A Novel Clustering Based Near Duplicate Video Retrieval Model

Dr Rajesh Doss

Assistant Professor in Computer Science, National Defence Academy (NDA), Pune -411023,
extrajesh@gmail.com

ABSTRACT

There are various video related tools, such as video sharing, recording, marketing, and consulting. Ironically the video-related tasks are uploading, sharing, posting and searching the video content as per the need and requirement. There is an increase in the usage of video posting, sharing and uploading upon the social media for various purposes. The increase in the usage of near-duplicate videos (NDVs) created on the Internet are in various forms, such as simple reformatting, to specific purchases, transformations, editions, and mixtures of distinct impacts. Joint multi-view hashing is a promising alternative for translating videos into compact and low-dimensional binary codes, inspired by observing multiple retrieval models. With this form of hashing video storage and time usage in the recovery process is increased. The main purpose of this work is to build a novel clustering model to maintain multiple feature structures and to reduce the retrieval process's time consumption.

Keywords: Clustering, Histogram Joint Multi-view, Near Duplicate, Video Retrieval

1. INTRODUCTION

The videos are the part of very strong and communicative media that are able to capture and present data. The frequency in the utilization of digital videos has guided towards the development of advanced research factors for efficient viewing, annotation and video data recovery. Video retrieval involves how to return comparable video clips (or events, shots and frames) to a user who has submitted a video request. There are two key groups of current jobs. One is to first retrieve main frames from the video information, then use picture recovery approaches to indirectly obtain the video information. While easy to implement, there is the obvious problem of losing the temporal dimension. The other approach integrates movement data (sometimes tracking objects) into the retrieval process. Although this is a better approach, the computationally expensive job requires movement evaluation. If object trajectories are to be followed this is harder [1].

Consumers are more prone to content-based searches than text-based searches. Those help search, index and rate the database according to human visual perception. The amount of time that the user spends filming, editing, uploading, scanning, and watching video has risen to a large or massive level. Massive publishing and distribution of videos gives rise to an already large amount of almost duplicate content. It includes pressing criteria as a major role in assignments such as video search, defense of video rights, video recommendations, and much more. Consequently, processing of almost duplicate video has drawn a lot of attention lately. And work into a range of clustering methods for detecting near-detection is also important. [2].

Duplicates in this study refer to the videos that are semantically and visually similar. The meaning of Video duplicity is said to have the same features like genre, story, scenario etc. whereas a duplicate image shares exactly the same semantics and scenes with an origin. The image duplication varies only in visual presentations. The video segment collectively obtained from another file, usually by various transformations such as addition, elimination, modification of colour, contrast and encoding. The same scene can differ significantly close to doubling stocks. Near-duplicate videos show slight misalignments. Near-duplicate videos typically present large semantic and visual misalignment at the level of entire videos. [3]. This paper further continues with the Section II as Background work , Section III as the Review for the proposed system model and its performance analysis
.

2. RELATED STUDY

The Emerging online video related technologies such as video sharing, video streaming, video reviews are increasingly evolving the user habits and finding their participation in video related behaviors such as editing, uploading, searching, downloading, and viewing. The Data found in the recent study conducted by one of the most leading digital environments measured as 76.8% of video transactions utitlized within a month. There was another survey that projected that in the current Pademic days, Internet viewers were fully dependent over the online videos and watched 14.8 billion videos online, with an average view count of 101 videos and an average viewing time of 356 minutes per user. It also indicates clearly the growing demand for online videos. The strong evidence have provoked that the view count grew by 4 percent and 15 percent over the average view time before pre-pandemic situations. We have referred and used the Near-Duplicate Video Retrieval (NDVR) as a definite role that occurs in various modern video applications. These extracts are acting as an emerging as a subject of great interest for this study. This paper is provided with a description and its insightful discussions relating its size, efficiency and accuracy. The Progress towards this research in the recent is wide and successfully applied in various sectors and domains.[5].

The rapid growth in the usage of video-related apps and their services are contributing continuously due to exponential rise of the online video content. According to survey as mentioned earlier on the video utilization, the amount of total internet videos have increased from 12,677,063,000 of Feb 2020 to 16,831,607,000 of May 2020, representing an increase of 17 percent in 3 months.The average length per clip increased from 3.2 minutes to 3.5 minutes in the same time period. Both figures are expected to have continous increase to rise. The Mainstream media evidently moves to the Internet, with Web apps and providers providing their video goods to customers. A few names are, the official artists' outlets are posting music videos on YouTube, TV shows are broadcasting on the TV broadcast company's websites, film trailers are being previewed on websites such as the Avatar film website. The editing and republishing of modified versions of the original video is often common with the author, company or business entity of a third party.[6].

