

A Review Of Video Retrieval Systems: Tiny Videos

Aditi Nain, Prof. K.S Bhagat, Dr.D.K Kirange
Student, EXTC Dept, J.T M College of Engineering, Maharashtra, India
Professor, EXTC Dept, J.T M College of Engineering, Maharashtra, India
HOD ,Computer Dept, J.T M College of Engineering, Maharashtra, India

Abstract: In this review paper, we summarize all the different video retrieval techniques available in the literature. The Tiny video is one of the simplest approaches that we find for retrieval of the videos from large dataset like youtube. We also present a summary table of all other methods along with their advantages and disadvantages. This paper will allow authors to come up with the best suiting retrieval system. We plan to work with tiny videos where kernels of videos are used to retrieve the video. This video retrieval system has combined approaches like content-based with semanticretrieval.

I. Introduction:

The tiny video is a technique of video retrieval, In the current scenario video retrieval, is very important. The effectiveness of a system of video retrieval depends to a great extent on the choice of the subjacent text and the components of searchpicture. The unique ownership of the video collections (e.g., Numbers, characters and temporal relations) suggest to examine the performance of these methods of retrieval in such multimode environment and to identify the relative importance of the subjacent components of recovery. In this article, we examine a variety of approaches to recovery of text/picture as well as their individual components in the context of the video of actuality.



Figure 1. Showing video retrieval system with three different search engines.

Numerous elements of recovery of text/picture were discussed in detail, including models of recovery, sources of text, methods of temporal expansion, methods of extension of requests, characteristics of the picture and of measurements of similarity. For every element, we perform a series of experiments to retrieve the video. In this paper, we discuss different video retrieval techniques and their merits and demerits.

II. Litrarure Review:

Karpenko et al. started upon his papers circa 2007 when YouTube and similar content distribution platforms were on the rise. He showcased that using a video classification algorithm, he could find related content for a database of videos without having to depend on user intervention. In 2009, he published another

paper detailing the use of multiple image sets for single videos to classify similar content, simplifying the need of scrubbing through entire videos for correlation to similar content. In 2011, he had improved his dataset of images and was downscaling them for compression, but still lacked the same level of detail recognition that is required for higher quality videos on the platform, hence making it a little obsolete at this day and age. In this year, Song et al. created an algorithm that performed multiple feature hashing and could accurately retrieve a near duplicate video from a set. Even though this was a novel approach towards content retrieval, it was a little too complex in its approach.

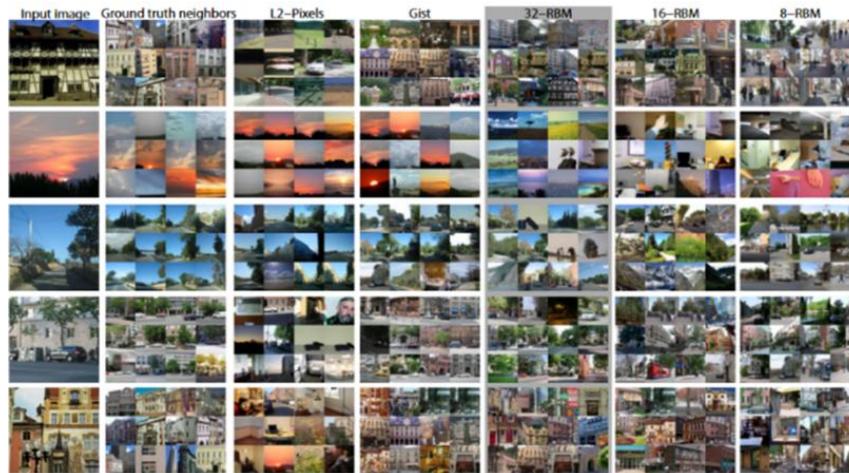


Figure 2. Shows 32RBM is best as compared to L2, Gist, and lower RBM

In 2013, Liu et al. had a similar technique of video retrieval in the form of NDV or near duplicate video. This algorithm showed a scalability issue. Douze et al., published a paper on Event Retrieval using circulant temporal encoding. This method was useful only for matching similar videos into clusters for specific events. Kazimierz had a paper on “Video Structure Analysis and Content-Based Indexing in the Automatic Video Indexer AVI”. It also showcased the various methods of video retrieval among the ones already published.

