Research Scholar, Quaid-e-Millath College for Women (Autonomous) Chennai*

[2]*Associate professor, Quaid-e-millath Govt. College for women (Autonomous), Chennai*

**Abstract**

*In this paper made unsequenced images during a sequenced form with the assistance of SPADE algorithm. In this first we have converted a video into images form. A video can contain thousand images in it. So first we detected which type of image is blurred or unsequenced. So that we can correct it with the help of sequencing algorithm. Every image has its feature extraction by which we will detect the standard of that image. So you can see that this paper is a combination of image processing and data mining , so that you can call it image mining .At last we have found out the graph for CPU time utilization and image forms.so at last we have distortion less image. In this type of project the biggest task is that in a pair of thousand images how you can find which image is distorted or fade and so as to simplify it we are using the sequencing algorithm. In this paper, Extracting the Features of image frames and Identifying the image , to find the sequence which part of the image frame we put sequential pattern mining, By using the sequential pattern mining then we find the sequence of frames and we make a sequence of frames.*

***Keyword: SPADE, CPU, Sequenced, UnSequenced, Noise, Denoise***

## 1. Introduction

In our lifestyle, each folks in particular receives processes, and scrutiny a huge quantity of varied sorts of in sequence. Any choice has supported this scrutiny. Image is an incredibly dominant medium which represents the knowledge and process during a packed in and well-organized way. it's not only major source of in sequence, but is additionally used for communication among people and for interaction between humans and machines.

Common digital images contain huge amount of data in very small space. a picture can take and send by using any communication average during a small number of seconds contains the maximum amount information. This is often why there's an important need for repeated and influential image analysis methods. Analysis and understanding of a picture is that the input difficulties in many request areas like remote sensing, astronomy, medical imaging, etc. The

digital image of picture isn't same as actual; it's such a lot dreadful conditions by any mean like color, focus, lightening, etc.

Segmentation is that the key steps before performing other operations like explanation, acknowledgment, view considerate, indexing etc. Image segmentation is that the classification of homogeneous regions within the image. The target of image segmentation is to panel a picture into regions and decay an opportunity into its mechanism. This is often skilled by segmenting a picture into subsets and later assigning the individual images to classes. The mixture of segmentation-based approaches is proposed to precise the thing and its spatial information. Some approaches perform complex segmentation, either mechanically or manually. the matter of those approaches is that, if the query image segmentation result's not accurate enough, the retrieval performance are going to be affected. Segmentation is performed to acknowledge objects from the image using features.

Features are often intensity value, edge, corners, texture, shape etc. generally, segmentation processes are supported one among two basic properties of the intensity values: Discontinuity and Similarity. Within the first category, Segmentation algorithms for monochrome images generally are supported of two basic properties of image intensity values: discontinuity and similarity. Within the first category, the approach is to partition a picture supported abrupt change in intensity, like edge. The principal approaches within the second category are supported portioning.

**Sequential mining:** Today, many companies have already used computing system as data storage transaction recording, and reporting. Processing in small scale are often done by using simple database or spreadsheet e.g Microsoft Excel. Report which is made from that application is enough for analyzing marketplace for deciding. However, for giant company which sell products in large scale, which consists of hundred or thousand quite product and selling type, that application are hardly to manage. There are often a missing knowledge from those data, which is critical for deciding, for instance the pattern of the customer’s purchasing. Sequential pattern works by identifying or analyzing all the sequences that always appears on an item (certain paint) purchased by the customer. With the info mining of sequential data on the acquisition of paint, it'll produce knowledge for paint sales.

Knowledge are often useful for companies to get information on any paint if purchased simultaneously and paint what is going to be purchased during a sequence in order that it can generate relationships among items also as what proportion paint is purchased during a sequence that actually different .Basically data processing is closely associated with data analysis and use of software to seek out patterns and similarities in data collection. Retrieve valuable information which is completely unexpected to extract patterns is an unseen pattern.