At the other hand, several publishers or individuals are permitted to obtain and publish identical videos about the same event. All these factors contribute to a large percentage of almost duplicates in online videos. There is a sudden increase in the NDVs ratio upto 93 percent based on some of the user queries searched from the Internet.The presence of vast NDV information puts heavy demands on Near-Duplicate Video Retrieval (NDVR), as it is crucial for many new applications including copyright infringement detection, video monitoring, web video re ranking, video recommendations, etc.A common scenario for naming a few may be when a content owner releases a copyright video on their own YouTube channel and tries to enforce its copyright rights by identifying and removing infringing copies of their original YouTube version.Another example may include a company that has invested in a TV commercial, having to verify whether the commercial is broadcast for the right counts during the appropriate time period. All these tasks were solved based on the NDVR techniques in order to achieve to results automatically.[8].

3. SYSTEM MODEL

The algorithm presents a simple way to categorize a given dataset as a parameter in a certain number of clusters, k clusters. The following steps are taken by the algorithm:

[1] Calculate the histogram finding the differences between the frames.

[2] Find the average of differences and consider it as threshold such that the difference
    of frames having greater than the threshold is selected as a cluster center.

[3] Assign each pattern to the representative cluster center.

[4] Using the current membership of the cluster, the cluster centers could be recomputed.

[5] If you wish not to attain a convergence criterion, go to step 2.

[Typical the convergence conditions are: no. (to be minimum) pattern re-assignment
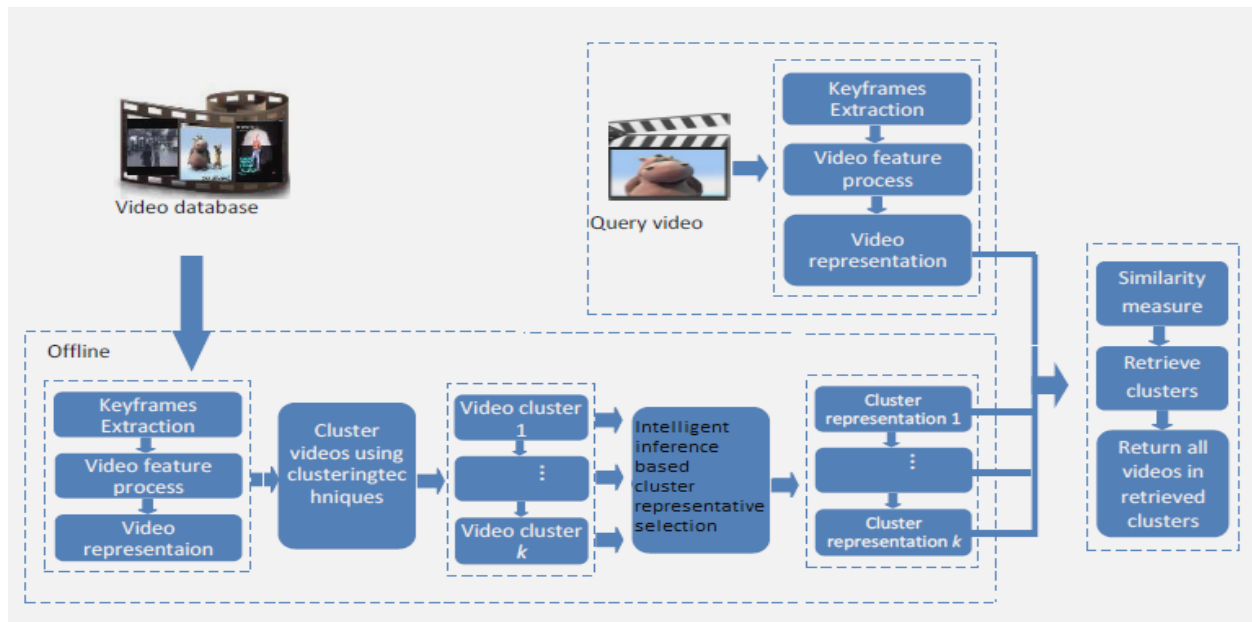 to new cluster centers, or limited squared error reduction.]

