Classification and Analysis of Web-based Multimedia Data using Metadata

Basavaraj A. Goudannavar, PhD
Department of Computer Applications
JSS Shri Manjunatheshwara MCA Institute, Dharwad, Karnatak

Jyoti G. Biradar, PhD
Department of Computer Applications
S.G. Balekundri Institute of Technology,
Belagavi,– 590 010, Karnataka

Anil Trimbakro Gaikwad, PhD
Associate Dean Academics
D Y Patil Agriculture and Technical University, Kolhapur - 416112, Maharashtra

ABSTRACT
The classification of multimedia data into conceptual categories, such "Entertainment," "News and Politics," "Sports," etc., typically calls for the combination of evidence from various multimedia components, including text, audio, video, and images. In this work, data mining classification techniques based on the metadata of the considered web multimedia data are used to identify and categorize domain-specific web multimedia data into various groups. Data mining classification algorithms, such as Decision Tree (DT) and Support Vector Machine (SVM) model, automatically classify web multimedia data into various categories based on extracted metadata. To extract knowledge from web multimedia data, the classification results are contrasted and examined.

General Terms
Data Mining, Web Multimedia Database, Classification

Keywords
Web Multimedia Mining, Metadata, Indexing, and Multimedia classification

1. INTRODUCTION
The advances in the digital and network technology have produced multimedia information on the Social media websites such as YouTube, Red Tube, and Face Book etc, automatic organizing of multimedia data into different classes is an emerging trend in the area of web multimedia research. In order to improve search ability, relevance, and tailored recommendations for multimedia content, including photographs, videos, and audio files, this technique uses algorithms and metadata to identify and organize the material. Important advancements consist of: Automation of Multimedia Classification: Evaluating metadata or the content itself, automated systems can classify content, saving time and improving accuracy. Content-Based Analysis: Exceeding typical metadata-based methods, content-based techniques classify media through picture, video, or audio analysis. Machine Learning: More precise and effective classification methods have been made possible by the development of sophisticated machine learning models, such as neural networks.

The amount of all three classes of multimedia data on the Inter net is increasing daily. It is getting harder to recognize and categorize multimedia content without knowing its content as it becomes more prevalent on the internet. An attempt is made to categorize Class 1 web multimedia data using a domain-specific methodology in this experiment.

Web videos have been selected as a Class 1 web multimedia data set for experimental purposes because the video domain consists of four fundamental components: audio, video, image, and text. The basic components of web videos will be separated for metadata extraction. The Classification is a supervised Machine Learning technique which assigns labels or classes to different objects or groups. Classification is a two-step process: First step is construction model which is defined as the analysis of the training records of a multimedia data. Second step is model usage; the model constructed is used for classification. The classification accuracy is estimated by the percentage of test samples or records that are correctly classified [1] [2]. Many classification models/algorithms and data mining and machine learning tools are developed in recent years. In this work, using KNIME data mining tool [3], the web multimedia-video metadata are extracted and classified based on available metadata of web multimedia-videos using Decision Tree and Support Vector Machine classification algorithms. The classification results are compared and analyzed. The rest of the paper is organized as follows: The section 2 represents related works on the classification of web multimedia videos, section 3 represents proposed web multimedia video classification methodology, section 4 represents performance evaluation analysis of classification models, and finally section 5 represents conclusion and future work.

2. RELATED WORKS
The aim of classification is to establish a collection of data models capable of precisely determining the class of various objects with high accuracy. In this paper, explore into two significant challenges or the adaptation of classifiers, specifically adaptive support vector machines (A-SVMs), which adapt auxiliary classifiers to a new dataset with limited labeled examples, and a technique for identifying the most suitable auxiliary classifiers for this adaptation process. The following observations from the experiments first, adapted classifiers trained by A-SVMs significantly outperform auxiliary classifiers and new classifiers trained from the labeled examples; Second, compared with other adaptation techniques, our approach achieves better performance than the ensemble approach and comparable performance to the aggregate approach while requiring 1/10 of the latter's training time; Third, selecting good auxiliary classifiers for adaptation is critical to the performance, and our selection method has proved to be effective [4]. During multimedia information retrieval, by using random forests, our method has all the advantages of tree classifiers (such as nonparametric and nonlinear), so it can effectively address the multimodal distribution of relevant objects [5].In this paper, the empirical study on the testing and fault-identification of multimedia systems by treating the issue as a classification problem. The experiment shows via empirical studies that classification techniques can learn non-deterministic characteristics from training data and identify the types of fault for multimedia systems. The relationships and memory

utilizations as the features for classification and also studied empirically whether the effectiveness of identifying errors may be affected by the types of fault, the system loading, and the classification techniques[6].

