

# An Adaptive Correlation Based Video Data Mining using Machine Learning



Mallikharjuna Lingam K, VSK Reddy

**Abstract:** With the immense growth in the multimedia contents for education and other purposes, the availability of the video contents has also increased. Nevertheless, the retrieval of the content is always a challenge. The identification of two video contents based on internal content similarity highly depends on extraction of key frames and that makes the process highly time complex. In the recent time, many of research attempts have tried to approach this problem with the intention to reduce the time complexity using various methods such as video to text conversion and further analysing both extracted text similarity analysis. Regardless to mention, this strategy is again language dependent and criticised for various reasons like local language dependencies and language paraphrase dependencies. Henceforth, this work approaches the problem with a different dimension with reduction possibilities of the video key frames using adaptive similarity. The proposed method analyses the key frames extracted from the library content and from the search video data based on various parameters and reduces the key frames using adaptive similarity. Also, this work uses machine learning and parallel programming algorithms to reduce the time complexity to a greater extent. The final outcome of this work is a reduced time complex algorithm for video data-based search to video content retrieval. The work demonstrates a nearly 50% reduction in the key frame without losing information with nearly 70% reduction in time complexity and 100% accuracy on search results.

**Keywords:** Video Data Mining, Key Frame Reduction, Adaptive Similarity, Video Retrieval, Multiple Dataset Performance.

## I. INTRODUCTION

The ease of effective computing is a research direction, which works towards empowering computing devices for human intelligence. One of the highly popular human mode of communication is expression with verbal communication. The expressions usually include hand gesture, body posture and the facial expression. For making the computing devices at par with the human intelligence, the inclusion of capabilities to interpret video contents. The work of R. Cowie et al. [13] have demonstrated that during a human machine interaction, the video components as an input feed is an essential part. Based on this phenomenon, the work by M. Scheutz et al. [11] demonstrates the proof of concept and

reports the improvements over other communication methods. During the training of these computing devices, the video content libraries are used. The fundamental purpose of the training process is to make the computing devices enabled for classification, clustering and detection of the expressions or context. These actions can be achieved by empowering the devices or the software agents or the applications to match with the similar video contents [Fig – 1].

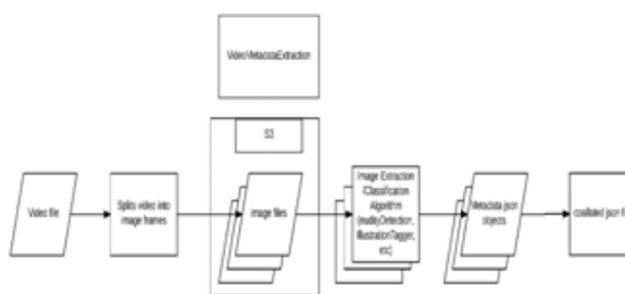


Fig. 1 Traditional Video Mining Process

The similarity measuring algorithms for the video content retrieval are highly popular and widely accepted. The YouTube’s video relevancy algorithm by P. Covington et al. [12] or the NetFlix’s movie recommendation algorithm by C. A. Gomez-Uribe et al. [2] are highly admired in the field of this research. Also, the content mining for audio is not very different from these two states of the art algorithms. The report by M. Schedl et al. [10] have also expressed that the challenges in music recommendation system are not very different from video content retrieval or content search.

The different research findings have reported the complexities of the research is multiple directions. The work by M. J. Duque et al. [9] have suggested that the video contents can be highly difficult to be classified or matched as in case of human presence in the content, the change in expression can lead to high change in the video content. Thus, in spite of similar content and context, the video frames can show different characteristics. In addition to that, many of non-computing researches have also demonstrated that, the use of emotion as a factor during video retrieval can be highly disastrous. The work by D. A. Worthy et al. [3] have confirmed to this theory by her case studies. Henceforth many of the parallel research attempts have suggested to ignore the human intelligent properties and rather focus on the image specific properties for the content retrieval. Later this theory was made further stronger with the case study by J. Etkin et al. [6]. Henceforth, this work attempts to reduce the time complexity of the retrieval process only using the physical properties of the key frames extracted from the video library contents.

