Multimedia Indexing and Retrieval: Optimized Combination of Low-level and High-level Features

Mohamed Hamroun[1], Henri Nicolas[2] and Benoit Crespin[1]
[1]Department of Computer Science, Limoges University, XLIM Laboratory, Bordeaux, France
[2]Department of Computer Science, Bordeaux University, LABRI Laboratory, Bordeaux, France

Keywords: Multimedia Indexing, Muti-Level Concepts, Multimedia Retrieval, Machine Learning.

Abstract: Nowadays, the number of theoretical studies that deal with classification and machine learning from a general point of view, without focusing on a particular application, remains very low. Although they address real problems such as the combination of visual (low-level) and semantic (high-level) descriptors, these studies do not provide a general approach that gives satisfying results in all cases. However, the implementation of a general approach will not go without asking the following questions: (i) How to model the combination of the information produced by both low-level and high-level features? (ii) How to assess the robustness of a given method on different applications ? We try in this study to address these questions that remain open-ended and challenging. We proposes a new semantic video search engine called "SIRI". It combines 3 subsystems based on the optimized combination of low-level and high-level features to improve the accuracy of data retrieval. Performance analysis shows that our SIRI system can raise the average accuracy metrics from 92% to 100% for the Beach category, and from 91% to 100% for the Mountain category over the ISE system using Corel dataset. Moreover, SIRI improves the average accuracy by 99% compared to 95% for the ISE. In fact, our system improves indexing for different concepts compared to both VINAS and VISEN systems. For example, the value of the traffic concept rises from 0.3 to 0.5 with SIRI. It is positively reflected on the search result using the TRECVID 2015 dataset, and increases the average accuracy by 98.41% compared to 85% for VINAS and 88% for VISEN.

1 INTRODUCTION

In the last decades, video occupies a remarkable place in the information retrieval community. This can be explained by the availability of digitized video data, which is becoming more and more voluminous. This constant growth in the mass of video data causes problems for indexation, retrieval, and navigation.

Finding a desirable video in a huge video archive is a very difficult task. In fact, users waste a lot of time trying to fulfill their needs. Research in the field of optimizing video data search and visualization tools will become a major area of research in the future. Indeed, the use of the video format is more and more common. Also, the need to have powerful tools to exploit this amount of data becomes important. Currently, the majority of different approaches is trying to improve these tools by relying on visual content and semantic content (or low-level and high-level descriptors). Although some works are efficient, they could be improved as for example only a few of them are based on multi-level fusion-based indexing, which implies a combination between low-level and high-level descriptors.

This article deals with the fusion of low-level descriptors and high-level descriptors. We have suggested a "SIRI" indexing approach to facilitate the retrieval and the data usage. The innovative aspect of the proposed system is the combination of low-level and high-level descriptors. The main goal is to guarantee both the accuracy of the results and of the semantics. More precisely, we combine the low-level descriptors we have proposed in other works, PMC and PMGA (Hamroun et al), with high-level ones. Moreover, we merge the descriptors that we have already proven to be effective by integrating the methods related to the ISE, VISEN and VINAS systems.

2 RELATED WORKS

Information retrieval is a set of operations used to retrieve a request through a user interface. First, the

Hamroun, M., Nicolas, H. and Crespin, B.
Multimedia Indexing and Retrieval: Optimized Combination of Low-level and High-level Features.
DOI: 10.5220/0011063000003179
In *Proceedings of the 24th International Conference on Enterprise Information Systems (ICEIS 2022) - Volume 1*, pages 194-202
ISBN: 978-989-758-569-2; ISSN: 2184-4992

user must formulate a query. This operation is evident for text, but it is difficult for images and even more difficult for videos. The query can be expressed by different media. It can be an image, a video, a sound or a sketch. The user can also include keywords. The definition of queries is considered a difficult problem with large-scale video databases. In the following we present several existing forms of queries: image, navigation, textual and conceptual queries.