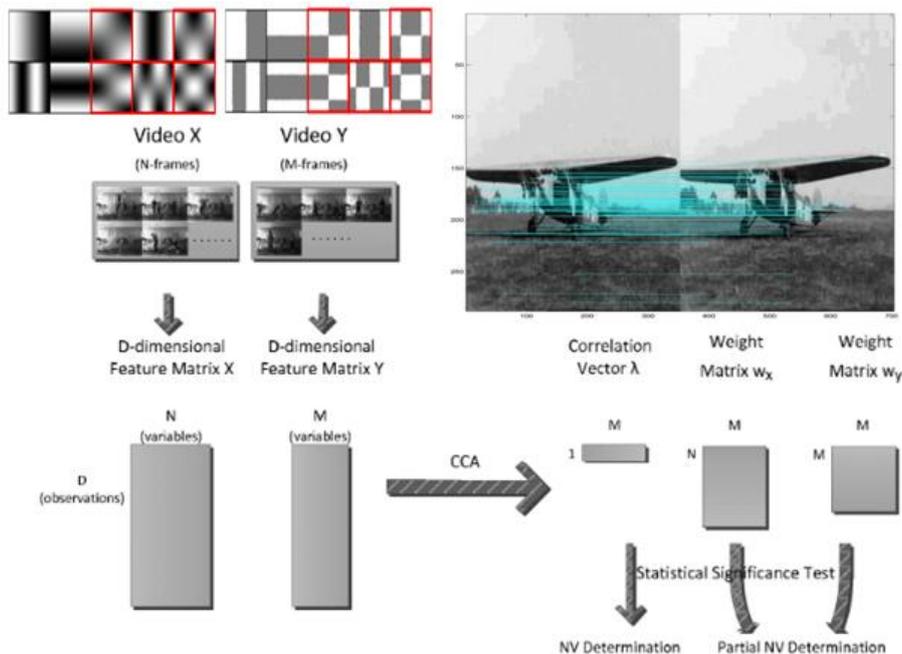


Figure 3. Showing x and y-axis affects retrieval

Karpenko et al in 2008 showed in that report, they have taken video samples of 40 x 30 pixels whereas our resolution on both the images as well as the videos is much higher, showing that it can be applied to a more modern dataset. Our dataset is limited to the amount of time it takes for computation but as we will be showing

you we can improve on this vastly since scaling is not an issue for this algorithm if we provide it enough computational power. Karpenko et al; 2009 worked on the similar algorithm. This paper showed the same limitation as for the previous paper of reduced resolution and hence inaccurate result fetching in case the resolution wasn't sufficient to "tag" a video. Also, we have expanded upon various other reports that showcase non-parametric search and keyframe-based video retrieval. The accuracy of our method is much higher and given the right amount of resources can be made better and faster in its search times as well as the number of hits it will generate. Karpenko et alin 2011 reported a high variance among the types of content presented but focuses on specific types, in this case, sports. Hence there is a bias present in the accuracy of certain videos which cannot be conclusively quantified. Liu et al in 2013 established the NDVR technique of retrieval is promising and is performed on a large dataset but is limited in application due to the complexity of its setup. Song et al in 2011 used a similar technique of NDVR but was limited to the amount of data required to perform an accurate match. It employed multiple feature hashing.

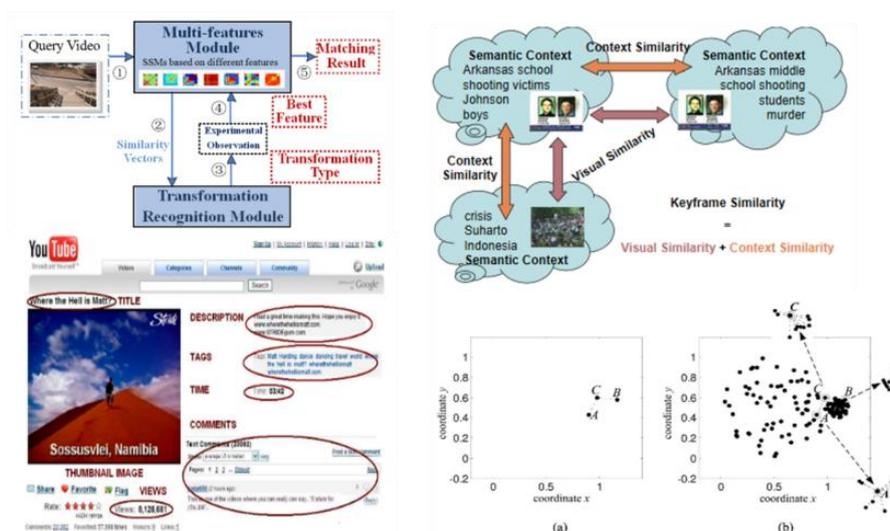


Figure 4. The importance of multimodal search. Context may match but still, retrieved video may not match

Revaud et al., 2013 designed a novel technique using circulant temporal encoding. It uses this technique to fetch videos that pertain to a set event that has a limited database. Kazimierz Choros, 2010. The article expands upon Automatic Scene Detection for tagging a video and also elaborates on the efficiency of this technique over traditional databasing. Choros et al; 2013. This is the same method applied to the previous article but it is applicable to TV sport news broadcasts rather than online content. Choros et al; 2010. This study uses temporal segmentation to classify content on the basis of information taken from the running scene. Y.J. Zhang et al; 2002. The article describes a similar method for pattern recognition to extract keyframes. It relies on flash detection to get accurate key frame images. It extracts multiple key frames instead of a single usable frame. Liu et al; 2006. This paper proposes an efficient story unit extraction method based on the result of unsupervised video scene clustering. In this method, cluster transition matrix is constructed to detect the key scene shots and extract the story unit. Wengang, 2003. The approach defines a scene as a sequence of semantically correlated shots near in time or location. The "shot cluster" is a basic unit of the scene and it defines the "shot cluster" as a group of shots similar in content and near in time or location. Zhou et al; 2009. The author details extraction of the local interest points with higher frequencies in a shot from its key frame to improve the performance of near-duplicate shot identification. Zhang et al; 2010. The proposed Self-Taught Hashing (STH) approach to semantic hashing is a general learning framework that consists of two distinct stages, first being the unsupervised learning of binary code. Second is the supervised learning of hash function. Yang et al; 2008. In this paper, different algorithms, including long term and short term, are proposed to utilize relevance feedback for refining the Laplacian media object space (LMOS) and the MMD semantic graph (MMDSG), mapping new items into the MMD semantic space and dynamically ranking search results to improve the retrieval performance.