**Image Segmentation**: Clustering may be a classification technique Let N be the Vector of measurements describing each image or group of images (i.e., region) in a picture, a similarity of the measurement vectors and thus their clustering within the N-dimensional measurement space implies similarity of the corresponding images or image groups. Therefore, clustering in measurement space could also be an indicator of similarity of image regions, and should be used for segmentation purposes.

## 2. Literature Survey

Hemalatha & Devasena (2011) proposed a research to find out the accurate images while mining an image (multimedia) database and developed an innovative technique for mining images by means of LIM dependent image matching method with neural networks. This approach is independent of several parameters setting to produce a robust solution. It is developed and implemented on MATLAB and is investigated with the images of several databases. Suitable measures were developed to estimate the performance of the system. The performances of the LIM dependent image matching method results were significant and comparable.

Victor & Peter (2010) put forth a new minimum spanning tree based clustering algorithm for image mining. The minimum spanning tree clustering algorithm is proficient of detecting clusters with irregular boundaries. The author presented a minimum spanning tree depending on the clustering technique using weighted Euclidean distance for edges, which is vital constituent in constructing the graph from image. The technique constructs ‘k’ clusters with segments. This approach is very much capable of protecting detail in low variability image regions while not considering detail in high variability regions which is the main advantage of this approach. This approach has handled the problems of undesired clustering structure and redundant huge number of clusters.

Sanjay et al. (2007) put forth an image mining technique using wavelet transform. The author proposed an image mining approach using wavelet transform. It uses common pattern identical, pattern identification and data mining models with the intention that a real life scene/image can be

associated to a particular category, assisting in different prediction and forecasting mechanisms. It is a three-step procedure i.e. image gathering, learning and classification. Since wavelet transform uses time frequency association, it can be utilized for image mining as a substitute of Fourier transform. Wavelet transform is utilized to decompose an image into dissimilar frequency sub bands and a small frequency sub band is used for Principal Component Analysis (PCA). Classification assists in recognizing the category to which an image relates with. They have constructed a prototype system for identification using DWT + PCA system. The conception of image mining as a consequence can be competently used for weather forecasting so that one can know the natural disasters that may occur in advance.

Effective research in the field of image retrieval and mining has turned out to be a significant research area because of significant applications in digital image databases. At present, a huge segment of information is in image form; it is necessary and certainly there is a significant requirement to search for images by means of content. Image mining has a wide range of applications in different sectors like medical diagnosis, space research, biology, remote sensing, etc.

## 3. Proposed Work

Image ARM cares with the appliance of ARM techniques to image sets. There are many reasons why we might need to analyse collections of images. Common example application areas where data processing techniques are applied to image sets include medical analysis, meteorology and oceanography. These application areas are addressed during a number of various manners but all include the recasting of the image set into a structured form which will facilitate data processing using established processes (in many cases the representation includes metadata). The challenge of applying ARM to image data is thus the transposing of the image data into a form that;

a) Attribute format is allowed and

b) Number of attributes are filtered by the manageable size.

The image analysis and retrieval community have undertaken a big amount of labor during this area and established a body of labor which the project team are going to be ready to draw on. during this context it's worth noting that the wants for image mining and image retrieval aren't identical. within the case of image mining where wish to supply a classifier for a limited number of predefined classes or cluster into a finite set of groups it's conjectured here that a way coarser image representation will suffice. Additionally it's worth remembering here that data processing can and is meant to figure with noisy data and thus the representations are often relatively crude.

Data mining techniques are used for a few times to get implicit knowledge in transaction databases. Especially, methods are available for determining the interesting associations among item sets over large numbers of transactions, like the products that are most often purchased together in market basket analysis. Achieving similar success with multimedia datasets remains a challenge, however, not only thanks to the dimensions and complexity of image and video data, but also the shortage of image equivalents for the association rule components, namely the things, the item sets, and even the principles. it's not straightforward to define, including detect, the things and item sets appropriate for locating the implicit spatial

In this proposed work, are trying to make our result better as the previous in combination with mining and processing. It gives better results in sequencing of image. Our basic concept is the find the feature extraction then finding the clusters of image with K-mean and after that finding the sequence of the image. In this we provide noisy image and finding the sequence of images. Sequence mining is a type of structured data mining in which the dataset and administrator look for sequences or trends in the data. This data mining is split into two fields. Sequence mining is different from regular trend mining, because the info are more specific, which makes building an efficient dataset difficult for database designers, and it can sometimes go away if the sequence is any different from the common sequence. Spade Algorithm for the sequencing of the data frames of video image are shown in Figure 3.1 below.