This article presents a possible explanation why multimedia retrieval and classification with huge real world data collections like web content stays for now behind the expectations that, in theory, the fusion of more information should lead intuitively to improved performance [7]. A new statistical model for the classification of structured documents and consider its use for multimedia document classification. In this classification model tested a particular instance of the model on the task of Web page filtering by considering two information sources: text and image. The model has been compared to baseline flat text and image classifiers. The experiments show that taking the structure into account increases the performance compared to a flat text classifier and that the integration of textual and image information via this structured document model still increases the performance [8].

3. METHODOLOGY

In this section we propose a effective methodology to extract the metadata from web multimedia files and classify them based on the extracted metadata by applying data mining techniques. Out of the entire metadata dataset, 60% is used for training purposes and the remaining 40% is used for testing the classification model that was created using SVM and Decision Tree classification techniques. The system model of the proposed system is represented in Figure 1. It consists of the following components:

i) Web multimedia-video metadata extraction
ii) Classification model
iii) Classification analysis

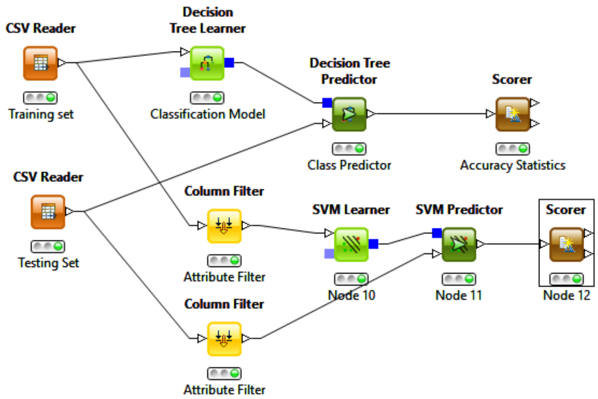

Figure 1: System model of the proposed methodology

The functionality of each component of the proposed system model is discussed in the following subsections.

3.1 Web Multimedia-Video Metadata Extraction and Pre-Processing

The metadata of web multimedia-video data are extracted using Mediainfo Extractor tool. Out of the 27 attributes, 22 were deemed to be significant for the planned work by experimental observation. Because the values of metadata attributes like codec id/info, frame rate mode, color space, scan type, and compression mode remain consistent for every tuple, they will not be included in the experiment. The remaining twenty-two metadata include: video duration, bit rate kbps, maximum bit rate kbps, pixels for width and height, display aspect ratio, bits/(pixel*frame), stream size mib, photo resolution, height, and width, text page, word count, character count, line count, paragraph count, size in kbps, and class. For experimental purposes, the extracted metadata will be stored as a CSV data file. The data have been pre-processed to add the mean or mode of each attribute to fill up any missing data [16].

3.2 Classification Model

In this experiment we adopt two classification model to classify web multimedia video data. The classification accuracy and efficiency will depend on the constructed classification model. This section represents detailed procedure to construct DT and SVM classification model.

3.2.1 Decision Tree Classification Model

The attribute selection measures provide specific criteria for each attribute describing the given tuples. As discussed in section 3.1, twenty two attribute class labels are considered for the dataset selected, and are listed in Table 1.

i) Attribute Selection Measures

Table 1: Attribute selection for classification

Multimedia metadata Attribute	Descriptions
Video Duration	Duration of Video component in times
Video Bit rate kbps	Bit rate of video component in kbps
Maximum bit rate kbps	Maximum bit rate of video component in kbps
Width Pixels	Width of video component in pixels
Height Pixels	Height of video component in pixels
Display aspect ratio	Display aspect ratio of video component
Bits/(Pixel*Frame)	Bits (quality) of video component in pixel
Stream size MiB	Stream size of video component in Mebibyte
Audio Duration	Duration of Audio component in times
Audio Bit rate kbps	Bit rate of audio component in kbps
Maximum bit rate kbps	Maximum bit rate of audio components in kbps
Stream size MiB	Stream size of audio component in Mebibyte
Image Resolution	Resolution of image component
Image Height	Height of image component in pixels
Image Width	Width of image component in pixels
Text Page	Page ordering and sequencing
Word count	Number of words in a document
Character count	Number of Character information
Line count	Number of line counts in a document
Paragraph count	Paragraph count in a document
Size in kbps	Size of text document in kbps
Class	Three different multimedia (Video Domain Specific) classes 1. Entrainment 2. News 3. Sports