2.1 Image Queries

This approach consists in introducing an example image which will be used to find similar content. For this kind of query, the indexing phase consists in extracting a number of image descriptors. Color features are among the most important recognition aspects, as color remains an unchanging parameter that is not altered by changes in the image orientation, size or placement (Kundu et al). Content-Based Image Retrieval (CBIR) systems use conventional color features such as dominant color descriptor (DCD) (Wang et al), color coherence vector (CCV) (Pass et al), color histogram (Singha et al) and color auto-correlogram (Chun et al). DCD is about quantifying the space occupied by the color feature of an image by placing its pixels into a measurable number of partitions and calculating the means and ratio of this placement. CCV partitions histogram bins into coherent or incoherent types; the results of this method are more precise since they not only emanate from color histogram classification but also from spatial classification. The accuracy of these results is more visible when it comes to images that contain rather homogeneous colors (Pass et al).

2.2 Navigation Queries

In the last decade several systems have been proposed favoring a search in videos collections by navigation in a tree of concepts. A recent review is given in (Schoemann et al). In (Ben Halima et al), the authors propose a video retrieval system by viewing and navigating in concepts. The proposed navigation module is based on a semantic classification. The strength of this system is the ability to integrate a personalized navigation module. The proposed module is based on a neuronal metaphor. For example, (Villa et al). present a system called "Facet Browser", where searching for videos is based on the simultaneous exploration of several search "facets". In (Ben Halima et al) (Hamroun et al), a suggestion tool is made available to assist the search of videos. Two types of suggestions are considered, textual or visual. The effectiveness of this system is proven by analyzing the log files and with user studies. (Schoeffmann et al). present an "instant video navigation" system that segments the videos starting from sequences visualized by the users. This tool offers two different views for searching videos, parallel or tree-based.

2.3 Textual Queries

Textual queries remain limited and are generally associated with families of specific visual documents, such as TV news. The most promising way in this context regarding videos consist in transcribing the soundtrack to determine its subject (Lefèvre et al), rather than exploiting the visual content.

2.4 Conceptual Queries

Conceptual queries are the subject of many research works. For example, the INFORMEDIA approach uses a limited set of high-level concepts to filter the results of textual queries. This system also creates groups of key-images, using the results of speech recognition to trace the collections of key-images at the re-associated geographic place on a map, and combines this with other visualizations to give the user an arrangement of the query result context. The suggested method in (Worring et al) is based on a process of semantic indexing. This system uses a large semantic lexicon divided in categories and threads to support the interaction. It defines a space of visual similarity, a space of semantic similarity, a semantic thread space and browsers to exploit these spaces. We can also mention the VERGE approach (Stefanos et al), which supports conceptual and high level visual research. This tool combines indexing, analysis and recovery techniques of diverse modalities (textual, visual and conceptual).

Many research works study the audiovisual data search domains on large databases and suggest tools based on diverse research forms. The limit of these techniques often lies at the level of user interactions. Enriching the search system with the past behavior of the users potentially offers more pertinent results. In the literature, only a few works focus on this aspect. The approach suggested in (Faudemay et al) is the first to put the user at the center of the search system. This idea represents a real improvement perspective to extend existing methods.

For the indexing part, most of the works presented above are based on either a low level or a high-level indexation and only a few are based on multi-level indexing techniques. Thus, a very challenging

