

Image Substance Extraction using Data Mining Clustering Method



D.Saravanan

Abstract. *Data retrieval is one of the key challenging factor for today. Because of increasing the volume of data sets every year due to various factors. Information extraction in image data sets are too multifaceted compare with normal text data recovery. Image data set consist of different attributes those attribute sets are normalized before it extract from the stored data base. This required additional burden to the user who wish to extract any information from this data sets. This key challenges invite more researchers in the field of image data mining. Today many of the data sets in the form of image it gives more accurate result and more outputs. For extracting any image data image attributes are properly trained for better result. The proposed work based on grouping the data sets using image attributes. The entire process of this work divided into two major separate operations. Experiments done against various data sets, and outputs verified proposed work gives more accurate results than the existing techniques.*

Key words: *Image Clustering, Image Histogram, Image Attribute Selection, Threshold Values, Mining Attribute Selection and Image Content Mining.*

I. INTRODUCTION

Increasing the demand of data sets extracting the correct data are more challenging today. Many research contributing an enormous effort for extracting the needed information from this huge data set. Extracting the hypermedia data are more challenging factors than normal text extraction because of the hypermedia data sets nature. Today image data sets are used for various application domains such as in medical patient test analysis, pattern matching, diagnosing, categorization and more. In law and order identify the previous record matching information especially in psycho type criminals, analysis the given data. Climate analysis, risk analysis and education and more. For all the above domain the image data sets plays very vital role [1]. From the above example information's are collected over period of time, and huge image attributes are used to analysis and stored the data based on the image properties. This data sets helps the user bring the new knowledge based on the analysis [2]. Because of the increasing in demand today most of the research work devoted in this domain.

Faculty of Operations & IT
ICFAI Business School (IBS), Hyderabad,
The ICFAI Foundation for Higher Education (IFHE)
(Deemed to be university u/s 3 of the UGC Act 1956)
Hyderabad-India.
Sa_roun@yahoo.com

Revised Manuscript Received on December 30, 2019.

* Correspondence Author

D.Saravanan*, Faculty of Operations & IT ICFAI Business School (IBS), Hyderabad, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

This image data sets satisfies the user query with greater accuracy. For performing this operations video data mining play major role in knowledge extraction process.

This require more additional domain knowledge because of the nature of the data sets. Like other searching user can specify the query, the corresponding information are retrieved and present to the user. But image processing required both training and storing phase. For extracting and storing any image data sets requires some additional knowledge it helps to brings the needed information more accurate and efficiently. For making this operation more automated way more knowledge skills are required for the users [3]. Existing may techniques and procedure works well for image data sets. But increasing of the technology today creating and storing of this data sets are easier for many users. From this huge respiratory extract the need content are really a challenging factor for the researchers. This new technology helps the users to extract the need content used in many application domain such as medical, education, training, security and more [4]. Any image extraction process done the following steps of operations. 1. Apprehend the input data sets- It helps the researchers and the users what type of data sets are there, what all are the attributes are used to define, store, extract this data sets for knowledge extraction. 2. Categorize the input data sets – This one of the most important steps for image extraction , image contents are stored based on any of the image attributes .This attributes not only used to categorize the data sets it also helps for retrieve the particular content. 3. Knowledge extraction- This step done either content based extraction or properties based extraction. In both technique image attribute sets help to extract the relevant information from the stored sets. 4. Knowledge representation – This step helps to present the extracted knowledge to the end user in terms of any visualizing technique. [5]. In most the existing image based technique works based on the image content based methods here any of the image information are consider for extraction process. The proposed technique image attribute are consider it helps to bring the more accurate result and more number of similar output also generated though the proposed technique. In comparison normally done based on the user input query but the proposed technique information extraction done both user side i.e. training the input data sets, due to extraction done effectively. The second phased in storage side here also contents were trained and stored for beings the better result [6]. Proposed method works separately attributes properties are treated separately for both user side and storage side. Based all this assumption the proposed technique categorized into structure conversion finding the repetition, grouping the relevant frames, user side training, storage side training, knowledge query, and knowledge extraction.

Image Substance Extraction Using Data Mining Clustering Method

Performing the structure conversation the motion operations are converted to static operation then image properties helps to segregate the structured information [7].