The following is presented on the process to measure attribute selections for the multimedia metadata on the web: Randomly chosen class-labeled tuples from an online multimedia metadata database comprise the training set D. The class label attribute has three distinct values namely, 'Sports', 'News' and 'Entrainment', therefore, there are three distinct classes (i.e., m=3). Let class C1 correspond to Sports, class C2 correspond to News and class C3 correspond to Entrainment. There are 85 tuples of class sports, 100 tuples of class news, and 62 tuples of class entrainment. A node N is created for the tuples in D. To find the splitting criterion for these tuples, the information gain of each attribute is computed as follows.

$$\text{Info (D)} = -\sum_{i=1}^{M} \frac{CiD}{D} \log_2\left(\frac{CiD}{D}\right) \ldots\ldots\ldots (1)$$

Next, to compute the expected information requirement for each attribute.For the attribute image resolution (IR) to look at the image resolution size for each category of IR. For the Image resolution of sports category there are out of 26 metadata tuples 26 tuples belongs the IR size 1182x720. The IR of news category there are out of 91 metadata tuples 61 tuples belongs the IR size 640x360. In the entrainment category there are out of 95 metadata tuples 77 tuples belongs the IR size 1280x720. Using Equation the expected information needed to classify a tuple in D if the tuples are partitioned according to image resolution is:

$$\text{Info\_A(D)} = \sum_{i=1}^{n} \frac{|D_i|}{|D|} * \text{Info}(D_i) \ldots\ldots\ldots\ldots\ldots\ldots(2)$$

Hence, the gain in information from such a partitioning would be

$$\text{Gain (A)} = \text{Info (D)} - \text{Info\_A(D)} \quad \ldots\ldots\ldots..(3)$$

Compute Gain (all attributes) in a similar manner to determine which attribute has the biggest information gain and designate i it as the splitting attribute. Node N is labeled with IR, and branches are grown for each of the attribute's values as shown in table 2.

Table 2: Information Gain of Multimedia Metadata Attributes

Multimedia metadata Attribute	Information Gain
Video Duration	0.116
Video Bit rate kbps	0.126
Maximum bit rate kbps	0.130
Width Pixels	0.111
Height Pixels	0.115
Display aspect ratio	0.125
Bits/(Pixel*Frame)	0.138
Stream size MiB	0.142
Audio Duration	0.116
Audio Bit rate kbps	0.125
Maximum bit rate kbps	0.128
Stream size MiB	0.140
Image Resolution	0.155
Image Height	0.115
Image Width	0.111
Text Page	0.109
Word count	0.116
Character count	0.118
Line count	0.099
Paragraph count	0.110
Size in kbps	0.121

The tuples are then partitioned accordingly, where, Di contains 22 attributes which are outcomes of data partitions D1, D2, D3…Dn, and Info (Di) can be calculated by using eq (1). Using Eq(1),(2) and (3) information gain of each attribute will be calculated and the attribute which has highest information gain will be labeled as splitting node[10]. The Decision Tree classification model's gain is shown in Table 2, where the attribute "Image Resolution" has the highest gain of all the selected attributes. As a result, the attribute is chosen as the tree's root node. In a similar manner, the gain will be computed at each node and the tree will be formulated appropriately.

ii) Classification Rules

Classification rules can be extracted from the tree structure of the classification model for the dataset chosen as shown in Figure 2.