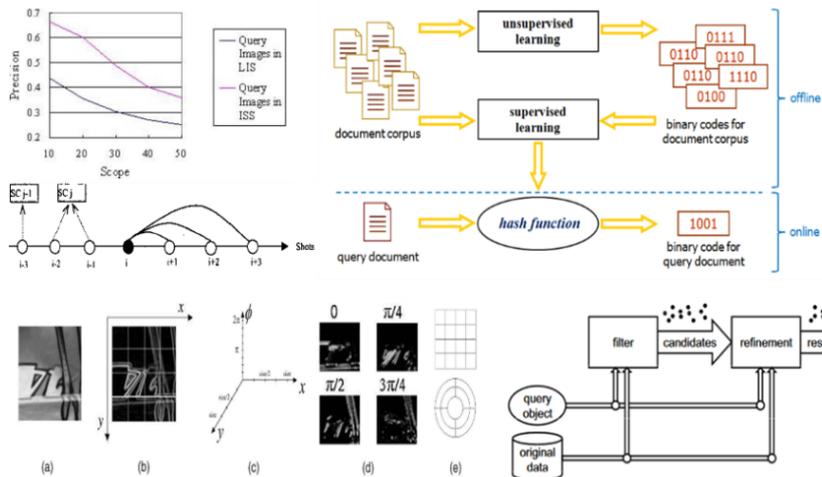


Figure 5. Shows impact of supervised learning, which helps in retrieval

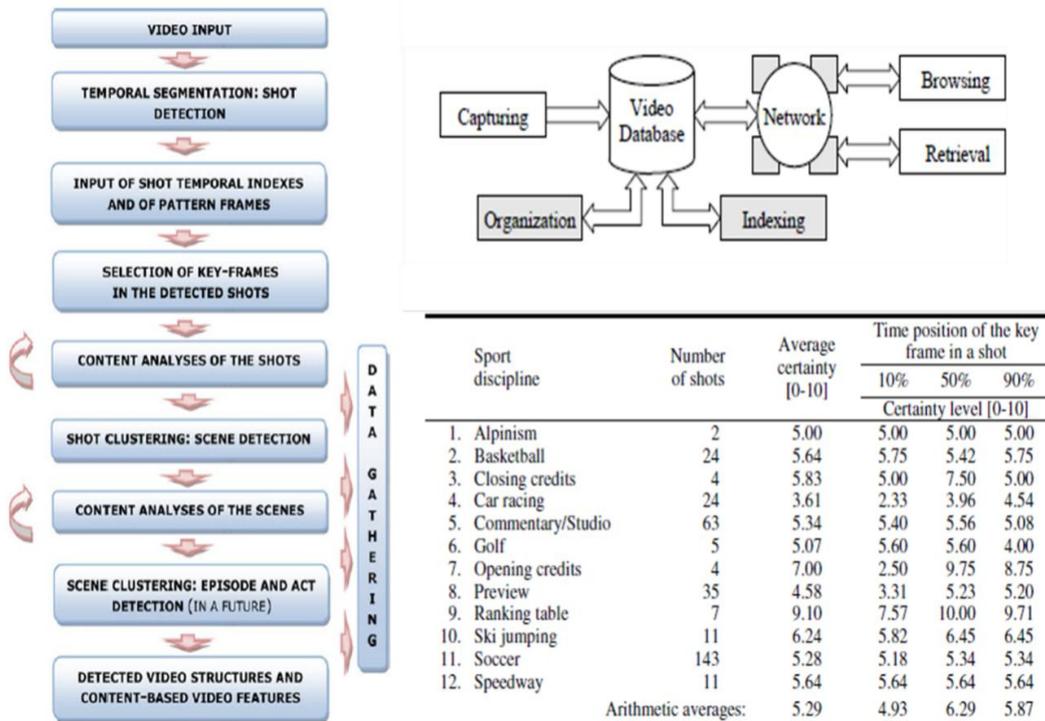


Figure 6. Shows effect of merging temporal retrieval with index-based retrieval

Z. Wu et al; 2009. This paper presents a Self-Similarity Matrix (SSM) based near-duplicate video matching scheme with a transformation recognition module. **X. Wu et al;** 2007. This article is described as follows: - The visual words are constructed by the offline quantization of keypoint descriptors. Under this representation, each keyframe is treated as a bag of visual words (BoW), analogous to the documents composed of text words. With BoW, it adopts the cosine similarity and language models respectively to evaluate the likelihood of near-duplicate between keyframes. **X. Wu et al;** 2009. Near-duplicate web videos are identical or approximately identical videos close to the exact duplicate of each other, which have a similar time duration/length, but different in file formats, encoding parameters, photometric variations, editing operations, and certain modifications.