Manuscript published on November 30, 2019.

\* Correspondence Author

**Mallikharjuna Lingam K,\*** Research Scholar, Faculty of Engineering, Lincoln University College, Malaysia

**VSK Reddy:** Professor, Faculty of Engineering, Lincoln University College, Malaysia.

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

The rest of the work is elaborated such that in Section – II, the standard video mining or video retrieval framework functionalities are elaborated, in Section – III, the outcomes from the parallel research works are analysed, in Section – IV, the problem is formulated using the mathematical models, in Section – V, the proposed algorithms are discussed with the purposes and the advantages, in Section – VI, the used datasets are discussed to demonstrate the purpose of use, in Section – VII, the obtained results are discussed and in the Section – VIII, the work presents the final research conclusion.

## II. VIDEO DATA MINING FRAMEWORK

In this section of the work, the fundamentals of video content retrieval process are mathematically presented. *This* analysis will help in realizing the bottlenecks of the existing systems. The initial component of the video data mining or retrieval is the search content. The searching content can be a limited frame or limited duration video data or an image file. Here assuming that the input searching content is a video data (V) containing multiple frames as  $v_i$ , then the relationship can be formulated as,

$$V = \frac{v_i}{\Delta t} \quad (\text{Eq.1})$$

Where,  $t$  is the minimum time duration to separate the frames in the video content.

The second important component is the video library ( $V'$ ), which again contains multiple video data sets. Hence, this relation can be formulated as,

$$V' \rightarrow [v_1, v_2, v_3, v_4, \dots, v_n] \quad (\text{Eq.2})$$

Again, the individual video contents ( $v_x$ ) must be separated based on the key frames as,

$$v_x = \sum_{j=1}^n v_j \quad (\text{Eq.3})$$

The video file frame contents must be processed for extraction of object information, in order to find out the eligibility of this frame for key frame. If the frame is identified as key frame, then this frame will take a key part in identification of the matching content.

The key frame extraction process defines the object extraction [Obj] process and for any two key frames ( $v_j, v_{j+1}$ ) can be represented as,

$$Obj_j = \Phi(v_j) \quad (\text{Eq.4})$$

And,

$$Obj_{j+1} = \Phi(v_{j+1}) \quad (\text{Eq.5})$$

Now, if the extracted contents from each frame are different, then the first frame will be considered as key frame due to the change in the objects in the frame.

Hence, here, the initial frame must be considered as key frame ( $v_j$ ).

With this principle, a number of key frames from each video content data will be extracted. As,

$$v_j = [Obj_1, Obj_2, \dots, Obj_n] \rightarrow \ddot{O} \quad (\text{Eq.6})$$

The similar action is also to be taken on the searching video content and that content will as well generate a set of objects in return, which can be formulated as,

$$v_j = [Obj'_1, Obj'_2, \dots, Obj'_n] \rightarrow \hat{O} \quad (\text{Eq.7})$$

Finally, based on the match between  $\hat{O}$  and  $\ddot{O}$ , the final search results can be generated.

The fundamental bottleneck in this process is the higher time complexity. During the key frame extraction process, a little change in the object shape or object location can also create a significant difference in identification of the key frames. As identified, a greater number of key frames can increase the time complexity of the process to a greater extent.

The solution to this problem is further analysed and presented in the forthcoming sections of this work.

## III. OUTCOME FROM THE PARALLEL RESEARCHES

In this section of the work, the recent and parallel research outcomes are analysed for better formulation of the problem in the next section of the work.