Figure 1.Proposed clustering NDVR framework

Table 1 Cluster centroid in iteration 1

	Centroid
Cluster 1	.0769 .0 .0 .0 .0 .0 .0 .0769 .1538 .0 .1538 .0 .0 .0769 .0 .0 .0 .0 .0 .0 .0 .0 .4615
Cluster 2	.1633 .0408 .0612 .1224 .0816 .0 .0 .0612 .0 .0 .0408 .0408 .0204 .0816 .0 .0 .0612 .0204 .0204 .102 .0 .0612 .0 .0204

Table 2 Video distance to cluster centroids in iteration 1

	croped_mini_subject_1_man	Croped_mini_subject_2_microwave	Croped_Mini_subject_3_mobile	Croped_Mini_subject_4_walk	Croped_Mini_subject_5_cleanwhiteboard
Cluster 1	.415	.524	.919	.73	.6204
Cluster 2	.48	.0	.0	.51	.534

The distance of all videos is determined to all new cluster centroids that are chosen as a threshold-based representative to further obtain the results. As a Next step, the

algorithm allocates the remaining video to the nearest cluster by the claws of the Euclidean distance between the video and the cluster centroid. As per the distance table of Euclid, Croped_mini_subject_5_ cleanwhiteboard is closest to the cluster 2, where the distance is .534.Croped_mini_subject_5_ cleanwhiteboard is assigned to the cluster 2 and then repeated in the same way, Croped_mini_ subject_3_microwave is assigned to cluster 1. until all the assignments are done. After the assignments, the center of each cluster has to be recomputed.

Table 4 Reassign videos to clusters

	Cluster 1	Cluster 2
Videos	Croped_mini_ subject_3_mobile	Croped_mini_ subject_5_ cleanwhiteboard

Continue repeating the operation in iteration 1 until it reaches or converges to the specified number of iterations. There are only two iterations and two clusters in this case, and the final converged cluster result has been achieved.

Table 5 Final clustering result

	Cluster 1	Cluster 2
Videos	Croped_mini_ subject_3_mobile	Croped_mini_ subject_5_ cleanwhiteboard

The proposed algorithm is a fine tuned, user friendly and very easy to implement. The time complexity is calibrated as O (nkl), where n depicts the size of the dataset, k is the number of clusters, and l is the number of iterations for the algorithm to converge. Predefining the number of clusters k is mandatory. In partial, we can random select or pick the minimum number as the k value to address the clusters. The threshold plays a prominent role in determining the head of the cluster, measuring their difference between histograms among two frames and taking the average as threshold. The Euclidean distance was applied to measure based on the nature of data.This algorithm proves that it reduces the number of clusters and gives more accurate cluster head in the lower number of iterations. This algorithm has also proved the reduction rates after raising the Time and space complexity.

Performance Evaluation

The ordinal function, global features, and the local features will be evaluated respectively as they have different features. They have different effects on NDV recovery. Traditional NDVR, NDVR based clustering and the proposed NDVR based smart inference are compared for review.The aim of this evaluation is to demonstrate that the proposed clustering-based NDV retrieval using different features is promising compared to traditional clustering-based and NDV retrieval, as the speed of retrieval is greatly reduced while the accuracy of retrieval plays comparable or better.