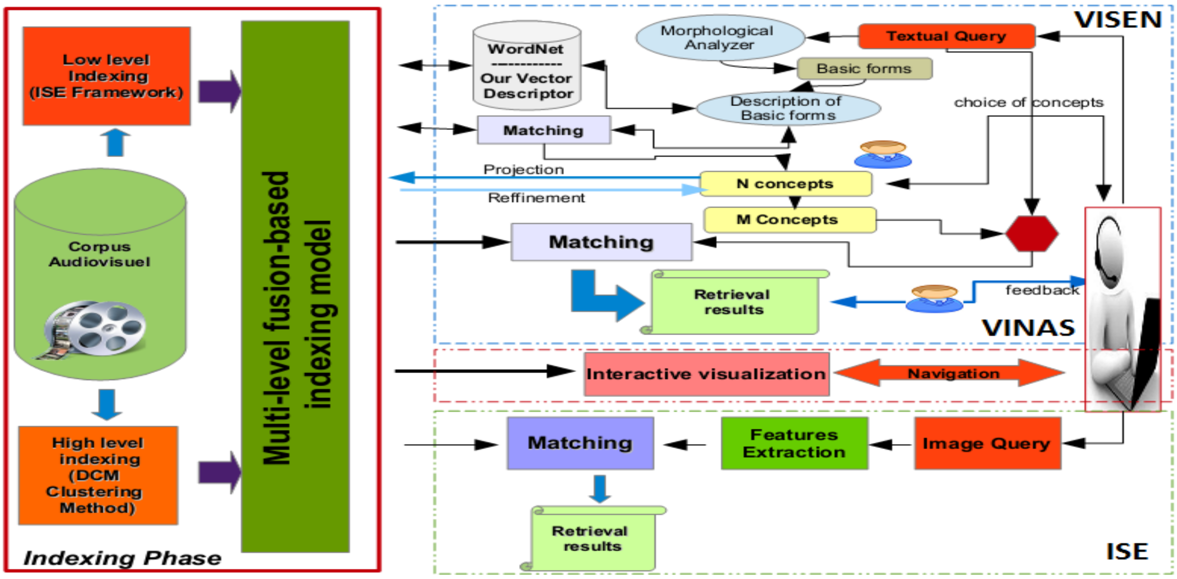

Figure 1: Conceptual architecture of the SIRI Framework.

perspective is to enrich indexing techniques with a multilevel-based indexing model. Information fusion is an area which is undergoing significant development, in particular for multimedia indexing and retrieval. In this field, the data source is multiplied with different characteristics (image, audio, text and movement, etc), but as each data source is generally insufficient, it is important to combine several descriptors in order to gain better knowledge. Several classes of combination methods have been proposed in the literature. In (Kuncheva et al), Kuncheva makes the difference between fusion and selection. Fusion involves combining all the outputs, while selection involves choosing the "best" from a set of possible classifiers to identify the unknown form. (Duin et al) distinguishes 2 types of combination methods in fusion: (l) combination of different classifiers and (2) weak classifiers. The latter has the same structure, but is trained on different data. Other classes are proposed by Xu, which distinguishes combination methods only by the output type of the classifiers (class, measure) presented at the input of the combination. Jain (Duin et al) constructed a dichotomy according to two criteria: the type of outputs from the classifiers and their learning capacity. This last criterion is also used by (Kuncheva et al) to separate different fusion methods. Larning methods allow to find and adapt the parameters to be used in the combination according to the database of available examples. The methods running without learning simply use the outputs of the classifiers without integrating in advance other information on the performance of each of the classifiers.

3 SIRI: TOWARDS A MULTI-LEVEL FRAMEWORK

Figure 1 shows the overall conceptual architecture of our SIRI framework. It includes (i) an indexing phase, which introduces a new merging approach detailed in the following section, (ii) a phase composed of three types of searches: textual, navigational and exemplary image. The textual search is similar to the VISEN system, while the navigation search is inspired by our VINAS tool, and the exemplary image is inspired by our ISE system.

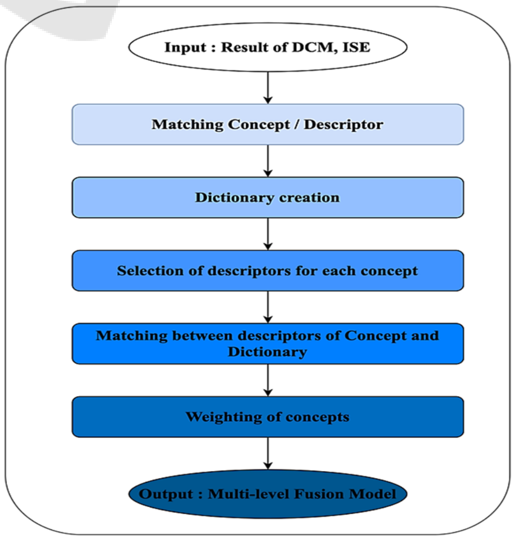

Figure 2: Our multi-level algorithm.

Figure 2 includes an algorithm that provides an overview of our multilevel fusion-based indexing method. The following sections detail each of the steps of this algorithm.