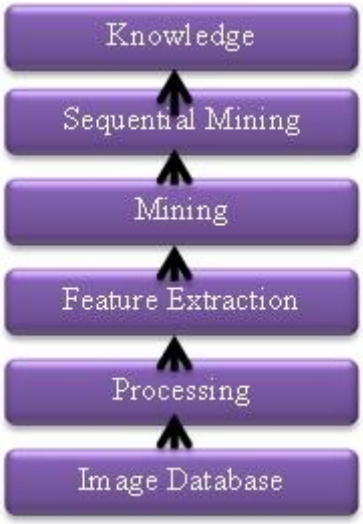



Figure 3.1 Process of SPADE Algorithm

**4. Algorithm**

The proposed algorithms are as follows

**Spade Algorithm**

ISBN : 978-81-943584-8-0

**Step 1:** Make the first pass over the sequence dataset D to yield all the 1-element frequent sequences

**Step 2:** Repeat until no new frequent sequences are found

**Step 3:** Merge pairs of frequent frames subsequences found in the (k-1) th pass to generate candidate sequences that contain k items

**Step 4:** Prune candidate k-sequences that contain infrequent (k-1)- subsequences

**Step 5:** Make a new pass over the sequence dataset D to find the support for these candidate sequences of the frames

**Step 6:** Eliminate candidate k-sequences whose actual support is less than minseq

**Proposed Sample codes for spade algorithm are**

A. Initialize the Video Image

B. Now convert the Video into the Frames

C. Finding the feature s of image through feature extraction

D. After it we also find the segment of the image so that it is easy to find the sequence of the colour image.

E. Now apply Sequence through the sequence mining algorithm

**F. Spade Process**

a. F1= set of frequent event images

b. F2= set of frequent 2- event-long sequence

c. For all equivalence frames {P i] €F1 in descending order

d. E2 =[Pi];

e. For (k=3;Ek-1!=Ø; K++)

f. For all classes [e] €E k-1

g. N=process_class([e]);

h. If(N!= Ø)

i. Ek=EkÙN;

j. Delete [e];

## 5. Implementation

This result shows the sequence of the image which is given by the image through example table 5.1 , table 5.2 as shown below

ISBN : 978-81-943584-8-0

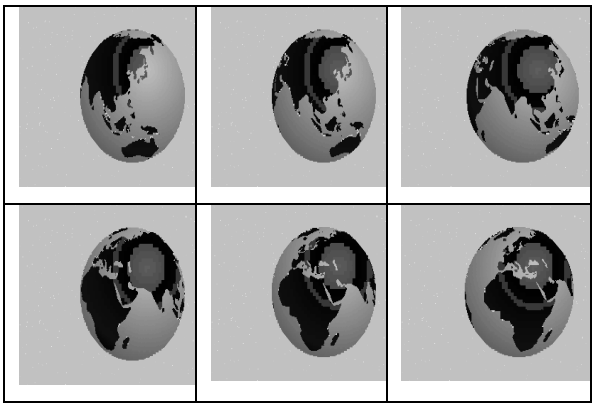
| Image | Noise | Denoise |
|---|---|---|
| 1.jpg | 1,2,3,4,6,8,9 | 1,2,3,4,5,6,7 |
| 2.jpg | 1,3,6,8,7,9,10 | 1,2,3,4,,6, |
| 3.jpg | 2,4,5,6,7,8 | 2,3,4,5,6,7 |
| 4.jpg | 1,3,4,5,6,7 | 1,2,3,4,5,6,7 |