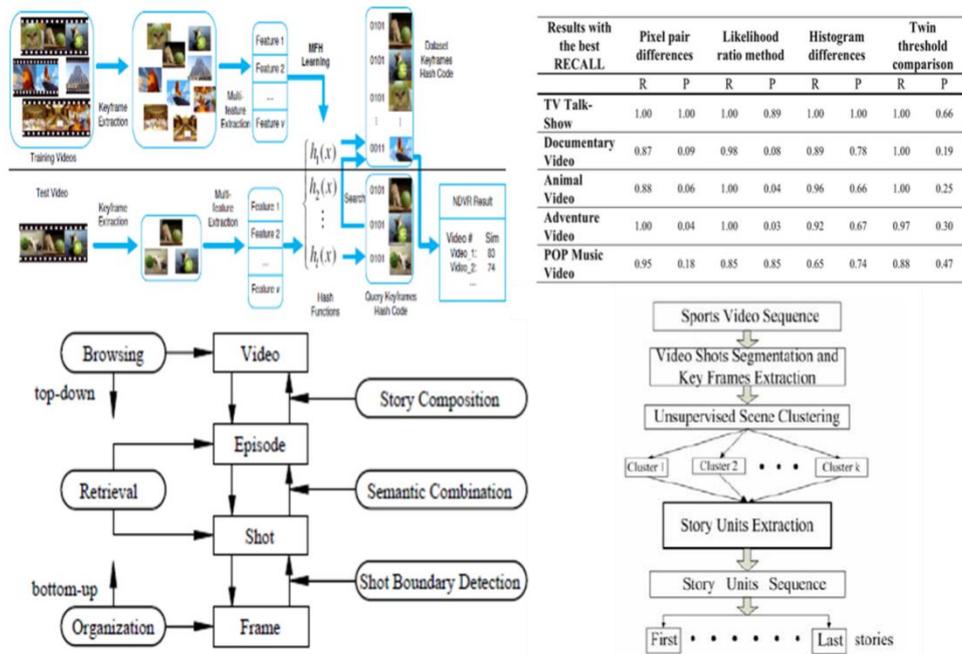


Figure 7. Different probability based retrieval models

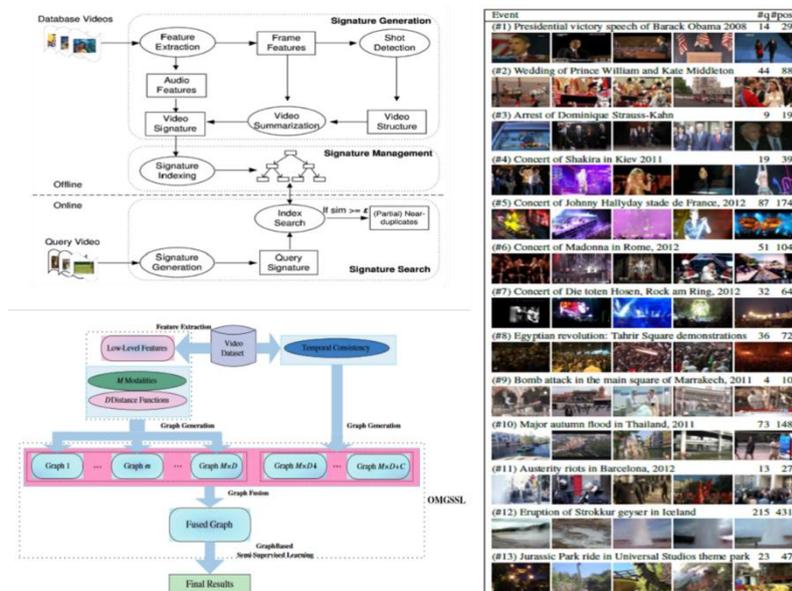


Figure 8. Shows probability based model, when merged with supervised learning, produces best results

Wichterich et al; 2008.In this article, the author proposes new techniques for dimensionality reduction of the Earth Mover’s Distance which significantly increases efficiency in query processing.**J Liu et al; 2011.**The article proposes Correlation-based Near-duplicate Video Retrieval (CNVR) based on information correlation analysis. It is designed to have strong tolerance to intensive content changes. Unlike SNVR, the nature of correlation analysis determines that in CNVR, thenon-trivial relationship between two near-duplicate videos has a great opportunity to be found even though the explicit content similarity is not high.**Weiss et al; 2009.**In this paper, we can see that the problem of finding the best code for a given dataset is closely related to the problem of graph partitioning and can be shown to be not hard.**M. Wang et al; 2009.**This paper proposes a similarity measure named neighborhood similarity which can explore this information. It consists of two components, namely, neighborhood sample similarity, and neighborhood label similarity, which are developed based on local sample distribution and local label distribution, respectively.**M. Wang et al; 2009.**The article proposes the OMG-SSL algorithm. Different from the existing graph-based learning techniques, which deal with only one graph, the OMG-SSL method optimally explores multiple complementary graphs in the manner of