3.1 Step 1: Matching Concept / Descriptor

Each Large scale Conceptual Ontology for Multimedia (LSCOM) is better represented by one or more low-level descriptors. For example, color descriptors may be more determinant for some concepts, such as sky, snow, water landscape or vegetation, but less determinant for studios and meetings. For this purpose, we propose to weight each descriptor at a low level according to its degree of discrimination in relation to the concept. Based on our previous indexing and semantic studies (LAMIRA and VISEN), for each concept we can select the most relevant video shots and then model this relationship in the form of a tree structure (figure 3) where the concept C_1 represents the root of tree A, the vertices represent the set of videos (V_1, V_2 and V50 in our example), and the leaves of A represent the set of shots (P_1, P_{19}, P_{25}, P_{50}, P_{51} in our example). This tree is limited to only 3 levels and has vertices that can be ordered according to their distance from C_1, withthe distance $D(C_1,P_i)$ defined as the weight of the concept C_1 in shot P_i.

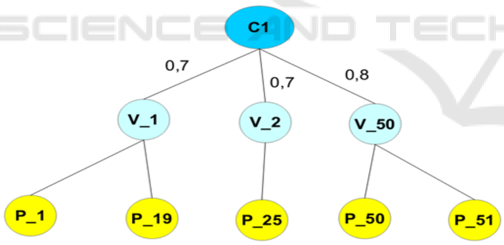

Figure 3: Initial representation of the concept.

After selecting the designs or the weight of C_1 >= 0.7 (value chosen based on an experimental research study), we selected the low-level descriptors that better represent each concept. These descriptors can be of several types, such as: Hsv Histogram, Gabor Texture, Camera Motion and Edge Histogram. Moreover, the choice of a descriptor for a concept is determined by its distance from that concept. This value is chosen according to the SVMs classification on the TRECVID 2015 data. For example, in figure 4, Edge Histogram, which is the most relevant descriptor for all concepts, is characterized by outlines, particularly for the concepts 'Car', 'Maps', Mountain, Sports, Waterscape, unlike Camera Motion, which influences only a few concepts. In fact, it can have the same performance as that of the rest of the descriptors for the 'Walking' and 'Running' concepts.

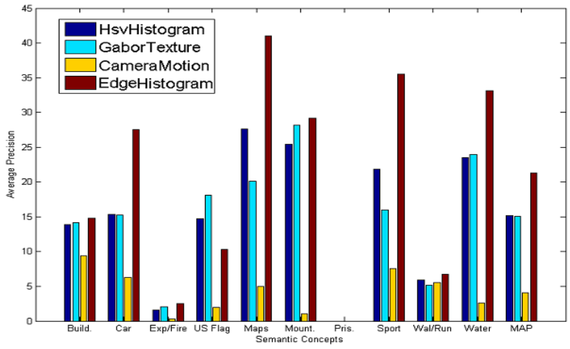

Figure 4: Performance of the SVM classification per concept for the four descriptors.

Table 1 shows the result of the concept/descriptor relationship where each concept is defined by a set of low-level descriptors.

Table 1: List of concepts according to the descriptors used.

Concepts Objet		
Ref.	**Concept**	**Used descriptors**
017	Bridges	Hough transform, PMGA, Gabor
047	Emergency Vehicles	Wavelet transform, SIFT- LBP
043	Dog	SIFT- LBP, PMGA, Gabor
004	Airplane Flying	DCD, CCV, SIFT- LBP, Gabor
123	Two people	Wavelet transform, SIFT- LBP, PMGA, Gabor
019	Bus	Wavelet transform, SIFT- LBP, PMGA
045	Driver	SIFT- LBP, Gabor
117	Phones	Sift, fourrier transform
041	Demonstration Or Protest	Wavelet transform et Gabor
059	Hand	DCD, CCV, SIFT- LBP, PMC, Gabor
081	Mountain	DCD, CCV, SIFT- LBP, PMC ,Gabor
015	Boat Ship	DCD, CCV, SIFT- LBP, PMC , Gabor
053	Flowers	DCD, CCV, PMC, Gabor

3.2 Creating a Visual Dictionary

The first step consists in the extraction of visual descriptors, such as colors, textures, etc, from each corpus shot. It is therefore necessary to submit a summary of the descriptors in a form more usable by our system. For this purpose, a visual dictionary is computed using the K-means clustering algorithm, which has been selected due to its simplicity of implementation and its speed of execution, which enables us to determine a predefined number of centers that best represent the set of descriptors. In our experiments we used 130 centers (also called visual keywords) related to semantic concepts. Figure 5 illustrates the principle of the visual dictionary construction.

