A SYSTEMATIC VIDEO INDEXING APPROACH USING DECISION TREE

M. Madheswaran
Centre for Research in Image and Signal Processing (CRISP),
Department of Electronics and Communication Engineering,
Muthayammal Engineering College (Autonomous),
Rasipuram, Tamilnadu, India.
madheswaran.dr@gmail.com

Alexander Muthurengan Murugaiyan
Department of Computing and Information Systems,
University of Seychelles,
Anse Royale, Seychelles.
alexander.murugaiyan@unisey.ac.sc

Submitted: *Aug, 15, 2022* ***Revised:*** *Oct, 04, 2022* ***Accepted:*** *Oct, 20, 2022*

Abstract: A systematic categorization approach for video indexing is presented in this paper. The amount of multimedia information that can be accessed over the internet continues to expand exponentially. Due to this growth and development of multimedia on the internet, particularly videos, there has been a rise in the need for video retrieval. The goal of this work is to discover subsets of characteristics that are suitable for indexing or categorizing video content. The features chosen from the frames of the video have dominant texture characteristics. Several statistical features have been applied for better performance with decision tree classification. The investigation included 1000 videos from different video contents, such as news, cartoon, advertisements, movies, and sports categories. Results show that the overall misclassification rate percentage was below 3%. The capability of indexing the video contents indicates the real power of the proposed system, which can enhance existing indexing services, thereby enriching the tools that are available for video indexing.

Keywords: Video indexing, decision tree, video categorization, texture features, pattern recognition, statistical features.

I. INTRODUCTION

The automated categorization of videos is expanding exponentially due to the enormous demand for them in the commercial market. This demand comes from various video collections, including cartoons, dramas, commercials, and news. It helps with adequate storage, fast browsing, and speedy retrieval of massive collections of video material. To aid computer vision and autonomous management systems, video categorization automatically interprets the video objects being seen. Because many different videos seem quite similar to one another, extracting cognitive content is regarded as a complex problem in categorizing sports genres.

A two-tiered automated system for classifying educational or non-educational videos is described in [1]. The video processing system is responsible for the first level of categorization. The deep learning algorithms are used to process the feature vectors obtained by using Inception V3 to extract them from each video frame. Text processing is employed in the metadata of YouTube videos

to complete the second level of categorization. These are the search terms used while looking for videos on YouTube.

The sports videos taken by random individuals using mobile phones are classified in [2]. A motion trajectory descriptor is utilized to effectively and efficiently depict a video because the movements involved in sporting activities are rather diverse from one another. In addition, the classification choice might be integrated throughout time via the temporal analysis of local descriptors.

A technique for automatically extracting headline sequences from news videos and a classification system for anchorperson scenes and reporter scenes based on visual elements is discussed in [3]. In the first step of the procedure, several characteristics are retrieved from keyframes, representing each shot once the process of shot boundary creation has been completed. In addition, some frames are categorized as part of the news headlines, part of the anchorperson, or part of the reporter based on a threshold calculation and a judgment made after combining numerous visual characteristics.

Video similarity assessment is described in [4] for classifying sports videos. It calculates the distance between each sequence of the training videos and the test videos by comparing each sampling frame of the training videos with all sampling frames of the test movies and then averaging the results by the value that was anticipated to be present. In addition, the colour histogram is used to represent each frame to facilitate the enhancement of feature reduction and the subsequent acceleration of data processing. Following that, the closest neighbour classifier is used.

A powerful combination of the two audio descriptors is used in [5] for classifying three video genres. They are signal energy and Mel Frequency Cepstral Coefficients (MFCC). The combination of features is fed to the Support Vector Machine (SVM) classifier for the classification. A fuzzy logic is used in [6] to determine which edge components are straight and which are curved for each video frame. Then, the local and global mass estimates are made by drawing a series of contiguous ellipses over edge images. In addition, features based on spatial closeness have been identified, and the SVM classifier is employed for the classification.

A collection of computational characteristics for the problem of automatically classifying videos is presented in [7]. These features were derived from the motion and colour in videos. A classification method for sporting events based on the Hidden Markov Model (HMM) is discussed in [8]. The rate at which colours change is determined for each video frame and sent into HMM as observation sequences in the classification process. A novel three-stage framework that integrates concept-level semantics and contextual features are described in [9]. It uses predictive and specific concept classifiers for categorizing the videos.