semi-supervised learning. **Torralba et al;** 2008. In this paper, we leverage recent results in machine learning to learn compact binary codes that allow efficient retrieval. **X. Wu et al;** 2007. In this paper, the author proposes a hierarchical approach combining global signatures and local feature based pairwise comparison to detect near-duplicate web videos. **Mikolajczyk et al;** 2005. In this paper, the comparison is carried out for different descriptors, different interest regions, and for different matching approaches.

III. Discussions:

As a discussion, we present a summary table which concludes all video retrieval systems in terms of their pros and cons.

Citation	Methodology	Pros	Cons
Karpenko, Alexandre, and Parham Aarabi. "Tiny videos: Non-parametric content-based video retrieval and recognition." <i>Multimedia</i> , 2008. ISM 2008. Tenth IEEE International Symposium on. IEEE, 2008.	The article presents a method for compressing the temporal dimension nonuniformly using affinity propagation. This method outperforms temporal sampling at uniform intervals.	Can use the algorithm on large databases and the method has been validated for good accuracy.	The content used for dataset comparison is of significantly lower quality than expected hence limiting usability
Karpenko, Alexandre, and Parham Aarabi. "Tiny videos: a large dataset for image and video frame categorization." <i>Multimedia</i> , 2009. ISM'09. 11th IEEE International Symposium on. IEEE, 2009.	This paper presents a new method for video and image categorization based on a database of over 50,000 videos collected from YouTube and down-sampled to tiny size.	A similar approach to previous paper but uses many tiny images to categorize a specific video.	The content used for dataset comparison is of significantly lower quality than expected hence limiting usability
Karpenko, Alexandre, and Parham Aarabi. "Tiny videos: a large data set for nonparametric video retrieval and frame classification." <i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i> 33.3 (2011): 618-630.	Here a compact representation called "tiny videos" that achieves high video compression rates while retaining the overall visual appearance of the video as it varies over time is used. This is paired with AP clustering.	This multiplexes the methods of the previous two papers and therefore improves image categorization to a video vastly over either of them.	The content used for dataset comparison is of significantly lower quality than expected hence limiting usability
Liu, Jiajun, et al. "Near-duplicate video retrieval: Current research and future trends." <i>ACM Computing Surveys (CSUR)</i> 45.4 (2013): 44.	The article explores Near-Duplicate Video Retrieval in its current form and against similar methods available.	The method used involves data mining and hence shows promise in terms of how accurate it is.	Data mining is a long and tedious process and requires high storage and computation to sort at the end.
Song, Jingkuan, et al. "Multiple feature hashing for real-time large scale near-duplicate video retrieval." <i>Proceedings of the 19th ACM international conference on Multimedia</i> . ACM, 2011.	This article proposes a new hashing algorithm, namely Multiple Feature Hashing (MFH), for accurate and efficient NDVR based on multiple visual features.	This reduces the storage requirements of NDVR to an extent and makes it fetch more accurately.	Still tedious and requires high levels of computation to do effectively.
Revaud, Jérôme, et al. "Event retrieval in large video collections with circulant temporal encoding." <i>Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition</i> . 2013.	This article first contributes in the following manner by encoding the frame descriptors of a video into a temporal representation and to exploit the properties of circulant matrices to compare videos in the frequency domain.	Extremely useful for categorization of videos into singular events of large scale without much need for NDVR type comparisons.	Complexity and inefficiency are represented which at times will give errors in categorization.
Choroś, Kazimierz. "Video structure analysis and content-based indexing in the Automatic Video Indexer AVI." <i>Advances in Multimedia and Network Information System Technologies</i> . Springer Berlin Heidelberg, 2010. 79-90.	The article makes the use of Automatic Scene Detection and Automatic Video Indexing to group similar content together into a cluster that can be stored and fetched later.	Accurate scene detection allows for good categorization for fast paced events and sports scenes.	Shows errors in case video content have been altered slightly with extra information like text or resizing.
Choroś, Kazimierz. "Automatic detection of headlines in temporally aggregated TV sports news videos." <i>2013 8th International Symposium on Image and Signal Processing and Analysis (ISPA)</i> . IEEE, 2013.	The article makes the use of Automatic Scene Detection and Automatic Video Indexing to group similar content together into a cluster that can be stored and fetched later. This study applies to TV video.	Accurate scene detection allows for good categorization for fast paced events and sports scenes.	Shows errors in case video content have been altered slightly with extra information like text or resizing.
Choroś, Kazimierz, and Piotr Pawlaczyk. "Content-based scene detection and analysis method for automatic classification of TV sports news." <i>International Conference on</i>	The article makes the use of Automatic Scene Detection and Automatic Video Indexing to group similar content together into a cluster that can	Accurate scene detection allows for good categorization for fast paced events and sports scenes.	Shows errors in case video content have been altered slightly with extra information like text or resizing.