During the video mining process, the behavioural analysis of the objects in the video frame can be considered in order to identify the key frames and the content similarity as projected by L. Ardissono et al. [8]. The further enhancement of this work was proposed by A. Hawalah et al. [1]. This work proposes to model the behavioural pattern of the objects and further utilize the same for pattern mining during video retrieval. The work uses a case study for user modelling. The video content retrieval for social media-based content can be easy as demonstrated by H. Yin et al. [5] as the profiling of the users and the relevant objects can be tracked and identified from the beginning of the existence.

Yet another dimension of the research is to incorporate the neural networks for classification and retrieval. The work by J. Li et al. [7] have considered the online business situation and proposed a cross breed encoder utilizing the structure of a neural system for session-based suggestion framework. Be that as it may, the quantity of things considered were extremely less with the nonattendance of a portion of the significant ones including value relations and thing classes.

Other research outcomes have demonstrated the use of sequential model for the content retrieval. This model for allosteric guideline of catalysts proposes that the subunits of multimeric proteins have two conformational states. The official of the ligand causes conformational change in different subunits of the multimeric protein. Even though the subunits experience conformational changes freely, the switch of one subunit makes different subunits bound to change, by decreasing the vitality required for resulting subunits to experience the equivalent conformational change. The work by E. Smirnova et al. [4] establishes the similar process and showcased higher performances.

Henceforth, with the detailed understanding of the parallel research outcomes, this work, elaborates the problem in the next section of the work.

IV. PROBLEM FORMULATION

With the detailed understanding of the fundamentals of video content extraction and the recent research outcomes, in this section of the work, the problem is formulated with the help of mathematical lemma.

**Lemma:** Reduction of the key frames without reducing the key content of the video data can reduce the time complexity of the video content matching and retrieval process.

**Proof:** In programming that help activity, particularly 3D designs, there are numerous parameters that can be changed for any one item. One case of such an article is light. Lights have numerous parameters including light power, shaft measure, light shading, and the surface cast by the light. Assuming that an artist needs the bar size to change easily starting with one esteem then onto the next inside a predefined timeframe, that could be accomplished by utilizing keyframes. Toward the beginning of the liveliness, a shaft estimate esteem is set. Another esteem is set for the finish of the liveliness. Consequently, the product program naturally inserts the two qualities, making smooth progress. Firstly, assuming that any video data V can be separated into multiple frames as set F. Thus, the formulation can be presented as,

$$V = F[] = \sum_{i=1}^n f_i \tag{Eq.8}$$

The total number of key frames are assumed to be n and for each key frame verification with the searching content can be presented as t. Henceforth, the total time, T, can be presented as,

$$T = n * t \tag{Eq.9}$$

In the due course, the reduction this time factor, T, must be realized in order to prove the lemma.

Further, each frame in the key frame set must go through a reduction process, where the similarity of any two frames must be considered based on the contour, C, of the object present, displacement, D, and finally the variation of image pixel density, P. Regardless to mention, the frames are considered in this situation with single object.

Henceforth, firstly for any two key frames, f<sub>i</sub> and f<sub>j</sub>, the object contours as C<sub>i</sub> and C<sub>j</sub>, displacement as D<sub>i</sub> and D<sub>j</sub>, and P<sub>i</sub> and P<sub>j</sub> respective, must be extracted. The formulations can be denoted as,

$$C_i \leftarrow \mathbb{I}(f_i) \tag{Eq.10}$$

And,

$$C_j \leftarrow \mathbb{I}(f_j) \tag{Eq.11}$$

Also,

$$D_i \leftarrow \xi(f_i) \tag{Eq.12}$$

And,

$$D_j \leftarrow \xi(f_j) \tag{Eq.13}$$

Further,

$$P_i \leftarrow \psi(f_i) \tag{Eq.14}$$

And,

$$P_j \leftarrow \psi(f_j) \tag{Eq.15}$$

Furthermore, if the frame object contours are nearly 50% similar to the consecutive frames, then the key frames can be considered as nearly similar frames and one of these two frames can be dropped.