II. EXISTING SYSTEM

Technology allows us to create huge volume of digital information's. Creations of any digital information's are easy but retrieve the exact information need a domain knowledge. Especially the multimedia data sets are too complex in nature. Extraction information from this complex data requires additional domain knowledge to the user community. Most of the existing techniques bring the content based on the image properties. Existing techniques find the relevant information by image evidence and opinion methods. [8]

A. Drawback of existing System

1. No single procedure
2. Grouping the similar structure takes more time.
3. Existing technique works well in specific input.
4. Image separation done with image content.
5. Efficiency factors are low.

III. PROPOSED SYSTEM

Proposed system works based on the image attributes property. Initially video files are transformed into static structures, based on the video file length this static structure is created. After creating the static structure number of duplication and unwanted files may present in the entire list. This list need to be cleaned else it create lot of time for any process and efficiency of the method got reduced. For this image attribute pixel values are calculated based on the threshold value comparison unwanted static structured is removed from the list. After eliminating the unwanted list rest of the input structured information are combined and stored separately for the future process. In this technique image pixel color values are consider this values are stored in the trained data base, it helps for image comparison and image retrieval [9]. For image comparison here hierarchical clustering technique uses, a comparative study undertaken for the existing hierarchical clustering technique from that identified that none of the existing clustering technique not used for all type of input files. Some of them works well in particular type of input files. This technique further explained in detail in experimental setup process. In using this clustering technique also done in user side process and storage side process. This helps to reduce the searching time and it brings the number of output effectively. Training process consume less time, but it helps in the searching process most effectively. So it is necessary this tow phases are compulsory for this proposed techniques. Experimental outputs verified that proposed technique time taken, number of output generated are improved with existing method. And this technique works well in all type of video files most effectively.

A. ADVANTAGE OF PROPOSED TECHNIQUE

1. Any type of video files the proposed techniques works.
2. Input data sets consist of trained information due that efficiency of the process improved.
3. Number of grouping based on the input data file are fast and more.

4. Brings the higher rate of outputs.
5. Searching time improved.

IV. EXPERIMENTAL SETUP

A. Video Mining

Data extraction is one of the most demand area for most of researchers today because of the increasing and availability of the huge data sets. Take out the useful information from this collection is really on of the key challenging factor. Every time in data mining user wish to get new knowledge or new rules from the stored data sets. This rules are unidentified previously based on the query the user wish to get new knowledge. Before user extract a new knowledge user or researcher to do the cleaning process in the stored data sets. Because of huge amount of information data's sets need to cleaned, unwanted, missed values, wrong values, repeated values may present in the data sets. It is necessary to extract correct data from the stored data sets, data sets need to be cleaned. After cleaning data's are arranged sequentially, it helps retrieve the efficient data quickly. In fig 1 the cleaning process shown clearly first the dynamic video file converted to static file then unwanted and repeated files are removed using video pixel values or video file handling technique. After the unwanted video file removed from the stored data set rest information's are integrated and stored in the data base for further operation.

B. Effectively used Data mining method

A. Simulated neural system

This similar for natural system construction. The entire functions are divided into various steps. In step on response level constructed with multi-level function. After this first level a hidden level is constructed. This hidden level based on the users operations it may presented. Based on the users operations any additional levels are added for bringing the effective result. This level architecture works well for most of the arrangement operations with tuned input data sets.

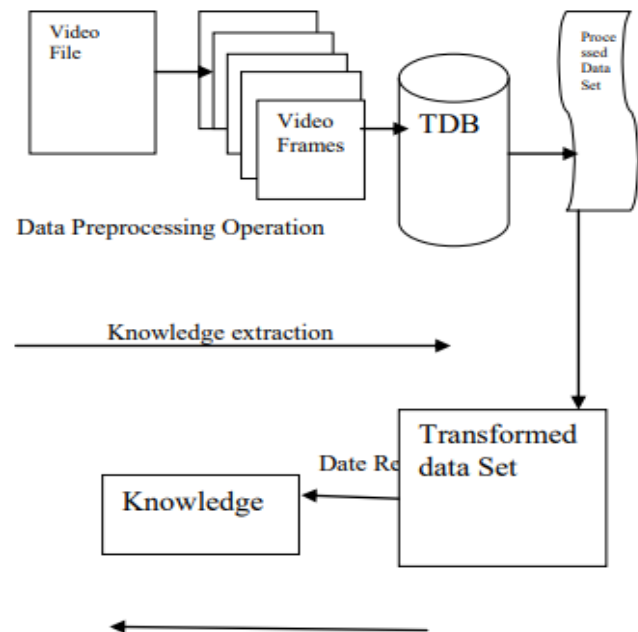


Fig. 1. Proposed Model for information retrieval

B. Resolution leaves

This structure helps to represent the data information in priority level architecture. This structure helps to segregate the data sets based on the priority they appearing in the data sets. Each level input data's are tested based on the testing outcome next level created. Like the levels get increased based on the decision outcome or finding the end solution. After constructing the level architecture unwanted information's are removed using error handling mechanism. This mechanism done in two ways first un wanted information are removed then level structure are constructed or after constructing the level errors are removed. Both methods helps to remove the unwanted information it helps to reducing the searching time.