Rough Sets and Current Trends in Computing. Springer Berlin Heidelberg, 2010.	be stored and fetched later. This study applies to TV video.		
Zhang, Yu Jin, and H. B. Lu. "A hierarchical organization scheme for video data." Pattern Recognition 35.11 (2002): 2381-2387.	This paper proposes a hierarchical structure and a process scheme for organizing video data to facilitate indexing, browsing, and querying.	Efficient in similar shot detection due to keyframe extraction.	High complexity due to a number of keyframes used for video comparison.
Liu, Chunxi, et al. "Extracting story units in sports video based on unsupervised video scene clustering." 2006 IEEE International Conference on Multimedia and Expo. IEEE, 2006.	A scene clustering method for sports video is adapted to automatically categorize the video shots into several scenes. Then, the clustering results are modeled by a transition matrix and then key scene shots are used to extract the story units.	Highly accurate due to these of several scenes from a large dataset and is effective in extracting similar videos from a sample.	High computation and storage requirements due to a large number of scenes being captured and compared against the given sample.
Wengang, Cheng, and Xu De. "A novel approach of generating video scene structure." TENCON 2003. Conference on Convergent Technologies for the Asia-Pacific Region. Vol. 1. IEEE, 2003.	We segment the video into shots with an improved twin-comparison method. Shots are grouped into shot clusters using a sliding shots window method, which can greatly reduce computation effort and time.	Correlation is used to generate video scene structure and hence the given method promises efficiency.	High complexity in implementation of the clustering method.
Zhou, Xiangmin, et al. "An efficient near-duplicate video shot detection method using shot-based interest points." IEEE Transactions on Multimedia 11.5 (2009): 879-891.	The basic idea is to eliminate the local descriptors with lower frequencies among the selected video frames from a shot to ensure that the shot representation is compact and discriminative.	Uses key descriptors from keyframe and hence reduces the storage requirement.	Comparatively inaccurate due to reduced data density and causes errors while fetching.
Zhang, Dell, et al. "Self-taught hashing for fast similarity search." Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. ACM, 2010.	This paper proposes a novel Self-Taught Hashing approach to semantic hashing: 1 - find the optimal 1-bit binary codes for all documents in the given corpus via unsupervised learning, 2 - train 1 classifiers via supervised learning to predict the 1-bit code for any query document unseen before.	Experiments show that the proposed approach using binarized Laplacian Eigenmap and linear Support Vector Machine outperforms state-of-the-art techniques significantly.	Some recently proposed techniques can generate high-quality codes for documents known in advance, obtaining codes for previously unseen documents is a very challenging problem.
Yang, Yi, et al. "Harmonizing hierarchical manifolds for multimedia document semantics understanding and cross-media retrieval." IEEE Transactions on Multimedia 10.3 (2008): 437-446.	The characteristics of media objects propagate along the MMD semantic graph and an MMD semantic space is constructed to perform cross-media retrieval. Different methods are proposed to realize this.	Proposed approaches are effective in video characterization and retrieval.	High complexity reduces the usage scenario for general everyday video search.
Wu, Zhipeng, Shuqiang Jiang, and Qingming Huang. "Near-duplicate video matching with transformation recognition." Proceedings of the 17th ACM international conference on Multimedia. ACM, 2009.	This paper focuses on the fact that a certain kind of feature may perform excellently to deal with one type of transformation while not be that good on another.	Improvement on existing NDVR techniques and better accuracy in detection.	Small sample size hence quantifying the effectiveness is a little difficult.
Wu, Xiao, Wan-Lei Zhao, and Chong-Wah Ngo. "Near-duplicate keyframe retrieval with visual keywords and semantic context." Proceedings of the 6th ACM international conference on Image and video retrieval. ACM, 2007.	This paper proposes a novel NDK retrieval approach by exploring both visual and textual cues from the visual vocabulary and semantic context respectively.	Use of keyframe comparison is an accurate way to achieve near duplicate video retrieval.	Suffers from expensive computation requirement due to the need of nearest online neighbor search.
Wu, Xiao, et al. "Real-time near-duplicate elimination for web video search with content and context." IEEE Transactions on Multimedia 11.2 (2009): 196-207.	This paper combines the contextual information from time duration, the number of views, and thumbnail images with the content analysis derived from color and local points to achieve real-time near-duplicate elimination.	Well tested on a large dataset and shows good accuracy compared to other NDVR methods.	The slight loss in performance.
Wichterich, Marc, et al. "Efficient EMD-based similarity search in	The Earth Mover's Distance (EMD) was developed in	The techniques are fully flexible in the number of	High complexity in employing the technique