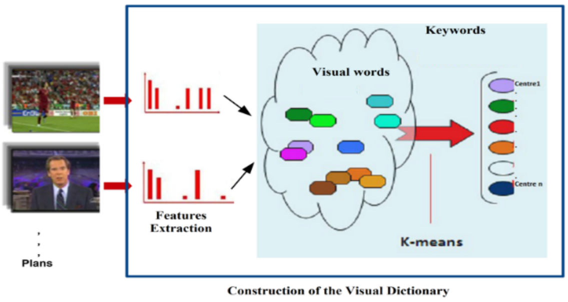

Figure 5: Construction of the Visual Dictionary.

After creating the visual dictionary, which contains all the descriptor vectors for each shot, we connected the closest descriptors (i.e. the descriptors that belong to the same class). Figure 6 shows an example of these relationships. Moreover, the descriptors represented in this example are CCV_{19}, $Histo_{19}$, PMC_1, MC_1, $Histo_1$, etc., which are examples of the descriptors for P_1, P_{19} and P_{50}. Therefore, the closest descriptors in the visual dictionary, such as the histograms 1 and 19, are related. Then, a simple Euclidean distance was computed for each pair of descriptors. Through this relationship and since these descriptors are determinant for P_1 and P_{19}, a relationship is implicitly created between plans 1 and 19.

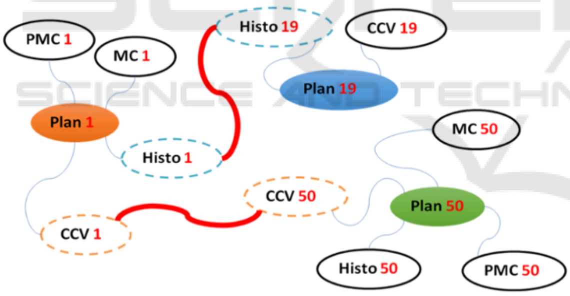

Figure 6: Extraction of the visual dictionary from concept C_1.

3.3 Selection of Descriptors for Each Concept

We are now able to determine the descriptors that best represent a concept for each shot. Therefore, in figure 7 we added the leaves CCV_{19}, $Histo_{19}$, PMC_1, MC_1, and $Histo_1$ to concept C_1 present in plans 1, 19, 25, 50 and 51, building a tree with 4 levels instead of 3.

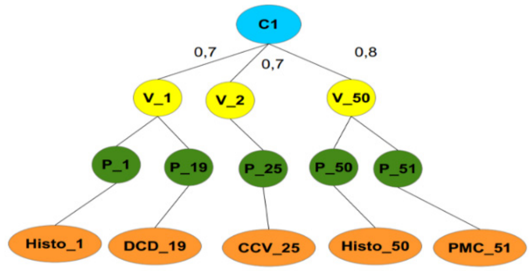

Figure 7: Selection of descriptors for each concept.

3.4 Matching between Descriptors

A matching is carried out between C_1 descriptors (such as $Histo_1$, DCD_{19}, CCV_{25} 1, $Histo_{50}$, etc.) and the dictionary based on the K-means classification algorithm. As shown in figure 8, the content of the dictionary is divided into 5 classes (number of C1 descriptors) with a threshold s <=0.3 chosen according to the experiment.

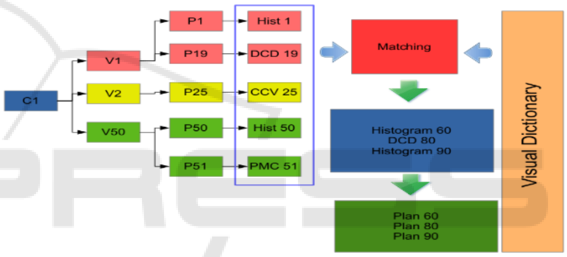

Figure 8: Matching between shots in the K classes.