Throughout the experiment, 500 videos are chosen from the CVonline image dataset. The 500 experimental dataset consists of 356 near duplicate videos and 144 noise videos. Instead of comparing the question video with all the videos on the server, only cluster members / centroids will be compared.Nevertheless, the assessment will demonstrate how the proposed clustering-based retrieval affects the retrieval accuracy. Since clusters are represented by representative videos or centroids with a query video, only 92 cluster representations / centroids need to be compared instead of comparing all 500 videos.We can observe that the clustering process decreases the size of the offline dataset dynamically in comparison with the online version after applying the indexing structure. While there is no substantial difference in average precision recall accuracy via t-test statistical review, at certain points of recall the proposed smart inference clustering-based technique is substantially more accurate, whereas the recovery speed is more than 5 times faster when all 1000 queries are processed.
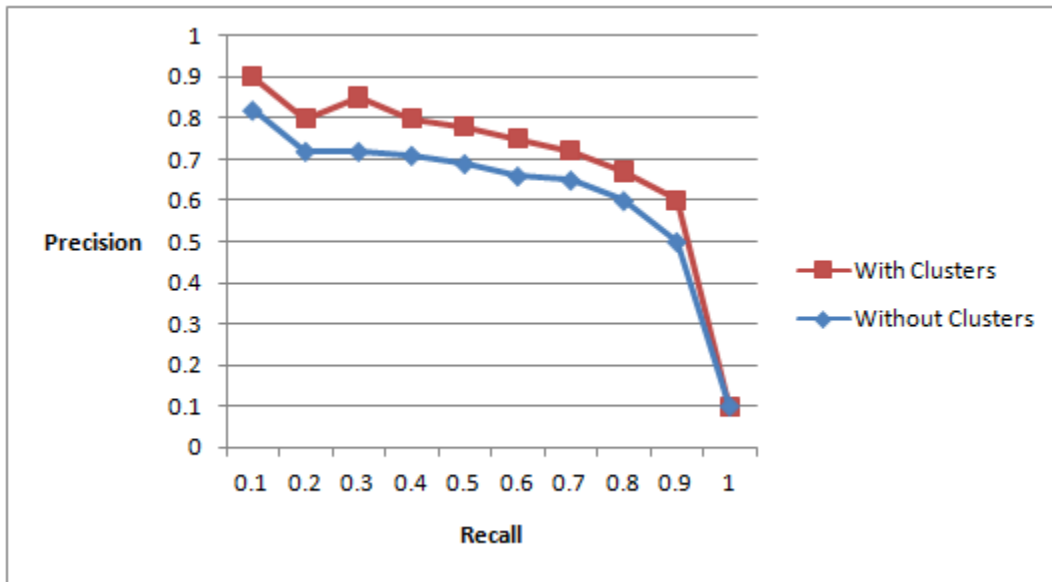
Figure 2.Precision-recall of Proposed-NDVR by using edge histogram
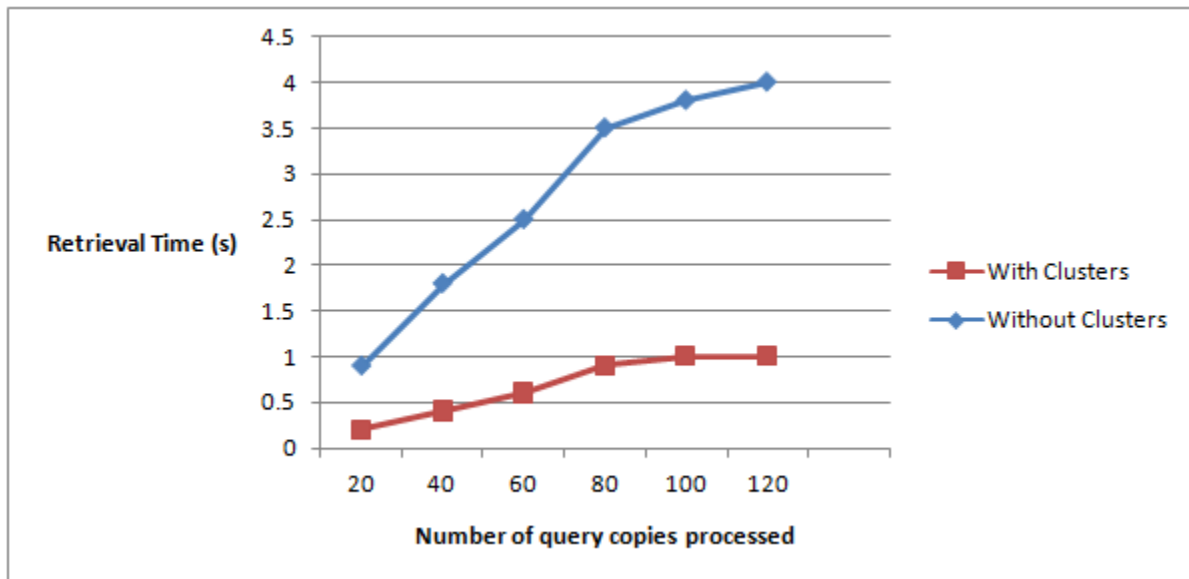