multimedia databases via flexible dimensionality reduction." Proceedings of the 2008 ACM SIGMOD international conference on Management of data. ACM, 2008.	computer vision as a flexible similarity model that utilizes similarities in feature space to define a high-quality similarity measure in feature representation space.	reduced dimensions, which is a novel feature in approximation techniques for the EMD.	to a viable dataset.
Liu, Jiajun, et al. "Correlation-based retrieval for heavily changed near-duplicate videos." ACM Transactions on Information Systems (TOIS) 29.4 (2011): 21.	The intuition is that near-duplicate videos should preserve strong information correlation in spite of intensive content changes.	More effective retrieval with stronger tolerance is achieved.	Experimental results show inaccuracies in video retrieval.
Weiss, Yair, Antonio Torralba, and Rob Fergus. "Spectral hashing." Advances in neural information processing systems. 2009.	Semantic hashing seeks compact binary codes of datapoints so that the Hamming distance between code words correlates with semantic similarity.	Both learning the code and applying it to a novel point is extremely simple.	Better than existing methods but still misses the mark in terms of expected accuracy.
Wang, Meng, et al. "Beyond distance measurement: constructing neighborhood similarity for video annotation." IEEE Transactions on Multimedia 11.3 (2009): 465-476.	This article proposes a novel neighborhood similarity measure, which explores the local sample and label distributions.	It is experimentally proven that this is superior to distance-based similarity.	Relies on machine learning hence computationally expensive and difficult to implement.
Wang, Meng, et al. "Unified video annotation via multigraph learning." IEEE Transactions on Circuits and Systems for Video Technology 19.5 (2009): 733-746.	OMG-SSL integrates multiple graphs into a regularization framework to sufficiently explore their complementation.	The method can be implemented through an efficient iterative process.	Unification while a novel approach tends to suffer from errors in implementation for edge cases.
Torralba, Antonio, Rob Fergus, and Yair Weiss. "Small codes and large image databases for recognition." Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008.	This approach uses recently developed machine learning techniques to convert the Gist descriptor to a compact binary code, with a few hundred bits per image.	Using the simple nearest neighbor technique yields powerful recognition results.	Highly computationally intensive.
Wu, Xiao, Alexander G. Hauptmann, and Chong-Wah Ngo. "Practical elimination of near-duplicates from the web video search." Proceedings of the 15th ACM international conference on Multimedia. ACM, 2007.	This paper outlines ways to cluster and filter out the near-duplicate video using a hierarchical approach.	Relatively small computational cost to reduce the amount of redundant results in the search output.	Hierarchical approach while thenovel is slow and tedious to implement for large datasets.
Mikolajczyk, Krystian, and Cordelia Schmid. "A performance evaluation of local descriptors." IEEE transactions on pattern analysis and machine intelligence 27.10 (2005): 1615-1630.	This evaluation uses as criterion recall with respect to precision and is carried out for different image transformations.	Implementing a SIFT descriptor outperforms the original method.	Lacks versatility and needs adjustment as per the given dataset.

IV. Conclusions:

Our work reviews upon multiple methods of video classification, retrieval, and recognition. Though image retrieval problem exists for a longer duration, most of above mentioned technique address to solve this issue. Some of them have successfully achieved their goal in terms of either speed or accuracy. Authors suggest the alternative approach of tiny videos which optimizes speed and accuracy both at a time. As a concluding remark authors would like to conclude that spatial domain approach is faster on other hand frequency domain approaches are very accurate.

References

- [1] Zhang, Hong Jiang, et al. "An integrated system for content-based video retrieval and browsing." Pattern recognition 30.4 (1997): 643-658.
- [2] Jiang, Yu-Gang, Chong-Wah Ngo, and Jun Yang. "Towards optimal bag-of-features for object categorization and semantic video retrieval." Proceedings of the 6th ACM international conference on Image and video retrieval. ACM, 2007.
- [3] Hanjalic, Alan, Reginald L. Lagendijk, and Jan Biemond. "Automated high-level movie segmentation for advanced videoretrieval systems." IEEE transactions on circuits and systems for video technology 9.4 (1999): 580-588.
- [4] Snoek, Cees GM, and Marcel Worring. "Concept-based video retrieval." Foundations and Trends in Information Retrieval 2.4 (2008): 215-322.
- [5] Naphade, Milind R., et al. "Probabilistic multimedia objects (multijet): A novel approach to video indexing and retrieval in multimedia systems." Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on. IEEE, 1998.
- [6] Amir, Arnon, et al. "IBM research TRECVID-2003 video retrieval system." NIST TRECVID-2003 (2003).
- [7] Aslandogan, Y. Alp, and Clement T. Yu. "Techniques and systems for image and video retrieval." IEEE transactions on Knowledge and Data Engineering 11.1 (1999): 56-63.
- [8] Geetha, P., and Vasumathi Narayanan. "A survey of content-based video retrieval." (2008).
- [9] Snoek, Cees GM, et al. "Adding semantics to detectors for video retrieval." IEEE Transactions on multimedia 9.5 (2007): 975-986.
- [10] Hauptmann, Alexander, et al. "Can high-level concepts fill the semantic gap in video retrieval? A case studies with broadcast news." IEEE transactions on multimedia 9.5 (2007): 958-966.