Table 5.2: Noise vs Denoise

| **Unsequenced images** | **Sequenced images** |
|---|---|
| 'si1.jpg' | 's1.jpg' |
| 's10.jpg' | 's2.jpg' |
| 's11.jpg' | 's3.jpg' |
| 's12.jpg' | 's4.jpg' |
| 's13.jpg' | 's5.jpg' |
| 's14.jpg' | 's6.jpg' |
| 's15.jpg' | 's7.jpg' |
| 's16.jpg' | 's8.jpg' |
| 's17.jpg' | 's9.jpg' |
| 's18.jpg' | 's10.jpg' |

Table 5.3 UnSequenced vs Sequenced

**Images viewed as segment and sequenced as shown in the Following Figure 5.1**



ISBN : 978-81-943584-8-0

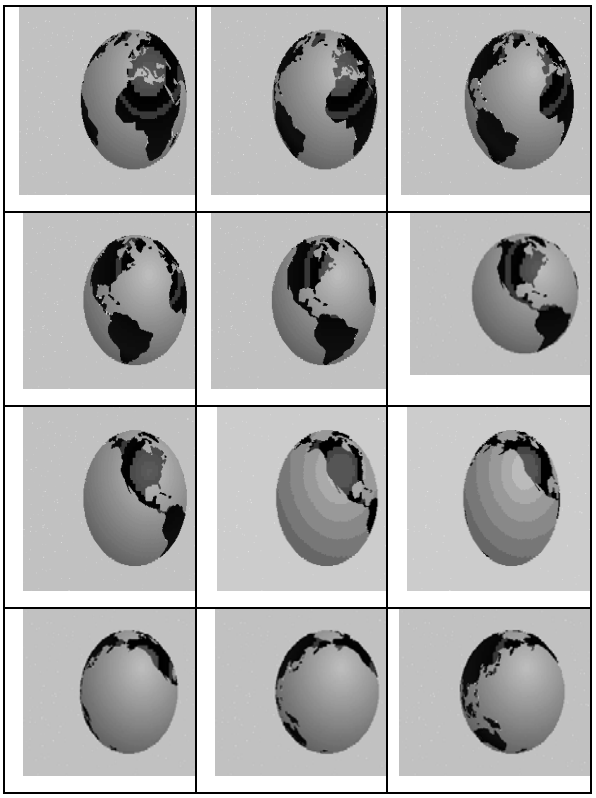


**Figure 5.1 : The Process of Sequenced Images**

Based on the spade algorithm threshold graph had been generated as shown in the Figure 5.2


**Figure 5.2 Threshold Graph for Image Frame**

**CPU Time for sequencing**

Using a SPADE Algorithm the CPU Time for Sequencing had been calculated and presented in the form of Figure 5.3 as shown below

ISBN : 978-81-943584-8-0

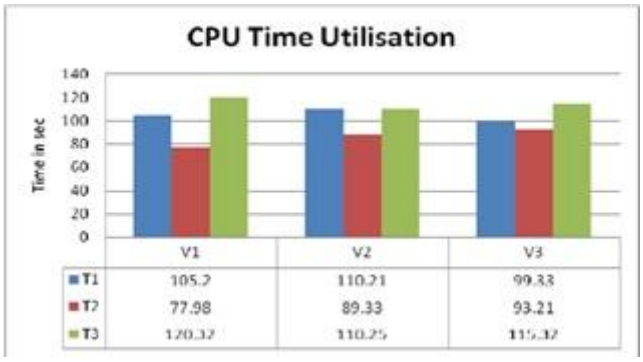



Figure 5.3 CPU Utilization during Sequencing the images

## 6. Conclusions

This proposed approach is to extract pattern from video images and large collections of images are derived, deals mainly with identification and extraction of unique features for a particular domain Now the idea is to use color shape in object classification field. The propose method is to compute size of color shape as a feature for each image to get information of the image. In this paper we are Extracting the Features of images and Identifying the image, after this we will put the frames, by putting the noise in image we get distorted image, to find which part of the image is distorted we put sequential pattern mining, By using the sequential pattern mining then we find the distortion less image and find the sequences of Image which is distorted.