Figure 3.Time consumed corresponding to the number of query copies processed by using edge histogram
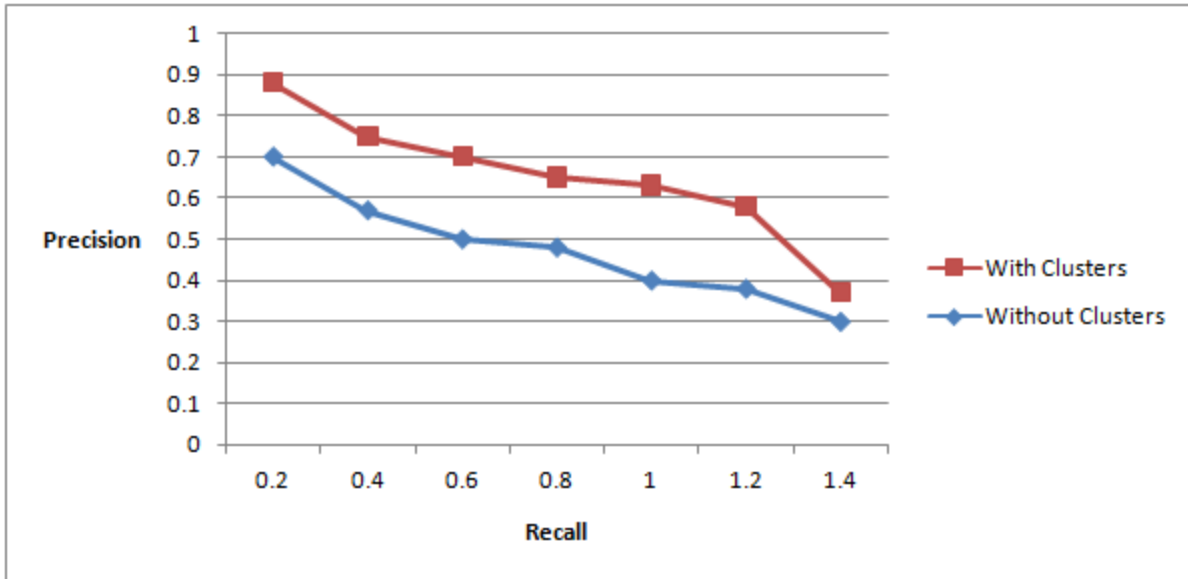
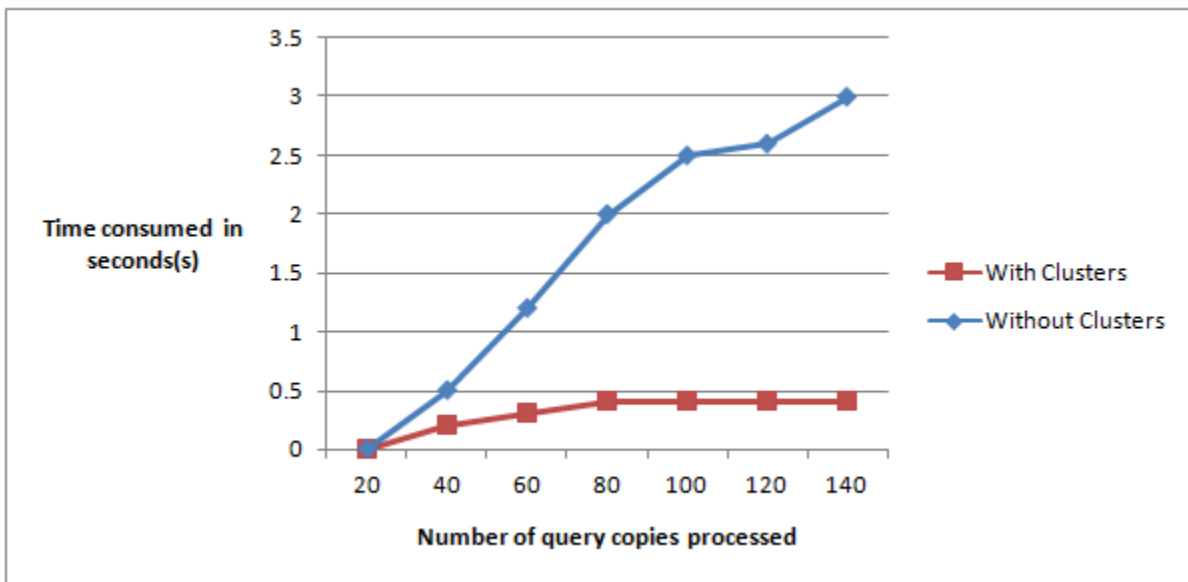Figure 4.Precision-recall of Proposed-NDVR by using scalable color

Figure 5.Time consumed corresponding to the number of query copies processed by using scalable color

Figure 3 shows the performance by using texture (edge histogram) in terms of retrieval accuracy and speed respectively. Figure 4 and Figure 5 shows Proposed-NDVR performance by using color (scalable color) in terms of retrieval accuracy and speed respectively.

