Effective and Efficient Content Redundancy Detection of Web Videos

***Dr. R. Kanagavalli, **Shahistakhan S, **Simren B S**

**Professor and Head of Department, **Students*
Dept. of ISE, The Oxford College of Engineering

[1]*Received: 15 March 2024; Accepted: 23 April 2024; Published: 28 April 2024*

ABSTRACT

Countless of videos are shared and kept on cloud storage platforms every day. This study's main objective is to create a comprehensive and efficient online near-duplicate video recognition system using material. As opposed to using advanced feature design, well-organized classifiers are implemented to improve accuracy. In the meanwhile, we increase the detection speed through employing simplistic features that have reduced dimensionality and taking utilization of detection architecture's parallelism. A significant amount of these videos are near-duplicate or identical. Building a content-based redundancy detection method that is both efficient and effective is therefore crucial, as this research has a wide range of applications. We demonstrate through many experiments that the recommended method of detection is efficient and accurate. The initial phase is assembling a dataset of webpages offering Creative Commons videos. The subsequent phase consists of contrasting hierarchical detection systems (HIER) with near-duplicate videos (NDVs).third-stage NDVD systems that depend entirely on conventional visual attributes.

INTRODUCTION

The main drivers of this expansion were the quick developments in multimedia technology and the rising demand for video sharing and hosting services like YouTube and Yahoo! Video. On the World Wide Web, videos have grown into an important kind of big data. Internet videos constituted 78% made up all the United States internet usage occurred in 2014, and according to Cisco Systems, that number is expected to rise to 84% in 2018. Duplications are a common side effect of the growing amount of video content. According to Wu et al.'s research [2], about 27% of the 13,129 films available online are closest duplicates. Therefore, one of the primary study objectives is to successfully detect near-duplicate videos (NDVs) on a broad scale, as this could enhance the performance of video hosting and sharing services in various manners. As an example, storage management and utilization.

From an examination of the NDV-related metadata, metadata pollution At current times, NDVs are frequently discovered via information, such as relevant descriptions, tags, or keywords. However, in terms of involves recognizing NDVs, descriptions and metadata are not as reliable as visual content. It is often the case for identical video clips to have different sets of pertaining tags, and even recordings with the same set of tags can differ considerably. As a result, it is recommended to use an NDVD system based on content compared to metadata [2], [4], [5].

However, for them to accomplish decent detection accuracy, these NDVD systems typically rely on high-dimensional feature representations and intricate algorithms, which trades efficiency for accuracy. Using this method for large-scale NDVD applications is not feasible. Every minute, 300 hours of video content are uploaded, according to the figures from YouTube [6]. The detecting speed of an NDVD system cannot match the uploading speed of videos if it is not efficient enough.

[1] *How to cite the article:* Kanagavalli R., Shahistakhan S., Simren B.S (April, 2024); Effective and Efficient Content Redundancy Detection of Web Videos; *International Journal of Advances in Engineering Research*, Apr 2024, Vol 27, Issue 4, 14-20

The Complexity of Data videos have a higher information density and level of complexity when compared to other big data types like records or logs. As a result, profiling a video using features is less successful than content-based duplicate document detection. Many changes are made to video content in the cloud to create NDVs, such as changes the encoding type or settings, adjustments to photometric data, or the addition or removal of frames. Every feature that has been found so far has its own disadvantages because it discards certain information about video footage.

However, in order to achieve good detection accuracy, these NDVD systems typically need complex algorithms and high-dimensional feature representations, which trades efficiency for accuracy. Using this method for large-scale NDVD applications is not feasible. According to data from YouTube, nearly three hundred hours of video footage are posted per minute [6]. The detecting speed of an NDVD system cannot match the uploading speed of videos if it is not efficient enough. It is therefore difficult to develop a workable NDVD system for two reasons listed below. However, because the development of these representations is exhaustive [4], this need runs counter to the process of utilising composite and high-resolution feature representations to capture motion pictures [4], [7].

As such, it is typically carried out despite their time-consuming constructions. Because of this, the focus of current research is on using feature fusion and high-dimensional feature design to detect NDVs [4], [7]. Nevertheless, in this work, the information entropy. In this section we demonstrate that a set of basic representations when combined with a composite feature representation can occasionally give additional details. Additionally, the informative may be further reduced by an increase in dimensionality. Therefore, we transfer emphasis from the design of advanced models to the design of platforms. We develop and put into operation Compound Eyes, an accurate and efficient NDVD system, served as the model for our concepts.

Even though a single miniature optical system is insufficient on its own, several of them combine to create a coherent eye sight that enables both the detection of rapid movement and an exceptionally wide seeing angle. The use of Compound Eyes requires ground-truth labels and a training set, both of which are outside the purview of this paper. We intend to move our system to more advanced cloud platforms in the future, such Spark, in order to get around the drawbacks of shared memory parallel computing architectures".