- [11] Dimitrova, Nevenka, and Mohamed S. Abdel-Mottaleb. "Video retrieval of MPEG compressed sequences using dc and motion signatures." U.S. Patent No. 5,870,754. 9 Feb. 1999.
- [12] Hu, Weiming, et al. "Semantic-based surveillance video retrieval." *IEEE Transactions on image processing* 16.4 (2007): 1168-1181.
- [13] Sebe, Nicu, et al. "The state of the art in image and video retrieval." *International Conference on Image and Video Retrieval*. Springer Berlin Heidelberg, 2003.
- [14] Zhang, HongJiang, John YA Wang, and Yucel Altunbasak. "Content-based video retrieval and compression: A unified solution." *Image Processing, 1997. Proceedings., International Conference on*. Vol. 1. IEEE, 1997.
- [15] Del Bimbo, Alberto. *Visual information retrieval*. Morgan and Kaufmann, 1999.
- [16] Yan, Rong, Jun Yang, and Alexander G. Hauptmann. "Learning query-class dependent weights in automatic video retrieval." *Proceedings of the 12th annual ACM international conference on Multimedia*. ACM, 2004.
- [17] Yan, Rong, Alexander G. Hauptmann, and Rong Jin. "Negative pseudo-relevance feedback in content-based video retrieval." *Proceedings of the eleventh ACM international conference on Multimedia*. ACM, 2003.
- [18] Hauptmann, Alexander, Rong Yan, and Wei-Hao Lin. "How many high-level concepts will fill the semantic gap in news video retrieval?." *Proceedings of the 6th ACM international conference on Image and video retrieval*. ACM, 2007.
- [19] Chua, Tat-Seng, and Li-Qun Ruan. "A video retrieval and sequencing system." *ACM Transactions on Information Systems (TOIS)* 13.4 (1995): 373-407.
- [20] Song, Jingkuan, et al. "Multiple feature hashing for real-time large scale near-duplicate video retrieval." *Proceedings of the 19th ACM international conference on Multimedia*. ACM, 2011.
- [21] Campbell, Murray, et al. "IBM Research TRECVID-2007 Video Retrieval System." *TRECVID*. 2007.
- [22] Veltkamp, Remco, Hans Burkhardt, and Hans-Peter Kriegel, eds. *State-of-the-art in content-based image and video retrieval*. Vol. 22. Springer Science & Business Media, 2013.
- [23] Orsolini, Garry S., and Richard D. Bowers. "Audio/video retrieval system that uses keyword indexing of digital recordings to display a list of the recorded text files, keywords and time stamps associated with the system." U.S. Patent No. 5,794,249. 11 Aug. 1998.
- [24] Ponceleon, Dulce, et al. "Key to effective video retrieval: effective cataloging and browsing." *Proceedings of the sixth ACM international conference on Multimedia*. ACM, 1998.
- [25] Mc Donald, Kieran, and Alan F. Smeaton. "A comparison of score, rank, and probability-based fusion methods for video shot retrieval." *International Conference on Image and Video Retrieval*. Springer Berlin Heidelberg, 2005.
- [26] Hori, Tetsuro, and Kiyoharu Aizawa. "Context-based video retrieval system for the life-long applications." *Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*. ACM, 2003.
- [27] Li, Ze-Nian, Osmar R. Zaiane, and Zinovi Tauber. "Illumination invariance and object model in content-based image and video retrieval." *Journal of Visual Communication and Image Representation* 10.3 (1999): 219-244.
- [28] Shah, Ankit, and Ninad Mehendale. "Decrypted Steganography." *International Journal of Scientific & Engineering Research* 4 (2013): 1154-1156.
- [29] Mehendale, Ninad Dileep, and Snehal Ajit Shah. "Image fusion using adaptive thresholding and cross filtering." *Communications and Signal Processing (ICCSP), 2015 International Conference on*. IEEE, 2015.
- [30] Karpenko, Alexandre, and Parham Aarabi. "Tiny videos: a large data set for nonparametric video retrieval and frame classification." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.3 (2011): 618-630.
- [31] Karpenko, Alexandre, and Parham Aarabi. "Tiny videos: Non-parametric content-based video retrieval and recognition." *Multimedia, 2008. ISM 2008. Tenth IEEE International Symposium on*. IEEE, 2008.
- [32] Karpenko, Alexandre, and Parham Aarabi. "Tiny videos: a large dataset for image and video frame categorization." *Multimedia, 2009. ISM'09. 11th IEEE International Symposium on*. IEEE, 2009.
- [33] Lau, Nelson C., et al. "An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*." *Science* 294.5543 (2001): 858-862.
- [34] Choroś, Kazimierz. "Video structure analysis and content-based indexing in the Automatic Video Indexer AVI." *Advances in Multimedia and Network Information System Technologies*. Springer Berlin Heidelberg, 2010. 79-90.
- [35] Park, Dennis, and Deva Ramanan. "Articulated pose estimation with tiny synthetic videos." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2015.
- [36] Yoshida, Kenichi, and Noboru Murabayashi. "Tiny LSH for content-based copied video detection." *Applications and the Internet, 2008. SAINT 2008. International Symposium on*. IEEE, 2008.