4. CONCLUSION

The efficiency of the Proposed-NDVR is with a pre-processing clustering stage is very productive in the bi-fold stages. They work better in terms of recovery, more cost-effective on one side and provides an higher recovery accuracy in a precision-recall system on the other side..This paper helps to study the recovery in integrating the intelligent inferences where the clustering method is also compared with the recovery of NDVR based clustering and conventional NDVR. The output has resulted such that the NDVR based on the proposed clustering outperforms that of other clustering-methods. It has overcome

the traditional NDVR using common global and local features of their accuracy and recovery rates. By projecting the outcomes, it shows that improving NDVR by clustering using common global and local features is feasible on the dataset.

REFERENCES

1. Kordopatis-Zilos, Giorgos & Papadopoulos, Symeon & Patras, Ioannis & Kompatsiaris, Ioannis. (2019). Finding Near-Duplicate Videos in Large-Scale Collections. 10.1007/978-3-030-26752-0_4.
2. Shen, Heng & Liu, Jiajun & Huang, Zi & Ngo, Chong-Wah & Wang, Wei. (2013). Near-Duplicate Video Retrieval: Current Research and Future Trends. Multimedia, IEEE. 45. 1 - 1. 10.1109/MMUL.2011.39.
3. Song, Jingkuan & Yang, Yi & Huang, Zi & Shen, Heng & Hong, Richang. (2011). Multiple feature hashing for real-time large scale near-duplicate video retrieval. MM'11 - Proceedings of the 2011 ACM Multimedia Conference and Co-Located Workshops. 423-432. 10.1145/2072298.2072354.
4. K. Karu, A. K. Jain, and R. M. Bolle. Is there any texture in the image? In 13th International Conference on Pattern Recognition, pages B770–774, Vienna, August 1996.
5. T. P. Minka and R. W. Picard. Interactive learning using a 'society of models'. Technical Report TR-349, MIT, 1995.
6. A. Pentland, R. W. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. Proc. SPIE Vol 2185: Storage and Retrieval for Image and Video Databases II, February 1994.
7. T. Randen and J. H. Husoy. Multichannel filtering for image texture segmentation. Optical Engineering, 33:2617–2625, August 1994.
8. H. Rowley, S. Baluja, and T. Kanade. Neural network based human face detection. In Proc. Computer Vision and Pattern Recognition, San Francisco, July 1996.
9. A. Vailaya, Y. Zhong, and A. K. Jain. A Hierarchical System for Efficient Image Retrieval. In 13th International Conference on Pattern Recognition, pages C356–360, Vienna, August 1996.
10. B.-L. Yeo and B. Liu. Rapid scene analysis on compressed videos. IEEE Transactions on Circuits and Systems For Video Technology, 5(6):533–544, December 1995.
11. M. Yeung and B.-L. Yeo. Time-constrained clustering for segmentation of video into story units. In 13th International Conference on Pattern Recognition, pages 375–380, Vienna, August 1996.
12. H. J. Zhang, A. Kankanhalli, and S. W. Smoliar. Automatic partitioning of full-motion video. Multimedia Systems, 1(1):10–28, 1993.
13. H. J. Zhang, S. W. Smoliar, and J. H. Wu. Content-based video browsing tools. In Proc. SPIE Conference on Multimedia Computing and Networking, San Jose, CA, February 1995.
14. Y. Zhong, K. Karu, and A. K. Jain. Locating text in complex color images. Pattern Recognition, 28(10):1523–1536, October 1995.
15. G. Hristescu and M. Farach-Colton, "Cluster-preserving embedding of proteins," Tech. Rep. DIMACS 99-50, Rutgers University, Piscataway, USA, 1999.
16. A. P. Berman and L. G. Shapiro, "A flexible image database system for content-based retrieval," Computer Vision and Image Understanding, vol. 75, no. 1/2, pp. 175–195, July/August 1999.
17. L. Cieplinski, S. Jeannin, M. Kim, and J.-R. Ohm, "Visual working draft 4.0," Tech. Rep. W3522, ISO/IEC JTC1/SC29/WG11, July 2000.
18. A.Z. Broder, S.C. Glassman, M.S. Manasse, and G. Zweig, "Syntactic clustering of the web," in Sixth International World Wide Web Conference, Sept. 1997, vol. 29, no.8-13 of Computer Networks and ISDN Systems, pp. 1157–66.
19. S. Theodoridis and K. Koutroumbas, Pattern Recognition, Academic Press, 1999.