OBJECTIVE

Building a content-oriented redundancy detection system that is both efficient and effective is therefore crucial, given the broad range of applications of this research. Despite advancements in this area, the conflicting demands of speed and precision make it challenging to develop a workable system for detection for web videos.

RELATED WORK

We provide a quick overview of the methods and strategies used in media duplicate detection and retrieval in this section. These methods involve fusing features, improving speed efficiency, and representing multimedia object (such as a picture or video clip) as an able to be processed data type (such as vectors). Videos and photos are different from one another, yet because a video is made up of frames, its representations are usually transferable. An image can be represented in two steps: first, its visual features are extracted, and then those features are described using data types. Both the characteristics and the representations can be divided into both global and local information according to the granularity.

1. X. Wu, A. G. Hauptmann, and C.-W. Ngo, "Practical elimination of near-duplicates from web video search," in Proceedings of the 15th international conference on Multimedia. ACM, 2007, pp. 218–227.

The emergence of Web 2.0, the number of web videos is growing exponentially, and current video search engines are returning a lot of nearly identical videos. The variety of near-duplicate movies, which varies from straightforward formatting to intricate combinations of various editing techniques, makes the task of near-duplicate video recognition difficult. In order to balance the demands for speed and performance, we suggested combining global signatures with local pairwise measures using a hierarchical approach. Initially, evident near-duplicate movies were identified with high confidence using global signatures on colour histograms, and videos that were clearly dissimilar were filtered out.

For movies hard to category as either new identification, Video demonstrate that the hierarchical method is capable of detecting a wide variety of nearly identical videos and significantly minimizing redundant material.

2. T. Brants, F. Chen, and A. Farahat. A System for New Event Detection. ACM SIGIR'03, Canada, Jul. 2003.

Text information retrieval has investigated novelty/redundancy. It shares a strong relationship with Topic Detection and Tracking's New Event Detection and First Story Detection, which look into various aspects of automatically organizing news items in text areas. Finding the first story discussing an incident that hasn't been reported before is NED's task. Comparing news reports to story clusters from previously recognized events is a popular way to address NED. In order to quantify the level of originality exhibited in words, the majority of novelty detection techniques for document and sentences rely statistical language models and vector space models.

To enhance the search results, the concept of novelty detecting has also been used to web searches. Using language models, affinity graphs, and maximal marginal relevance, the documents and pages have been reranked based on the combination of question relevance and information novelty. These methods, however, mostly rely on textual data.

3. Y. Ke and R. Sukthankar, "Pca-sift: A more distinctive representation for local image descriptors," in Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, 2004.

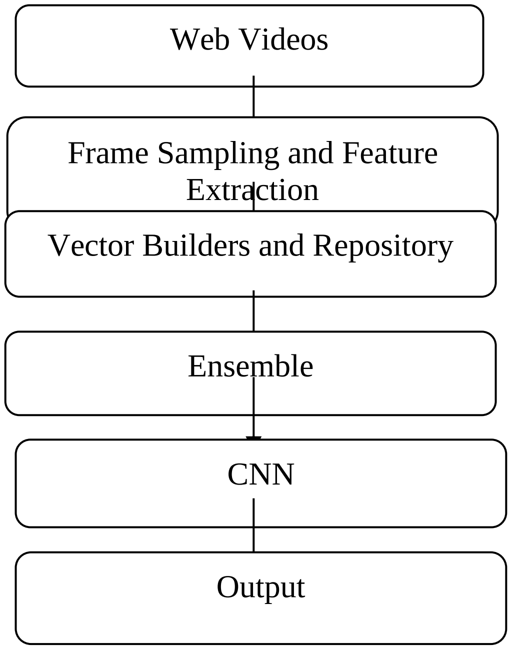
The evaluation of the SIFT matching algorithm's effectiveness against different types of picture distortion, including rotation, scaling, fisheye, and motion distortion, was conducted. The results showed that the method produced accurate and false positive rates for a substantial number of image pairs. Additionally, the distribution of the difference in keypoint orientation between correct and wrong matches was shown. The outcomes of this study will be applied in subsequent research to improve the accuracy of SIFT matching.

4. Y. Chen, W. He, Y. Hua, and W. Wang, "Compoundeyes:Near-duplicate detection in large scale online video systems in the cloud," 2016.

In order to detect near-duplicate images, we proposed a coarse-to-fine pattern matching approach that makes use of both global and local features. The suggested approach enables accurate and real-time near-duplicate picture recognition by utilising both local and global CNN features. We extract global characteristics to swiftly filter out the majority of irrelevant photos during the coarse matching stage. In order to get the final detection results, we extract and match the suggested local CNN features using the saliency map after detecting CFMs in the fine matching stage.