$$\frac{\mathbb{I}(f_i)}{\mathbb{I}(f_j)} \rightarrow \nabla \mathbb{I}(C_i) < \frac{\int_{i=1}^{50} \mathbb{I}(f_i) + \int_{j=51}^{100} \mathbb{I}(f_j)}{\mathbb{I}(f_i) + \mathbb{I}(f_j)} \tag{Eq.16}$$

Similarly, for the displacement, the formulation can be presented with the respect to the complete frame size.

$$\frac{\xi(f_i)}{\xi(f_j)} \rightarrow \nabla \xi(D_i) < \frac{\frac{d}{\delta Len \rightarrow 50} f_i + \frac{d}{\delta Len \rightarrow 100} f_j}{\mathfrak{F}f_i + \mathfrak{F}f_j} \tag{Eq.17}$$

Correspondingly, for the image pixel intensity, the formulation can be presented with respect to the rate of pixel intensity change of the key frames,

$$\frac{\psi(f_i)}{\psi(f_j)} \rightarrow \nabla \psi(P_i) < \frac{|I(f_i) - I(f_j)|}{I(f_i) + I(f_j)} \tag{Eq.18}$$

Henceforth, due to these processes, the number of key frames, m, will reduce. Further the time complexity must be analysed.

Assuming that the contour detection time, t<sub>1</sub>, the deviation of object detection time, t<sub>2</sub>, and the pixel intensity rate of change, t<sub>3</sub>, thus the total time for the processing, T', can be presented as,

$$T' = m * t_1 + m * t_2 + m * t_3 \tag{Eq.19}$$

It is natural to realize that the number of key frames after the reduction process will reduce to 50% to be at the least. Thus,

$$T' = \frac{n}{2} * t_1 + \frac{n}{2} * t_2 + \frac{n}{2} * t_3 \tag{Eq.20}$$

Further, assuming that,

$$t_1 + t_2 + t_3 \rightarrow k \tag{Eq.21}$$

Hence, From Eq. 20,

$$T' = \frac{3n}{2} * k \tag{Eq.22}$$

Realizing the fact that k << n, From Eq. 22,

$$T' = \frac{3n}{2} \tag{Eq.23}$$

Further, this work applies the machine learning methods to obtain the similarity scales, hence these three tasks can be performed in parallel. Hence, the final time complexity can be viewed as,



$$T' = \frac{n}{2} \quad (\text{Eq.24})$$

Hence, it is natural to realize that,

$$T' \ll T \quad (\text{Eq.25})$$

Thus, the reduction of the key frames without reducing the information from the key framesets, can significantly reduce the time complexity of the process.

Further, realization of the algorithms proposed in this lemma is presented in the further section of this work.

## V. PROPOSED ADAPTIVE CORRELATION ALGORITHM FOR CONTENT SIMILARITY

In this section of the work, two novel algorithms are presented.

Firstly, the algorithm for image key frame similarity is presented. The distinction as for different methods referenced already, for example, MSE or PSNR is that these methodologies gauge supreme blunders; then again, SSIM is an observation-based model that considers image debasement as saw the change in auxiliary data, while likewise joining significant perceptual marvels, including both luminance concealing and differentiation covering terms. Basic data is the possibility that the pixels have solid between conditions particularly when they are spatially close. These conditions convey significant data about the structure of the articles in the visual scene. Luminance veiling is a marvel whereby image twists (in this unique situation) will, in general, be less unmistakable in brilliant areas, while differentiate covering is a wonder whereby bends turned out to be less obvious where there is the noteworthy action or surface in the image.