C Nearby grouping technique

Points or values are grouped based on their relevance. In the groups points which are very closer are put into one groups, points are faraway they put into other groups. Similarity between the groups are differ. Making this group used need information about the data points. Finally based on the users requirements similar groups are combined and make one big group for the operations convenience.

Regulation generation

In hierarchical structure information's are extracted based on instruction based function. Train the data sets, generation of output everything based on the instruction based information. After generating the instruction user need to identify which instruction best of execute the particular operations. Based on the applications and usage this instructions are defined same set of instruction used for further operations. This applications used in many real world operations such as identify the wrong operations in banking sector, find the bestselling product in the market based on customer purchase power , history. The main advantage of this method it works well even in small data sets.

C. Video grouping

Grouping of data based on the property or value of the data available in the data set. This grouping normally done based on the similarities between the data. If items or data's are similar they are all put in to one groups. Relations between this data in the groups are very closer among the data. Items or values are not similar they are all group into other category. Similarity between this data sets are not too similar. Like very similar items are kept in one groups other items are kept in other group [10]. This grouping of information's helps the user to identify the closeness among the given data sets

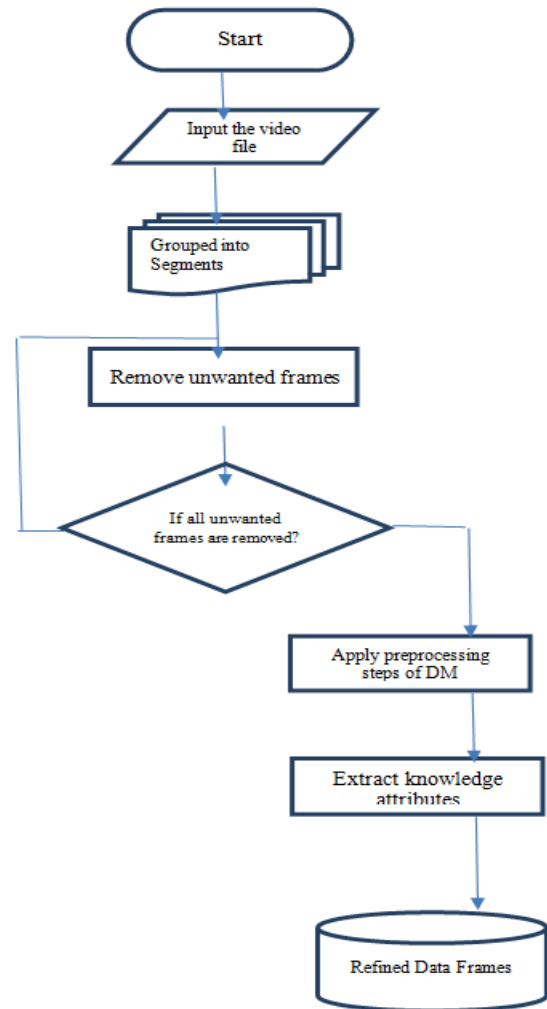


Fig Flow chart for knowledge extraction

It helps for the researchers and users to retrieve the similar items or information easily and also improve the searching speed. Variety of grouping mechanism are available, based on the different property of the data data sets. Most of this technique works based on space between the items. Based on this space difference are calculated to form the grouping [11]. Normal grouping mechanism is not suit for video grouping because here video information's are complex in nature. Property of this information's are very difficult to separate. For this reason still number of research work carried in this particular filed and it gives more challenging for the researchers this process had shown in the fig 2, 3, 4 and Table 1.

D. Video Classification

After successful after preprocessing steps cleaned data sets are integrated and stored separately for further operations. Here the data's are assigned object identifier it helps to retrieve the information more quickly. This also improve the speed of searching and retrieval process [12]. After this information's are assigned a catalogue it helps to extract need information more quickly. Each group user need to create a key file are selected through help of this information are identified and extracted

Pseudo code for knowledge extraction:

Step1: Find the input vide file.