According to the experimental results, our method performs well in terms of both precision and efficacy, which makes it a suitable option for real-world applications involving content-based image detection and retrieval. Although pre-trained CNNs were initially intended for image classification, our approach uses them directly for near-duplicate picture identification. Therefore, using transfer learning techniques to create an optimised CNN algorithm for near-duplicate picture identification may prove to be more efficient.

ARCHITECTURE DIAGRAM



Explanation:

The flowchart offers a thorough understanding of CompoundEyes's operational architecture. CompoundEyes is a cutting-edge technology designed to identify nearly identical movies on the internet. It describes in detail the successive phases that are involved in the procedure, starting with frames sampling and ending with the final assessment of video similarity. The process begins with the methodical extract of frames from the input videos, which signifies the start of the phase of feature extraction. In this case, a wide range of features, including temporal, spatial, colour, appearance, and object size parameters, are carefully extracted from every frame. These features are converted into compressed feature vectors, which make the analysis and comparison in the space of vectors is domain more effective.

After feature extraction is finished, CompoundEyes enters the critical phase of group learning. Various machine learning classifiers work together to examine the variety of feature vectors and use their distinct viewpoints to evaluate video resemblance or duplication in a thorough way. This cooperative strategy encourages a strong decision-making process, which improves the system's capacity to identify minute details in the video material. Making decisions is where all of this complex work comes to an end. The individual decisions made by the group of classifiers are carefully combined and analysed to create an overall video similarity score. By means of this thorough analysis, CompoundEyes provides an efficient detection of nearly identical videos inside the large volume of internet multimedia information.

Essentially, the flowchart is a visual representation of CompoundEyes expertise and effectiveness, demonstrating how well it can handle the challenges of near-duplicate video identification. CompoundEyes is an essential tool for multimedia analysis and content management because of its careful structure and collaborative learning methodology. It is the only tool of its kind for video duplication detection.

METHODOLOGY

Our near-duplicate video detection solution, CompoundEyes, was developed using a technique that takes a methodical approach to guaranteeing the efficacy and efficiency of near-duplicate video detection. The first step in the process is having a complete grasp of the problem domain, which includes knowing how often it is to find near-duplicate movies online and how difficult it can be to discover them. This knowledge guides the creation of precise goals and specifications for the system, taking into account elements like precision, effectiveness, scalability, and flexibility.

The next step is a thorough analysis of the literature, which explores current approaches and research in the discipline of near-duplicate video identification in order to spot areas for advancement and creativity. Using this information, the CompoundEyes architectural framework is created, integrating essential components including ensemble learning, frame the sampling process, feature extraction, and decision-making.

Data collection and preprocessing are undertaken next, involving the acquisition of diverse datasets of web videos and the standardization of formats to ensure consistency in feature extraction. After implementing feature extraction methods to extract different visual aspects of the video clip, a diverse ensembles of artificial intelligence classifiers is chosen for ensemble learning.

The decision-making framework is developed to aggregate the individual predictions of the ensemble classifiers, resulting in a conclusive assessment of video similarity. Established metrics with iterative refinement based on insights gained from the evaluation process.

The methodology, implementation details, test outcomes, and findings are maintained throughout the process to aid in dissemination through academic publications, conversation presentations, and technical papers and move forward analysing in multi-media analysis and content management. After implementing feature extraction methods to extract different visual aspects of the video clip, a diverse ensembles of artificial intelligence classifiers is chosen for ensemble learning.

CompoundEyes promises to overcome the challenges and complexities involved in multimedia analysis of web content by offering a dependable and effective near-duplicate video recognition solution through the use of this methodical technique.

RESULT AND ANALYSIS

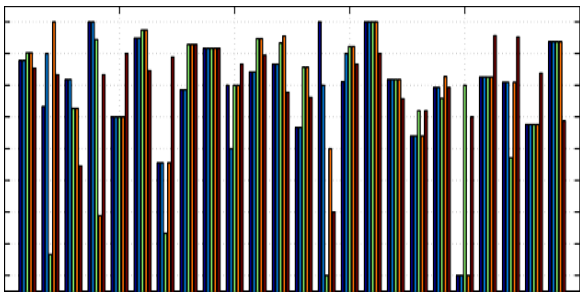
The CompoundEyes system employs the combination of ensemble learning, extraction of features, frame sampling, vector representation, and decision-making approach in a multi-layered manner to identify video duplication. The system has a notable speed and accuracy advantage over equivalent systems, according to experiments and research. CompoundEyes improves accuracy significantly, primarily as a result of its students' collaborative efforts. Using seven feature extraction algorithms, the system captures different aspects of video content such as colour, texture, object size, spatial, temporal, and so on, providing an extensive variety of features for learning. The range of feature representation and the diversity of representation both increase prediction accuracy. Furthermore, the conscious choice layer's hierarchical learner design enables the aggregation of individual predictions, resulting in more accurate and dependable final classifications. CompoundEyes's two key features allow it to recognise objects faster. First off, compact representations speed up the process of extracting features and obtaining nearby vectorial representations. Secondly, the system leverages both function synchronisation amongst vectors builders and the weak learners, as well as frame-level data parallelism within vector builders. In addition to facilitating efficient vector repository calculation, this parallelism contributes to an overall improvement in detection performance. CompoundEyes in-situ updating capability allows classifiers to be updated with fresh data continually and without incurring computational overhead, thus enhancing its effectiveness and versatility.