## 7. References

[1]. J. Fernandez, N. Miranda, R. Guerrero and F. Piccoli (2007), "Appling Parallelism in Image Mining,"

[2]. Abhi Gholap, Gauri Naik, Aparna Joshi and CVK Rao (2005),"Content-Based Tissue Image Mining", IEEE Computational Systems Bioinformatics Conference - (CSBW'05), pp.359-363.

[3]. Sanjay T. Gandhe, K. T. Talele and Avinash G. Keskar (2007), "Image Mining Using Wavelet Transform", Knowledge-Based Intelligent Information and Engineering Systems, Springer link book chapter, pp. 797-803.

[4]. Manikandan, R. and Latha, R., "A Literature Survey of Existing Map Matching Algorithm for Navigation Technology", IJSRT, International Journal of Engineering Sciences & Research Technology, ISSN: 2277-9655, Vol. 03, CODEN: IJESS7, pp.326-331, 2017.

[5]. Sabyasachi Pattnaik, Pranab Kumar Das Gupta and Manojranjan Nayak (2008), "Mining images using clustering and data compressing techniques", International Journal of Information and Communication Technology, vol. 1, no. 2, pp. 131-147.

[6]. Lu Kun-Che and Yang Don-Lin (2009), "Image Processing and Image Mining using Decision Trees", Journal of information science and engineering, vol. 25, no. 4, pp. 989-1003.

[7]. L. Jaba Sheela and V. Shanthi (2007), "Image Mining Techniques for Classification and Segmentation of Brain MRI Data," Journal of Theoretical and Applied Information Technology," vol. 3, no. 4, pp. 115-121

[8]. Aura Conci and Everest Mathias M. M. Castro (2002), "Image mining by content", Expert Systems with Applications, vol. 23, no. 4, pp. 377-383.

[9]. Selim Aksoy and R. Gokberk Cinbis (2009), "Image Mining Using Directional Spatial Constraints," IEEE Geosciences and Remote Sensing Letters, vol. 7, no. 1, pp. 33-37.

[10]. Hui Jiang and Chong-Wah Ngo (2003), "Image Mining Using Inexact Maximal Common Subgraph of Multiple ARGs", http://vireo.cs.cityu.edu.hk/papers/vis03-jiang.pdf.

[11]. Wei Sun, David Taniar and Torab Torabi (2008), "Image Mining: A Case for Clustering Shoe prints," International Journal of Information Technology and Web Engineering, vol. 3, no. 1, pp. 7084.

[12]. P. Rajendran and M. Madheswaran (2009), "An Improved Image Mining Technique For Brain Tumour Classification Using Efficient classifier", International Journal of Computer Science and Information Security, vol. 6, no. 3, pp. 107-116.

[13]. Kun Qin, Zequn Guan, Deren Li and Xin Zhou Wang (2003), "Methods of remote sensing image mining based on concept lattice", Third International Symposium on Multispectral Image Processing and Pattern Recognition, vol. 5286, pp. 254-259. 13.

[14]. S. P. Victor and S. John Peter (2010), "A Novel Minimum Spanning Tree Based Clustering Algorithm for Image Mining," European Journal of Scientific Research, vol. 40, no. 4, pp. 540-546.

[15]. M. Hemalatha and C. Lakshmi Devasena (2011), "A Hybrid Image Mining Technique using LIMbased Data Mining Algorithm," International Journal of Computer Applications, vol. 25, no.2, pp. 15.

[16]. Keiji Yanai (2003), "Web Image Mining toward Generic Image Recognition", World Wide Web Conference Series.

[17]. Sanjay Silakari, Mahesh Motwani and Manish Maheshwari (2009), "Color Image Clustering using Block Truncation Algorithm", IJCSI International Journal of Computer Science Issues, Vol. 4, No. 2.

ISBN : 978-81-943584-8-0

[18]. Lukasz Kobylinski and Krzysztof Walczak (2007), "Color Mining of Images Based on Clustering", Proceedings of the International Multiconference on Computer Science and Information Technology pp. 203–212.

[19]. Rajshree S. Dubey (2010), "Image Mining using Content Based Image Retrieval System", (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 07, pp. 2353-2356.