For accurate classification, a combination of features generated from visual and audio are utilized in [10] with modern classifiers such as Gaussian Mixture Models and SVM. MFCCs, edge, colour, and histogram features are extracted. The characteristics from both spatial and temporal are generated for video classification in [11] using 20 seconds videos. SVM classifier is used to classify cartoons, ads, and sports videos. Different sports video classification systems are described in [12-14] using machine learning and deep learning approaches.

II. PROPOSED SYSTEM

A typical pattern recognition system for any image classification is shown in Figure 1. The recognition system may be divided into two distinct modules. The initial component of the system is the feature extraction module, which is

responsible for the parameterization of the raw input data. The front-end interface is the feature extraction module between the raw data and the image classifier. The image classifier is included as the second module. This creates a map between the incoming feature vectors and the appropriate underlying pattern by taking input from the feature extraction module and classifying the patterns.

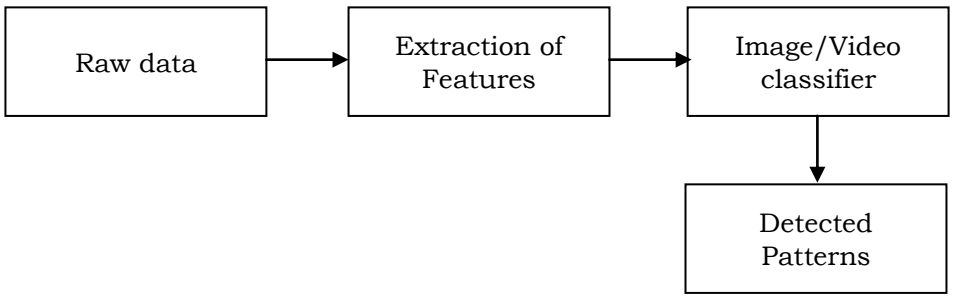

Fig. 1 Typical pattern recognition system

A. Feature Extraction Module

The feature extraction module can be considered an intermediary between the raw data and the classification module. The properties of the extracted features are of the utmost importance to the performance of the classification module. The ability of the classifier to correctly classify the videos will be aided by the presence of a "good" collection of features. A feature extraction module not only provides the classifier with "excellent" features but also, the features are insensitive to any corruption.

An ordered collection of coefficients acquired by spectral analysis on an image is known as a feature vector. Each feature vector represents the characteristics of the patterns, and such feature vectors provide a parametric representative that describes the texture patterns. When doing classification, it is important to keep the dimensionality of the feature vector as low as possible to reduce the classifier's processing resources. Numerous feature extraction strategies have been designed as potential methods for characterizing images [15].

Three types of statistical features are extracted from each video frame. They are

1. Histogram features
2. Co-occurrence features
3. Intensity difference features

A histogram is one of the simplest methods that may be used when describing the texture. Because it is generated from the distribution of the image's histogram, it is a straightforward calculation. Even though it does not include the relative information of pixels, it has been adopted in many applications because of its simple implementation. Co-occurrence features represent the spatial connections of grey levels in an image, and thus they are superior to other features. It requires computing a great number of matrices, which might increase a significant amount of processing time. There is also a need for essential direction in selecting characteristics obtained from the co-occurrence matrices, which might be very difficult. In this approach, all these features are combined for oral cancer diagnosis.

B. Classification Module

The term “image or video classification" refers to classifying images or videos into various classes based on the kind or severity of the underlying condition. Accurate and automated categorization of images or videos helps to decrease the amount of work that has to be done by humans when making choices. Classification is a popular use of the decision tree induction algorithm. It is utilized extensively in computer vision applications. A decision tree is a chart structured like a tree, with each internal node representing an attribute, each branch representing the test's result, and each leaf node representing the distribution of classes.