CompoundEyes combines sophisticated ensemble learning with enhanced feature representation to provide a potent and effective solution for video duplicate detection. The system maintains a high detection speed and achieves amazing accuracy improvements because to its hierarchical organisation and systematic architecture, which makes it a viable tool for a variety of a multimedia presentation analytics and content management applications.

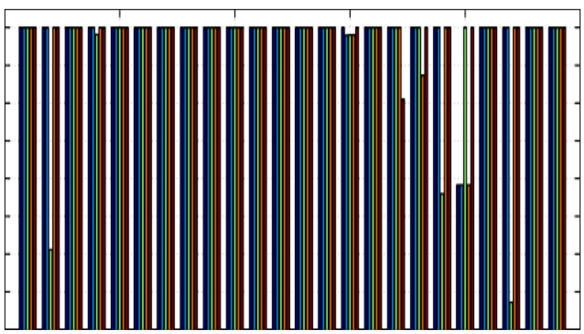
CONCLUSION

We suggested and created Compound Eyes, a powerful and functional NDVD system. This system's architecture replaces the complex feature representation with a bag of more straightforward feature representations, using a novel detection paradigm. Through the use of functionality decomposition, the ideas of the systems approach may be used to the design of the system's structure. This allows for the utilisation of the parallelism and lower level of complexity of each component to reduce the temporal overhead associated with NDVD activities. In the meantime, the efficient merging of data in features maintains a respectable level of detection accuracy. The trial and analysis findings confirm that CompoundEyes not only outperforms other modern feature fusion NDVD/NDVR systems in terms of detection accuracy, but also performs on par with feature-centered systems that rely on CNN and BoW features. CompoundEyes performs better than any other system in the interim in terms of peak memory utilisation and temporal complexity. To sum up, CompoundEyes performs adequately well enough for extensive NDVD activities involving web videos.

RESULT:



Accuracy



Average Precision

REFERENCE

[1] M. LOPES. Videos may make up 84 percent of internet traffic by 2018: Cisco. [Online]. Available: http://www.reuters.com/article/2014/06/10/us-internet-consumerscisco-systems-idUSKBN0EL15E20140610.

[2] X. Wu, A. G. Hauptmann, and C.-W. Ngo, "Practical elimination of near-duplicates from web video search," in Proceedings of the 15th ACM international conference on Multimedia, 2007.

[3] J. S. Pedro, S. Siersdorfer, and M. Sanderson, "Content redundancy in youtube and its application to video tagging," ACM Transactions on Information Systems (TOIS), 2011.

[4] J. Song, Y. Yang, Z. Huang, H. T. Shen, and R. Hong, "Multiple feature hashing for real-time large scale near-duplicate video retrieval," in Proceedings of the 19th ACM international conference on Multimedia, 2011.

[5] M. Hefeeda, T. ElGamal, K. Calagari, and A. Abdelsadek, "Cloud-based multimedia content protection system," 2013.

[6] YouTube. Statistics. [Online]. Available: http://www.youtube.com/yt/press/statistics.html

[7] L. Shang, L. Yang, F. Wang, K.-P. Chan, and X.-S. Hua, "Real-time large scale near-duplicate web video retrieval," in Proceedings of the 18th ACM international conference on Multimedia, 2010.

[8] Y. Chen, W. He, Y. Hua, and W. Wang, "Compoundeyes: Near-duplicate detection in large scale online video systems in the cloud," in Computer Communications, IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on, 2016.

[9] A. Amir, M. Berg, S.-F. Chang, W. Hsu, G. Iyengar, C.-Y. Lin, M. Naphade, A. Natsev, C. Neti, H. Nock et al., "Ibm research trecvid- 2003 video retrieval system," NIST TRECVID-2003, 2003.

[10] B. Thomee, M. J. Huiskes, E. Bakker, and M. S. Lew, "Large scale image copy detection evaluation," in Proceedings of the 1st ACM international conference on Multimedia information retrieval, 2008.