Stept2: Segment input file into frames

Image Substance Extraction Using Data Mining Clustering Method

Step3: Stored the segmented frames.

Step4: Using image comparison remove unwanted file.

Step5: Check all the frames are compared.

Step6: Repeat step 3-5 until all frames are tested.

Step7. Stored the cleaned frames.

Step8. Used data mining preprocessing step

Step9. Extract knowledge information.

Step10. Repeat step 8-10 until get the need information.

Step 11. Stop

VI. IMAGE MINING

Increasing the demand of image files today number of research work done on this side. This one of the most thruster area for many researchers. During the image processing two important function need to be consider first similarity factor for input and extracting frames, which domain this application going to applied [13]. Image processing differs from traditional mining operations, every image mining done with training processes, query process. Before we extract any information from the stored content, information needs to be trained. Training the image data base one of big challenge for the researchers, number of technique currently available. It shows in the fig 5. Performances of the proposed system frame count Vs time taken are shown in the table 2, performance graph shown in the fig 6.

A. Image content extraction procedure

Image content extraction procedure done in four steps:

1. Attribute mine: Motion objects are converted to frames, after successfully eliminate the duplicates it is necessary to find unlikeness of every frames. For this any image property are consider. With help image property need to construct building dictionary of image frames.
2. Entity classification: After the process one, it is necessary to create entity identification. Each frame treated single entity. This step also identifies the difference between successive entities. It helps the user classify the entities in to different groups.
3. Generate supporting descriptions: Create descriptors for every image frames with help of mining.
- 4 Repeat the step up to all frames are properly trained.

VI. EXPERIMENTAL OUTCOMES



Fig. 2 Motion object converted are converted as static objects.



Fig. 3. Grouping the static objects

Frames	Cluster	Mins	Sec
1	6544	0	6
2	6530	0	11
3	6689	0	17
4	6719	0	23
5	6901	0	28
6	6911	0	33

Fig 4. Frame Vs Time taken in millisecond for different static objects.

TABLE 1. Grouping the static files Vs Period in Sec

	id	frame	numclus	time
	1	0	35204	3:17:55 PM
	2	1	35175	3:18:01 PM
	3	2	29162	3:18:07 PM
	4	3	29762	3:18:14 PM
	5	4	30475	3:18:20 PM
	6	5	30975	3:18:26 PM
	7	6	31351	3:18:33 PM
	8	7	31867	3:18:39 PM
	9	8	32339	3:18:45 PM
	10	9	32427	3:18:52 PM
	11	10	32482	3:18:58 PM
	12	11	32440	3:19:04 PM
	13	12	32842	3:19:11 PM
	14	13	32964	3:19:17 PM
	(AutoNumber)	0		

Image Substance Extraction using Data Mining Clustering Method

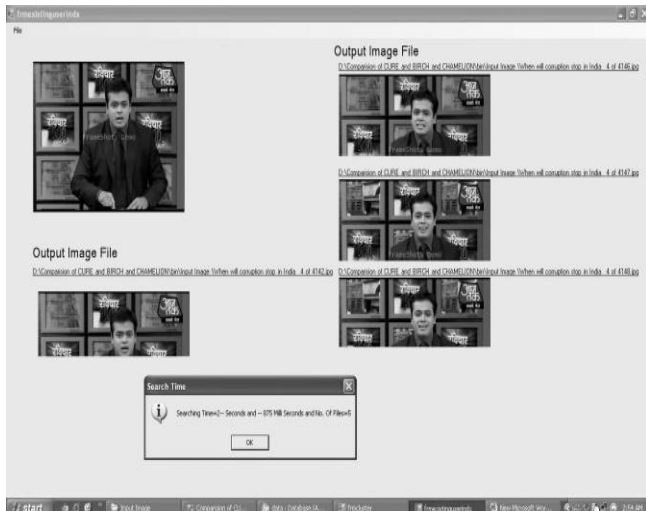


Fig. 5. Image extraction using image attributes values.