The decision tree learning technique is a kind of supervised learning. It is most effective in situations involving instances represented by attribute-value pairs, with the goal function having discrete output values [16]. Instances are categorized using the Decision Tree method, sorting them in ascending order from the tree's root node to its leaf nodes. Each node that is not a leaf is linked to a test that divides the possibilities associated with that node into subgroups based on the various test outcomes. And the directions that each branch points in are determined by the outcome of a certain test, with the leaf node of each branch being linked to a different group of potential solutions. After the classification tree has been built, it is simple to develop the classification rule. As a result, the decision tree induction consists of the process shown in Figure 2.

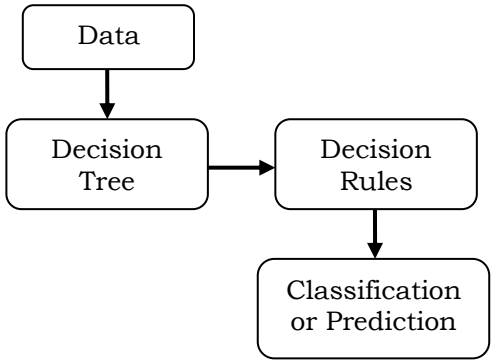

Fig. 2 Decision Tree Approach

The process in which a decision tree is constructed from a dataset is one of the most critical components that must be completed. Defining the sequence of attributes, which is helpful when building little trees, is the construction process's first and most important phase. The idea of average entropy is brought up whenever there is a need to choose between several attributes. Entropy is a metric that originates from information theory; it describes the level of purity or homogeneity that a collection of samples has. The strategy entails determining the average entropy of each characteristic and selecting the characteristic with the lowest value of average entropy. It is necessary to understand the concepts of *nb*, *nbe*, and *nt* to calculate the average entropy. *nb* refers to the total number of instances in branch *b*, *nbe* refers to the total number of instances in branch *b* of class *c*, and *nt* refers to the total number of instances in all branches. It is defined by

$$\sum_{b}\left[\left(\frac{nb}{nt}\right)\times\left(\sum_{c}-\left(\frac{nbc}{nb}\right)\log_2\left(\frac{nbc}{nb}\right)\right)\right] \quad (1)$$

Because this characteristic has the lowest value relative to the average entropy, it was chosen as the basis for the decision tree. It is recommended that a pruning algorithm be used for the decision tree to enhance the system's accuracy by deleting tree branches. After completing this stage, decision trees can be turned into classification IF-THEN rules suitable for classification.

III. RESULTS AND DISCUSSIONS

The proposed system discussed in this paper is designed for the categorization/classification of videos into news, cartoon, advertisements, movies, and sports categories. The system's performance is analyzed in great detail using 1000 videos (200 videos per category). All videos are downloaded from the youtube channel with a duration of only two seconds. Sample videos are shown in Figure 3.

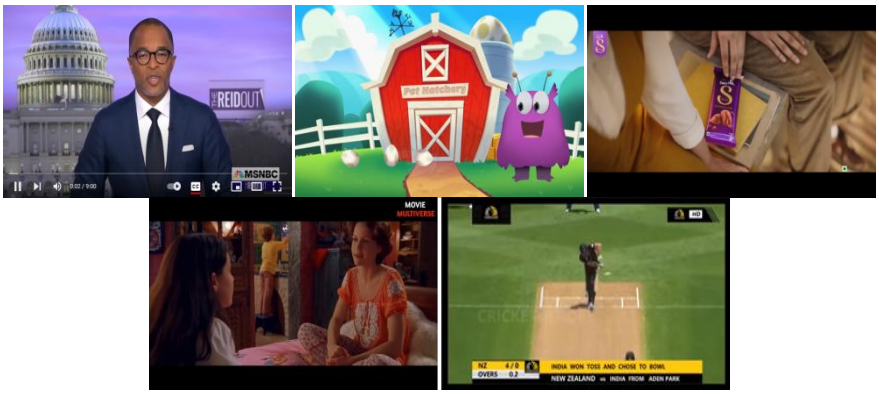

Fig. 3 Sample Videos

Using a hold-out approach of cross-validation is considered to be the most fundamental. During the training process, a portion of the data set is kept away and is later used exclusively for performance evaluation. This delivers a more trustworthy test. However, the quantity of data used to develop models reduces. This is an actual example of inefficient data use. Another option s to use the whole data set except for a single pattern for each round. The model constructed using the N-1 data is tested on the one example that was not included, one for each pattern. The leave-one-out-cross-validation method relies on this idea. Figure 4 shows the cross-validation approach used in this work.