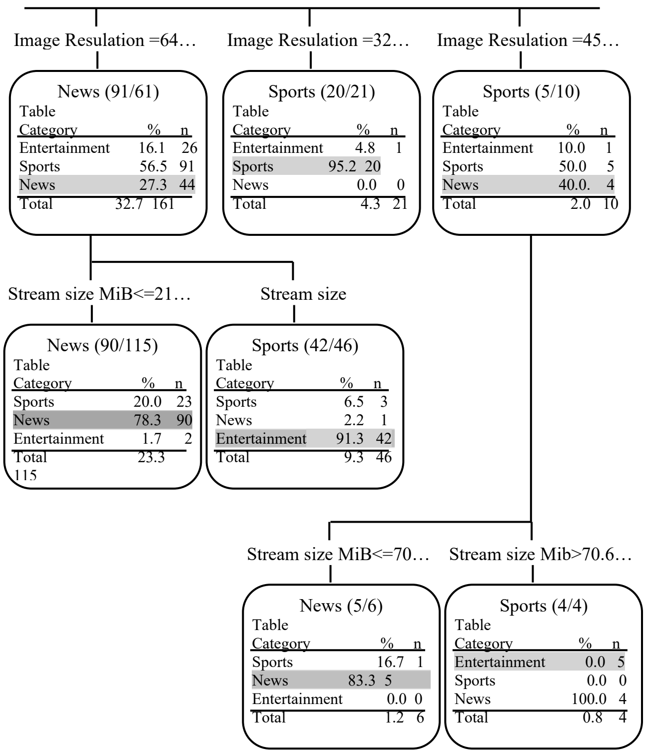

Figure 2: Tree structure result of DT classification model

The above tree can be converted to classification rules by traversing the path from root node to each leaf node in the tree. Figure 2 shows the root node constructed using the splitting values for the "Image Resolution", attribute every node holds class label information regarding instances that are appropriately and incorrectly classified.

3.3 Support Vector Machine Classification Model

The support vector machine (SVM) is a supervised classification system that uses a hypothesis space of linear functions in a high dimensional feature space in order to learn separating hyperplanes. As such, SVM classification attempts to generalize an optimal decision boundary between classes. Labeled training data in a given space is separated by a maximum margin hyperplane through SVM classification. The appeal of SVMs is based on their strong connection to the underlying statistical learning theory. That is, an SVM is an appropriate implementation of the structural risk minimization method [11] [12].

Consider the problem of separating the set of training vectors belonging to three separate classes, (x1, y1),…., (xl, yl), where xi € Rn is a feature vector and yi € {-1;,+1} a class label, with a hyperplane of equation wx + b = 0. Of all the boundaries determined by w and b, the one that maximizes the margin (Figure 3 would generalize well as opposed to other possible separating hyperplanes. A canonical hyperplane has the constraint for parameters w and b: minxiyi(w . xi + b) = 1. A separating hyperplane in canonical form must satisfy the following constraints, yi [(w . xi) + b]≥1, i = 1,…., l. The margin is $\frac{2}{\text{II wII}}$ according to its definition [13]. Hence the hyperplane that optimally separates the data is the one that minimizes ϕ (w) = $\frac{1}{2}$1 II w II 2.

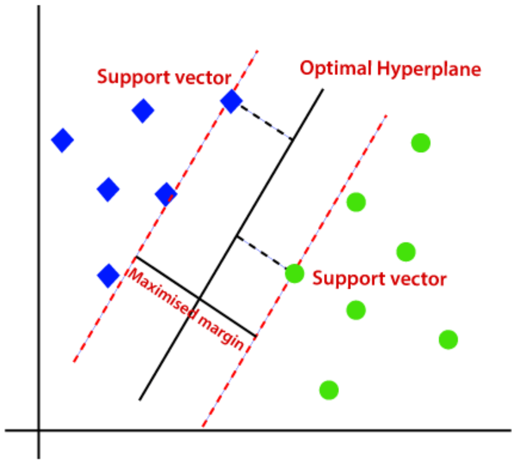

Figure 3: A Liner Support Vector Machine

The classification task involves training and testing data, which consist of web multimedia metadata instances. Each instance in the training set contains target value" (class labels) and several "attributes" (features). The goal of SVM is to produce a model, which predicts target value of data instances in the testing set that are given only the attributes. Training vectors are mapped into a higher (maybe infinite) dimensional space by a learned function. Then SVM finds a linear separating hyperplane with the maximal margin in this higher dimensional space.

Every instance in the training set has multiple "attributes" (features) and a target value (class labels). Creating a model that forecasts the target value of data instances in the testing set that are merely given the attributes is the aim of the support vector machine (SVM). A learnt function maps training vectors onto a higher dimensional space, which may even be infinite. Next, in this higher dimensional space, SVM locates a linear separating hyperplane with the largest margin.

3.4 Classification Analysis

In this section, performance evaluation measures such as TP, FP, precision, recall and F-Measure will be calculated to measure classification accuracy and efficiency of DT and SVM classification model. Also the classification accuracy of DT and SVM will be compared. The quality of the DT and SVM classification models will be represented in the form of confusion matrix [15].