The K-means algorithm helps group homogeneous shots into K classes. In this method, each class C_i is represented by its gravity center, which is the average of the descriptor vectors of the shots belonging to this class.

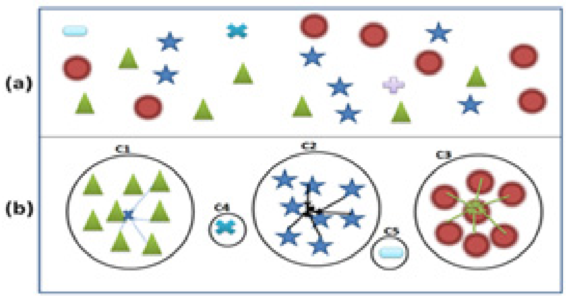

Figure 9: Classification with K-means.

Figure 9 shows how the K-means algorithm operates. We can distinguish 5 classes of shots of which C_1, C_2 and C_3 are three classes that represent similar shots. The gravity centers of these three classes are presented in the form of three vectors reflecting the mean of nearest neighboring descriptors. The other classes C_4 and C_5 group together isolated shots which are not similar. In fact, the objective of this method is to select the class centers so as to reduce the sum of the intra-class inertia within all the classes. The intra-class inertia of class C_i (i<n, where n is the number of classes), denoted Inertia (C_i), is the mean distance between the vectors of the class and its center of gravity g_i. Let P_j be a shot, such that j < m where m is the total number of shots in the set:

$$Inertia(C) = \frac{1}{|C|}\sum_{C_i \in C} Inertia(C_i)$$

$$Inertia(C_i) = \frac{1}{|C_i|}\sum_{I_j \in C_i} d(g_i, I_j)$$

$$g_i = \frac{1}{|C_i|}\sum_{I_j \in C_i} I_j$$

Figure 10 shows the tree resulting from the application of the descriptor classification, where P_{60}, P_{80}, and P_{90} are the new shots indexed by C_1.

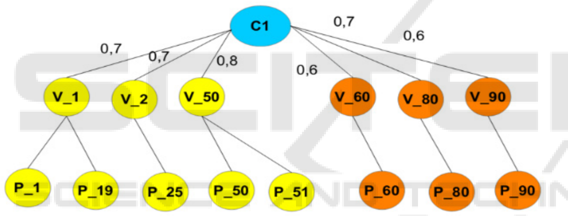

Figure 10: The result of the new classification.

3.5 Weighting

3.5.1 Concept Weighting in a Shot

Furthermore, we propose a new approach to calculate the conceptual weighting by adding the concept of the relationship between low-level and high-level descriptors.

$$W(Ci/Pj) = \frac{\text{Number of plans indexed by Ci and in relation to Pj}}{\text{Total number of plans indexed by Ci}} * \frac{\text{Number of concepts similar to Ci}}{\text{Number of concepts in Pj}}$$

For our example, this value obtained is :

3.5.2 Weighting a Concept in a Video

$$W(Ci, Vj) = \frac{\sum W(Ci/Pj)}{N}$$

with N the total number of shotes in V_j

4 EXPERIMENTATION AND RESULTS

4.1 Datasets

For the TRECVID 2015 Semantic Indexing task there are two data sets provided by the National Institute of Standards and Technology (NIST): one for testing and one for training. The training dataset named IACC.1.tv10.training contains 3200 internet archive videos (50GB, 200 hrs) while the test dataset named IACC.1.A contains approximately 8000 internet archive videos (50GB, 200 hrs), annotated with 130 semantic concepts.

4.2 The User's Interface

The interface shown in figure 11 represents the main menu interface of our image research system. Part (A) helps the user to search per keywords in English, French or Arabic with a relevance feedback mechanism. Then, in part (B), the user can navigate through a hierarchical structure (context/ concept/ image) to get the desired image. Finally, part (C) enables the user to search through an image query from an image database containing 10000 examples.