### *Algorithm - 1: Frame Similarity Extraction & Reduction using Parallel Kernel Calls Algorithm (FSER-PKC)*

- Step - 1. Accept the key frames from video data as  $F[i]$   
 Step - 2. Initiate three parallel kernel calls using CUDA framework  
 Step - 3. Cuda\_Kernel Call\_1: For each key frame  $F[i]$
- a. Calculate the gradients for each object
  - b. Apply Non-Maximization method
  - c. Increase the threshold of the object regions
  - d. Calculate the contour as  $C[i]$
- Step - 4. Cuda\_Kernel Call\_2: For each key frame  $F[i]$
- a. Calculate the object centroid
  - b. Calculate the object position as  $D[i]$
- Step - 5. Cuda\_Kernel Call\_3: For each key frame  $F[i]$
- a. Calculate the over all frame pixel density
  - b. Calculate the object region image density as  $P[i]$
- Step - 6. For  $(C[i], D[i], P[i]) \&\& (C[i+1], D[i+1], P[i+1])$
- a. Check for the threshold [Mention in the Lemma]
  - b. If similar,
    - i. then accept  $F[i]$  and discard  $F[i+1]$
- Step - 7. Report final key frames as  $F'[]$

The second algorithm is responsible for video content matching and extraction from the video library. Video content analysis (VCA) is another innovation. New applications are as often as possible found, anyway the reputation of various kinds of VCA contrasts generally. Functionalities, for example, movement location and individuals checking are accepted to be accessible as business off-the-rack items with a respectable reputation, even freeware, for example, probiotics Flowstone can deal with development and shading investigation. In numerous spaces VCA is actualized on CCTV frameworks, either circulated on the cameras (at-the-edge) or incorporated on committed preparing frameworks. Video Analytics and Smart CCTV are business terms for VCA in the security area. In the UK the BSIA has built up a presentation control for VCA in the security space. Notwithstanding video examination and to supplement it, the sound investigation can likewise be utilized.

### **Algorithm - 2: Video Content Extraction using Reduced Key Frames (VCE-RKF)**

- Step - 1. Accept the key frames from searching video content  $F''[]$   
 Step - 2. Accept the video library key frames as  $F'[]$   
 Step - 3. For each  $F'[]$
- a. If  $F'[i]$  is Non similar to  $F''[j]$
  - b. Increase the dis-similarity count, DisSimCon.
- Step - 4. Calculate the Mean DisSimCon for the searching video  
 Step - 5. If frame DisSimCon is higher than video DisSimCon mean
- a. Then discard the frame
- Step - 6. Else
- a. Consider the frame for matched frame as  $F''[]$
- Step - 7. If Count of  $F''[] = F'[] * 95\%$
- a. Then return the video library as match result
- Step - 8. Else,
- a. Proceed to the next video library

Further, in the further section of the work, the obtained outcomes from these algorithms are analysed.

## VI. DATASET ANALYSIS

In this section of the work, the used standard datasets are discussed. For validating the proposed algorithms this work tests the algorithms on five different datasets. The enthusiasm for CBIR has developed due to the constraints intrinsic in metadata-based frameworks, just as the expansive scope of potential uses for proficient picture recovery. Printed data about pictures can be effectively looked utilizing existing innovation, however this expects people to physically depict each picture in the database. This can be unfeasible for huge databases or for pictures that are produced consequently, for example those from reconnaissance cameras. It is likewise conceivable to miss pictures that utilization various equivalent words in their depictions. Frameworks dependent on classifying pictures in semantic classes like "feline" as a subclass of "creature" can maintain a strategic distance from the

mis-categorization issue, however will require more exertion by a client to discover pictures that may be "felines" yet are just delegated a "creature". Numerous measures have been created to order pictures, however all still face scaling and mis-categorization issues.

The details of the datasets are listed here [Table – 1].