Although only the patterns that are not considered contributions to the error rate estimation in the cross-validation situation [14], the error rate estimations are still produced using all samples. A sufficient amount of experimental work to demonstrate the applicability of the proposed approach to index or categorize videos is provided. Table 1 shows the confusion matrix obtained for the system designed in section 2 using 50 decision trees.

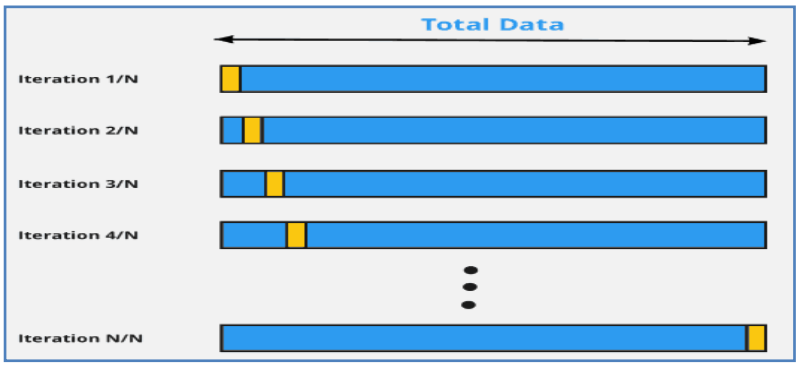

Fig. 4 Cross-validation approach used in this work

TABLE. 1 Confusion matrix of the proposed video indexing system

		Predicted class				
		News	Cartoon	advertisement	Movies	Sports
Actual class	News	200	0	0	2	2
	Cartoons	0	200	0	0	0
	advertisement	0	0	194	0	0
	Movies	0	0	4	196	0
	Sports	0	0	2	2	198

It can be seen from the confusion matrix that the system provides average classification accuracy of 98.8% using the combined statistical features. Among the different types of videos, news and cartoons provide 100% classification accuracy, while others are classified with a lesser miss rate. The system's performance depends on the statistical features and the number of decision trees used in this work. Figure 5 shows the obtained average accuracy of the video indexing system with a different number of decision trees.

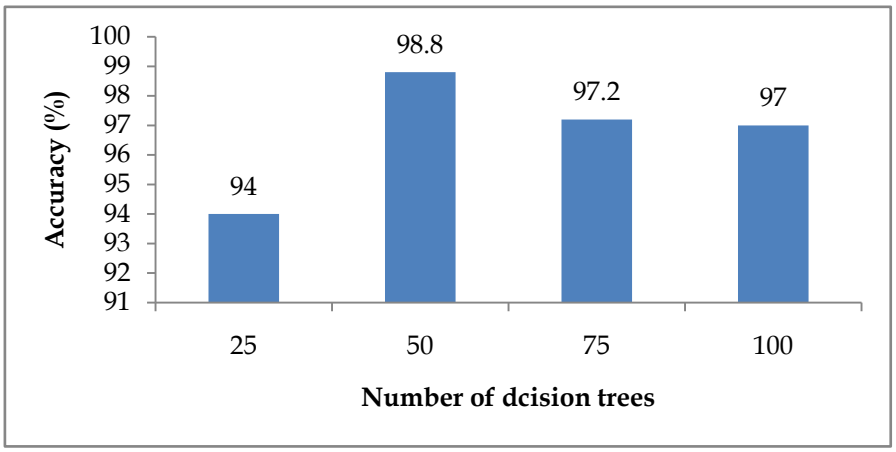

Fig. 5 Performance of the proposed video indexing system with different numbers of decision trees

It is observed from Figure 5 that increasing the number of decision trees increases the average classification accuracy of the system. However, more

decision trees used for the classification increases the computational complexity of the system.