TABLE 2. Static structure Vs Period

ferment	milliseconds	category
25	2586	News
50	2640	News
75	2625	News
100	3062	News
125	3375	News
150	3421	News

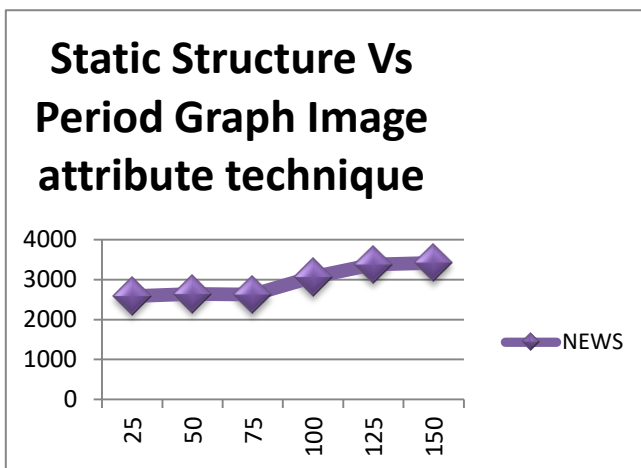


Fig 6. Efficiency chart Static structure Vs Period (millisecond)

VII. CONCLUSION

Images acting a significant character in data extraction. This technology helps the researchers and user to bring the truthful evidence in small amount of time. This research outcome verified image extraction are more operative than

any current technique available. In upcoming research works is to expand the work with other image property are combing and bring the number of outputs with short time duration.

REFERENCES

1. Antonios Oikonomopoulos, Ioannis Patras, and Maja Pantic (2006). Spatiotemporal salient points for visual recognition of human actions, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 36, NO. 3, JUNE 2006
2. D.Saravanan “ Effective text data extraction using Hierarchical clustering technique” Int. Journal of Engg., and Advanced Technology(IJEAT),Feb 2019, Vol. 8, Issue 3-S, Pages 741-743(ISSN 2249 8658)June 2019
3. Yi Yang , Feiping Nie , Dong Xu , Jiebo Luo , Yueting Zhuang , Yunhe Pan(2012). A multimedia retrieval framework based on semi-supervised ranking and relevance feedback, IEEE Trans. Pattern Anal. Mach. Intell. 34 (5) (2012) 723–742.
4. D.Saravanan, ”Information retrieval using image attribute possessions” Soft computing and signal processing , Advances in Intelligence systems and computing 898, Springer. DOI:10.1007/978-981-13-3393-4_77, Pages 759-767. March 2019.
5. J Y. Wang C.-S. Chua(2005) Face recognition from 2D and 3D images using 3D Gabor filters, Image Vis. Comput. 23 (11) (2005) 1018–1028.6.
6. M.Fayyad , G. Piatetsky- Shapiro, P. Smyth , and R.Uthurusamy, advanced in knowledge discovery and data mining. AAAI/MIT press, 1996.
7. D.Saravanan,Dr. Dennis Joseph,” Image data extraction using image similarities”, Lecture notes in Electrical engineering, Volume 521, Pages 409-420, ISBN:978-981-13-1905-1,Nov 2018.
8. Till Quack, Vittorio Ferrari, Luc Van Gool(2006) Video mining with frequent itemset configurations, in: Conference on Image and Video Retrieval CIVR, 2006.
9. D.Saravanan,” Efficient Video indexing and retrieval using hierarchical clustering techniques”, Advances in Intelligence systems and computing, Volume 712, Pages 1-8, ISBN:978-981-10-8227.6, Nov-2018
10. Raymond T. Ng and Jiawei Han . member,IEEE computer society CLARANS: A method for clustering objects for spatial data mining . IEEE transactions on knowledge and data engineering, vol.14, No.5, sept-oct.2002.
11. D.Saravanan” Segment Based Indexing Technique for Data file “, Procedia of computer Science,87(2016), Pages 12-17,ISSN: 1877-0509, June 2016
12. D.Saravanan, ”Various Assorted Cluster Performance Examination using Vide Data Mining Technique” , Global journal of pure and applied Mathematics , Volume 11, No.6(2015), Nov 2015,ISSN 0973-1768, Pages 4457-4467
13. M. Shyu, Z. Xie, M. Chen, and S. Chen, “Video semantic event/concept detection using a subspace-based multimedia data mining framework,”IEEE Transaction on Multimedia, Vol. 10, pp 252-259, Feb2008

AUTHORS PROFILE

D.Saravanan did his M.E in computer science and Engg.,and completed his Doctor of philosophy in the same area. He had 20.5 years of teaching experience. His area of interest is Data mining, knowledge extraction, image mining and Information Retrieval.