TABLE I DATASET DESCRIPTION

Serial Number	Dataset Name	Year	Generic Characteristics
1	BBC Eastenders	2018	Video Dataset with small length contents for instance retrieval
2	IACC.3	2018	Video Dataset with contents for Ad-hoc detection
3	i-LIDS	2017	Video Dataset with surveillance video contents for key frame detection based on events
4	Flickr	2012	Video Dataset with small length contents for instance retrieval
5	BBC Rushes	2011	Video Dataset with small length contents for instance retrieval

VII. RESULTS AND DISCUSSION

In this section of the work, the highly satisfactory results obtained from the proposed algorithms are analysed. The results are discussed in four different sections as the result of key frame reduction method, key frame-based error threshold, the retrieval process accuracy and finally the time complexity.

A. Key Frame Reduction

Firstly, the key frame reduction process results are elaborated here [Table – 2]:

TABLE II KEY FRAME REDUCTION RESULTS

Dataset Name	Content Name	Initial Extracted Key frames	Final Key Frames after Reduction	Percentage of Reduction
BBC Eastenders	Sample - 1	79	35	55.70
IACC.3	Sample - 2	187	90	51.87
i-LIDS	Sample - 3	86	38	55.81
Flickr	Sample - 4	63	30	52.38
BBC Rushes	Sample - 5	122	63	48.36

The result is also visualized graphically [Fig – 2].

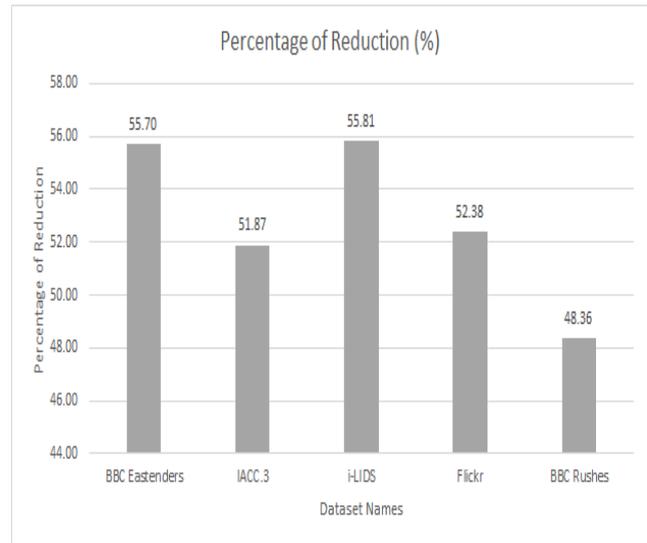


Fig. 2 Key Frame Reduction

It is natural to visualize that the reduction is nearly 50% for majority of the datasets.

The detailed advantages of this process are elaborated in the prior section of this work.

B. Video Library Based Error Threshold

Secondly, the video library-based error threshold results are elaborated here [Table – 3]:

TABLE III ERROR THRESHOLD ANALYSIS

Searchin g Data	BBC Eastender s Dataset	IACC. 3 Dataset	i-LIDS Datase t	Flickr Datase t	BBC Rushes Datase t
Sample - 1	1.0554	1.0907	1.0966	1.0050	0.5415
Sample - 2	0.8135	1.0128	1.0134	1.0079	1.0966
Sample - 3	0.8143	0.6110	1.0143	1.0088	0.9143
Sample - 4	0.5949	1.0138	0.5402	0.5415	0.9210
Sample - 5	0.0714	0.0971	0.1275	0.0109	0.4565

The result is also visualized graphically [Fig – 3].

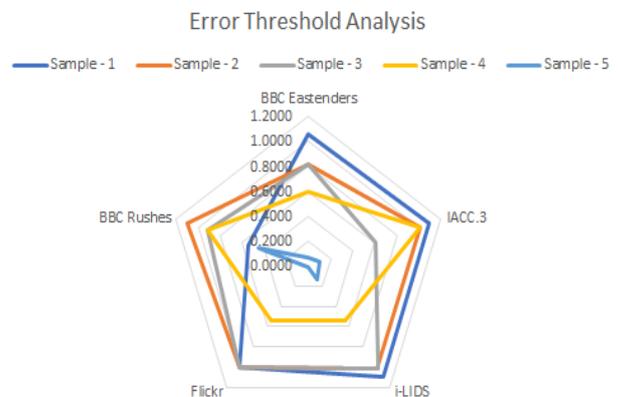


Fig. 3 Error Threshold Analysis

The detailed advantages of this process are elaborated in the prior section of this work.