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

4.1 Classification using Decision Tree model

The multimedia dataset, which consists of 247 web multimedia-video metadata instances, is taken from the data mining tool in order to assess the effectiveness of the classification models built using Decision Trees and Support Vector Machines. Number of properly categorized cases, number of wrongly classified instances, TP rate, FP rate, precision, recall, and F-score are used to evaluate the model's performance. The Table 3 represents classification result obtained by the Decision Tree classification model.

Table 3: Classification result of Decision Tree classification model

Sl.No	Class Labels	Total Instances	Correctly Classified	Incorrectly Classified	TP	FP	Precision	Recall	F-Measure
1	Sports	85	76	9	76	2	0.974	0.894	0.933
2	News	100	99	1	99	8	0.925	0.99	0.957
3	Entertainment	62	59	3	59	3	0.952	0.952	0.952
	Total	247	234	13	234	13	0.950	0.945	0.947

The Decision tree classification model correctly classifies 234 tuples out of 247 instances, while incorrectly classifying 13 tuples is observed based on the Decision tree experimental result. The class labeled "Sports" has the highest accuracy and precision. Additionally, there is a relatively low fall-positive rate in "sports," correspondingly. Out of 100 records in the "News" Class label, 99 were correctly classified by the DT model, while 1 was wrongly classified. In contrast to the other class labels, the "News" class label has a higher fall positive rate. Decision tree categorization has an overall efficiency of 94.7%.

4.2 Classification using SVM model

The support vector machine experimental result that, out of 247 instances, 220 tuples are correctly classified and 27 tuples are incorrectly classified by the SVM classification model. The class label 'Entertainment' has highest precision and accuracy.

Also the falls positive rate of 'sports' is very less with respectively. Out of 100 entries in the "News" Class label, 91 were correctly identified by the SVM model, while 9 were wrongly classified. Comparing the class labels "News" to the remaining class labels, the falls positive rate for News is higher.

According to the support vector machine experiment, the SVM classification model successfully classifies 220 tuples out of 247 instances, while erroneously classifies 27 tuples. The category label "Entertainment" has the highest accuracy and precision. Additionally, there is a relatively low fall-positive rate in "sports," correspondingly. Nine records in the 'News' Class label were identified wrongly by the SVM model, whereas 91 records in total were correctly classified. In contrast to the other class labels, the "News" class label has a higher fall positive rate.

The overall efficiency of SVM classification is found 89%. The Table 4 represents classification result obtained by the SVM classification model.

Table 4: Classification result of SVM classification model

Sl. No.	Class Labels	Total Instances	Correctly Classified	Incorrectly Classified	TP	FP	Precision	Recall	F-Measure
1	Sports	85	69	16	69	4	0.945	0.812	0.873
2	News	100	91	9	91	16	0.85	0.91	0.879
3	Entertainment	62	60	2	60	7	0.968	0.968	0.93
	Total	247	220	27	220	27	0.921	0.896	0.894

As a conclusion of the examination of the classification result achieved by the DT and SVM model, the quality of the model is represented by the terms of confusion matrix and is displayed in Table 5. The comparison of efficiency from the comparison analysis of classification models shows which model is more effective at classifying multimedia and video content from websites in Table 5.

Table 5: Confusion matrix of DT and SVM classification result

Confusion Matrix	
Support Vector Machine Model ===Confusion Matrix=== b b c ←Classified as 766 3| a= Sports 199 0| b= News 12 59| c= Entrainment	Decision Tree Classification Model ===Confusion Matrix=== b bc←Classified as 7663| a= Sports 199 0 | b= News 1259| c= Entrainment

The Graphical representation of comparison analysis is represented in Figure-4.

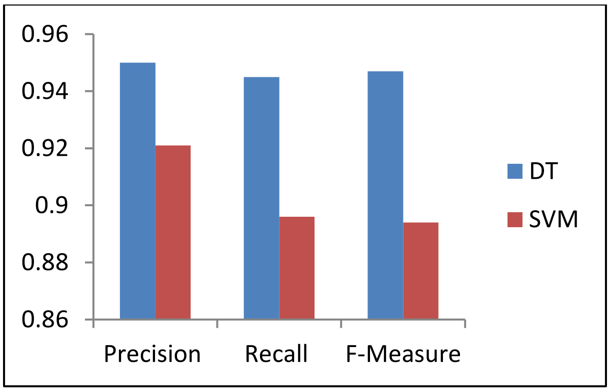

Figure 4: Comparison of classification result

The experimental result shows that, DN classification model works well as compared to SVM classification model. Every attribute is present in the online multimedia-video metadata databases as continuous values. This factor makes the DT classification model more accurate than the SVM classification model. The web multimedia-video metadata datasets contains all independent attribute as continuous values. Due to this factor the SVM classification has less accuracy than DT classification model.