IV. CONCLUSIONS

This paper presented an efficient approach for video indexing by extracting statistical features and a random tree classifier. Although video classification system is a well-studied topic in the pattern recognition and data analysis literature, several statistical features have been applied for better performance with random tree classification. All of the goals of this study have been accomplished. It has been determined that the video indexing system can be successfully applied and reach a high classification accuracy level. The newly created automated categorization system can correctly categorize the input videos. The results of the experiments showed that the statistical characteristics were the most effective for accurately classifying the video clips with more than 97% accuracy.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

REFERENCES

[1]. I. Ramesh, I. Sivakumar, K. Ramesh, V. P. P. Venkatesh and V. Vetriselvi, "Categorization of YouTube Videos by Video Sampling and Keyword Processing," International Conference on Communication and Signal Processing, 2020, pp. 56-60.

[2]. S. M. Safdarnejad, X. Liu and L. Udpa, "Genre categorization of amateur sports videos in the wild," IEEE International Conference on Image Processing, 2014, pp. 1001-1005.

[3]. H. Han and J. Kim, "An useful method for scene categorization from new video using visual features," Third World Congress on Nature and Biologically Inspired Computing, 2011, pp. 480-484.

[4]. P. Mutchima and P. Sanguansat, "A Novel Approach for Measuring Video Similarity without Threshold and Its Application in Sports Video Categorization," First International Conference on Pervasive Computing, Signal Processing and Applications, 2010, pp. 868-872.

[5]. N. Dammak and Y. Ben Ayed, "Video genre categorization using Support Vector Machines," 1st International Conference on Advanced Technologies for Signal and Image Processing, 2014, pp. 106-110.

[6]. S. Roy, P. Shivakumara, N. Jain, V. Khare, U. Pal and T. Lu, "New Fuzzy-Mass Based Features for Video Image Type Categorization," International Conference on Document Analysis and Recognition, 2017, pp. 838-843.

[7]. B. T. Truong and C. Dorai, "Automatic genre identification for content-based video categorization," International Conference on Pattern Recognition, 2000, pp. 230-233.

[8]. J. Hanna, F. Patlar, A. Akbulut, E. Mendi and C. Bayrak, "HMM based classification of sports videos using color feature," IEEE International Conference Intelligent Systems, 2012, pp. 388-390.

[9]. M. Afzal, X. Wu, H. Chen, Y. G. Jiang and Q. Peng, "Web video categorization using category-predictive classifiers and category-specific concept classifiers," Neurocomputing. vol. 214, 2016, pp. 175-190.

[10]. A.M. Barbancho, L.J. Tardón, J. López-Carrasco, J. Eggink and I. Barbancho, "Automatic classification of personal video recordings based on audiovisual features," Knowledge-Based Systems, vol. 89, 2015. pp. 218-227.

[11]. V. Suresh, C.K. Mohan, R.K. Swamy and B. Yegnanarayana "Content-based video classification using support vector machines," International conference on neural information processing, 2004, pp. 726-731.

[12]. S. U. Maheswari and R. Ramakrishnan, "Sports video classification using multi scale framework and nearest neighbor classifier," Indian Journal of Science and Technology. vol. 8, no. 6, 2015, pp. 529-535.

[13]. M. A. Russo, L. Kurnianggoro and K. -H. Jo, "Classification of sports videos with combination of deep learning models and transfer learning," International Conference on Electrical, Computer and Communication Engineering, 2019, pp. 1-5.

[14]. M. Ramesh and K. Mahesh, "Sports video classification with deep convolution neural network: a test on UCF101 dataset," International Journal of Engineering and Advanced Technology. vol. 8, no. 4S2, 2019, pp. 2249-8958.

[15]. K. Djunaidi, H.B. Agtriadi, D. Kuswardani and Y.S. Purwanto, "Gray level co-occurrence matrix feature extraction and histogram in breast cancer classification with ultrasonographic imagery," Indonesian Journal of Electrical Enggineeriung and Computer Science, vol. 22, no. 2, 2020, pp. 187-192.

[16]. H. Ferdous, T. Siraj, S.J. Setu, M. Anwar and M.A. Rahman, "Machine learning approach towards satellite image classification," Proceedings of International Conference on Trends in Computational and Cognitive Engineering, 2021, pp. 627-637.