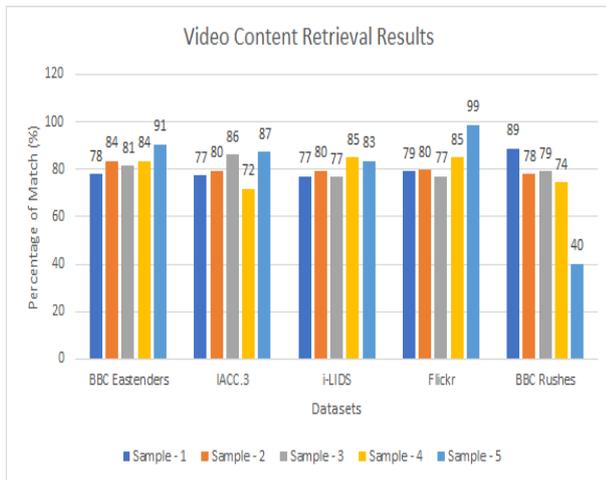
### C. Video Content Retrieval Accuracy

Third, the video content retrieval accuracy results are elaborated here [Table – 4]:

**TABLE IV**  
VIDEO CONTENT RETRIEVAL ANALYSIS BASED ON PERCENTAGE OF MATCHING

Searching Data	BBC Eastenders Dataset (%)	IACC.3 Dataset (%)	i-LIDS Dataset (%)	Flickr Dataset (%)	BBC Rushes Dataset (%)
Sample - 1	78	77	77	79	89
Sample - 2	84	80	80	80	78
Sample - 3	81	86	77	77	79
Sample - 4	84	72	85	85	74
Sample - 5	91	87	83	99	40

The result is also visualized graphically [Fig – 4].



**Fig. 4** Content Retrieval Accuracy

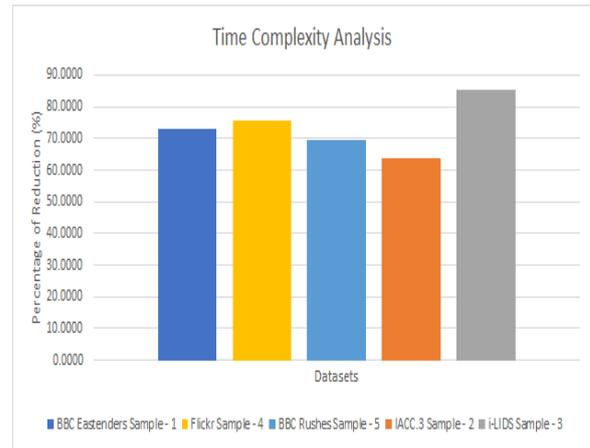
### D. Time Complexity Analysis

Finally, time complexity analysis results are elaborated here [Table – 5]:

**TABLE V**  
TIME COMPLEXITY ANALYSIS

Dataset Name	Content Name	Time Complexity without Proposed Algorithms (Sec)	Time Complexity with Proposed Algorithms (Sec)	Percentage of Reduction (%)
BBC Eastenders	Sample - 1	42.19	11.417	72.9365
IACC.3	Sample - 2	87.24	31.563	63.8186
i-LIDS	Sample - 3	59.66	8.6678	85.4709
Flickr	Sample - 4	44.80	10.98	75.4907
BBC Rushes	Sample - 5	85.60	26.1828	69.4109

The result is also visualized graphically [Fig – 5].