5. CONCLUSION AND FUTURE WORK

In this work, we used online multimedia metadata to classify web multimedia videos according to their categories. In order to classify the web multimedia-video metadata, it is extracted and kept in a database. The web multimedia-videos are to be classified using the Decision Tree (DT) and Support Vector Machine (SVM) methods. When DT and SVM classification models' classification results are compared, it is discovered that the DT classification model is more effective in classifying web multimedia videos using metadata. Additionally, depending on independent features, the SVM classification model performs less well for web multimedia and video. The accuracy of SVM classification model if found 89% and the accuracy of DT classification model is found 94.7%. In the future work, we will explore the effectiveness of SVM classification techniques in multimedia systems with a broader range of web multimedia videos objects and more complex temporal relationships. Additionally, we will analyze the root causes of the notable observations in this study to identify the most valuable attributes of multimedia systems for feature selection and enhance the overall classification process.

6. REFERENCES

[1] Ashwini S. Mane, P. M. Kamde, "Video Classification using SVM", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-2, Issue-3, July 2013

[2] S. Syed Shajahaan1, S. Shanthi2, V. ManoChitra3, "Application of Data Mining Techniques to Model Breast Cancer Data", International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 11, November 2013.

[3] Chen Gang "MediaInfo extractor – A Tool for Media Data Mining", 2011. http://mediaarea.net/en/MediaInfo.

[4] Jun Yang, Rong Yan, Alexander G. Hauptmann, "Cross-Domain Video Concept Detection Using Adaptive SVMs", MM'07, September 23–28, 2007, Augsburg, Bavaria, Germany.

[5] Yimin Wu, Aidong Zhang D, "An Adaptive Classification Method For Multimedia Retrieval" http://www.cse.buffalo.edu/DBGROUP/psfiles/yiminkwu/icme03.pdf

[6] M.Y. Cheng, S. C. Cheung, T.H. Tse," Towards the Application of Classification Techniques to Test and

Identify Faults in Multimedia Systems", To appear in Proceedings of the 4th International Conference on Quality Software, IEEE Computer Society Press, Los Alamitos, California (2004).

[7] Jana Kludas, "Multimedia Retrieval and Classification for Web Content", BCS IRSG Symposium: Future Directions in Information Access (FDIA 2007).

[8] Ludovic Denoyer, Sylvie Brunessaux, "Structured Multimedia Document Classification", DocEng'03, November 20–22, 2003, Grenoble, France.

[9] Lisa Merete Kristensen, Jannick Kirk Sørensen, "Classifying the News: Metadata as Structures

[10] Visibility and Compliance with Tech Standards", in book: Data Publics (pp.145-164), July 2023.

[11] Siddu P. Algur, Prashant Bhat, "The Role of Metadata in Web Video Mining: Issues and Perspectives", International Journal of Engineering Sciences & Research Technology, February-2015.

[12] Vapnik V, "The Nature of Statistical Learning Theory", Springer-Verlag, NY, USA, 2000.

[13] Multimedia classification of movie shots using low-level and semantic features. http://articles.ircam.fr/textes/Delezoide05a/index.pdf

[14] Vakkalanka Suresh, C. Krishna Mohan,R. Kumara Swamy, and B. Yegnanarayana, "Content-Based Video Classification Using Support Vector Machines", Springer-Verlag Berlin Heidelberg, LNCS 3316, pp. 726–731 2004.

[15] M. Hemashree , Shreya Banerjee , (Member, Ieee), And R. Rashmi, "Tagging of Multimedia Contents on the Web 3.0 Using Semantic Artificial Intelligence: A Systematic Literature Review", Digital Object Identifier 10.1109/ACCESS.2025.3597801 2025.

[16] Gennaro, Claudio. Automatic Metadata Extraction and Indexing for Reusing e-Learning Multimedia Objects. 2007. Gennaro, C. (2007). 14https://doi.org/10.1145/1290067.1290072