**Fig. 5** Percentage of Time Complexity Reduction

### VIII. CONCLUSION

Motivated by the popularity and growth of the video contents for various purposes, a good number of research attempts have tried to make the video retrieval process a less time complex process. Nonetheless, the advanced video content libraries demand video content-based searching, where based on the similarity of the input search video, the related video contents will be returns as search results. During this process, the searching video and the contents from the video library must be divided into key frames and these extracted key frames must be analysed and compared for similarity. For a video content bigger in length, this process can be highly time complex. Thus, reduction in time complexity without compromising on the accuracy of the search results is objective of recent research trends. This work mathematically proves that, reduction on the key frames without losing the information from the content can significantly reduce the time complexity of the retrieval process. Hence, this work deploys a novel approach to reduce the key frames using similarity analysis based on object contour, object displacement and finally using the pixel intensity rate of changes. The work also deploys CUDA based parallel processing methods to analyse the key frame similarities to reduce the time complexity to a greater extend. Further the similarity of the key frames extracted from the searching video and the library video contents are also analysed for generating the search results. The reduction of the key frames, the reduction of the searching time complexity and finally the accuracy of the video data retrieval process is highly satisfactory, and the improvement is nearly 70% over the parallel research outcomes. This work is a newer dimension on this research direction and makes the video content mining process ready for the mission critical data with reduced time complexity and higher accuracy.

### REFERENCES

- Hawalah, M. Fasli, "A multi-agent system using ontological user profiles for dynamic user modelling", Proc. IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Technol., vol. 1, pp. 430-437, Aug. 2011.
- A. Gomez-Urbe, N. Hunt, "The Netflix recommender system: Algorithms business value and innovation", ACM Trans. Manage. Inf. Syst., vol. 6, no. 4, 2016.
- A. Worthy, K. A. Byrne, S. Fields, "Effects of emotion on prospection during decision-making", Frontiers Psychol., vol. 5, pp. 591, Jun. 2014.



4. Smirnova, F. Vasile, Contextual sequence modeling for recommendation with recurrent neural networks, 2017, [online] Available: <https://arxiv.org/abs/1706.07684>.
5. H. Yin, B. Cui, L. Chen, Z. Hu, X. Zhou, "Dynamic user modeling in social media systems", ACM Trans. Inf. Syst., vol. 33, no. 3, 2015.
6. J. Etkin, A. P. Ghosh, "When being in a positive mood increases choice deferral", J. Consum. Res., vol. 45, no. 1, pp. 208-225, 2017.
7. J. Li, P. Ren, Z. Chen, Z. Ren, T. Lian, J. Ma, "Neural attentive session-based recommendation", Proc. ACM Conf. Inf. Knowl. Manage., pp. 1419-1428, 2017.
8. L. Ardissono, P. Torasso, "Dynamic user modeling in a Web store shell", Proc. 14th Eur. Conf. Artif. Intell., pp. 621-625, 2000.
9. M. J. Duque, C. Turla, L. Evangelista, "Effects of emotional state on decision making time", Procedia-Social Behav. Sci., vol. 97, pp. 137-146, Nov. 2013.
10. M. Schedl, H. Zamani, C.-W. Chen, Y. Deldjoo, M. Elahi, "Current challenges and visions in music recommender systems research", Int. J. Multimedia Inf. Retr., vol. 7, no. 2, pp. 95-116, 2018.
11. M. Scheutz et al., "Toward affective cognitive robots for human-robot interaction", Proc. Nat. Conf. Artif. Intell., pp. 61-66, 2005.
12. P. Covington, J. Adams, E. Sargin, "Deep neural networks for YouTube recommendations", Proc. 10th ACM Conf. Recommender Syst., pp. 191-198, 2016.
13. R. Cowie, E. Douglas-Cowie, K. Karpouzis, G. Caridakis, M. Wallace, S. Kollias, "Recognition of emotional states in natural human-computer interaction" in Multimodal User Interfaces, Berlin, Germany:Springer, pp. 119-153, 2008.