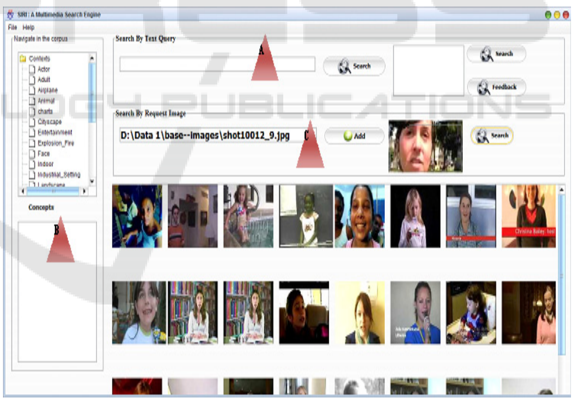

Figure 11: Main menu of the semantic research system.

4.3 Step 1: Our System versus Low Level Systems

The average precision for all the image categories of the Corel-1k dataset is reported in figure 12. To show the utility of our CBIR scheme, the results of nine other CBIR systems are also reported in this table. Since the average precision of our results is 99%, our CBIR scheme has the highest accuracy among the other state-of-the-art CBIR systems. In fact, our proposed CBIR system outperforms that of (Jhanwar

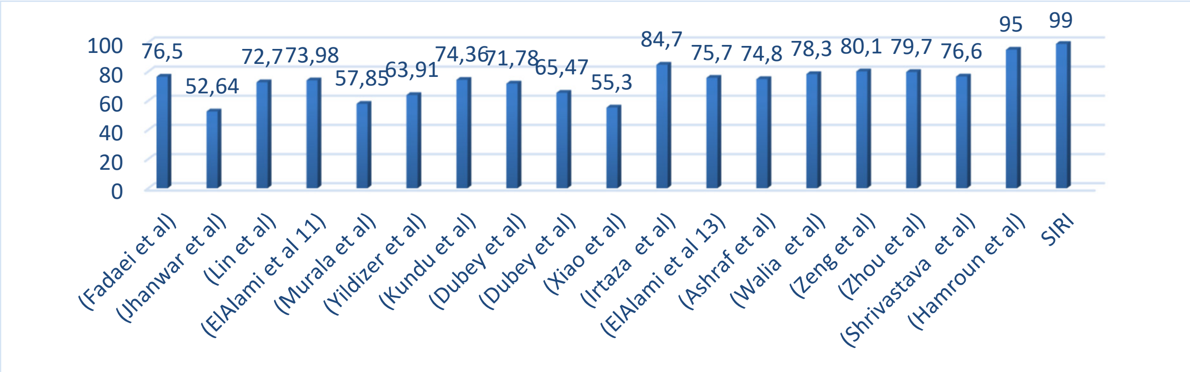

Figure 12: Comparison of the average precision of the previous methods and proposed method.

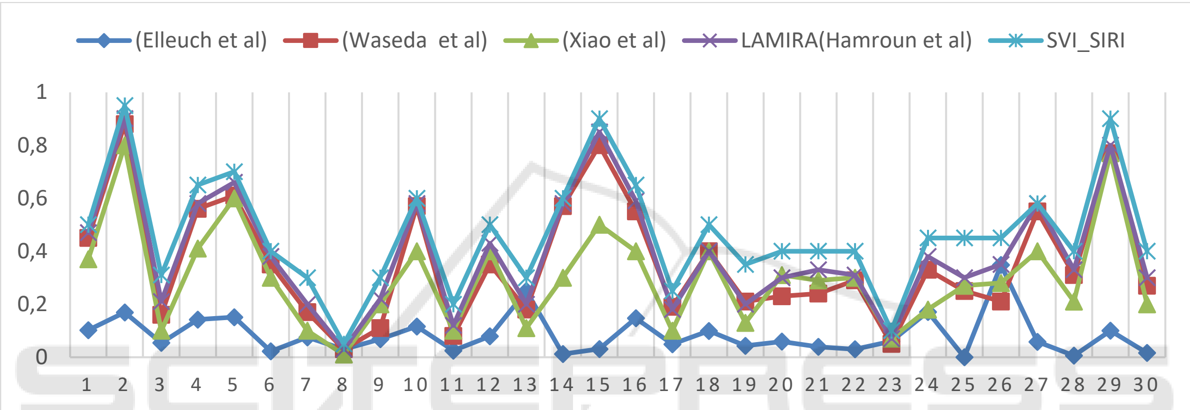

Figure 13: Comparison of the average precision of the (Elleuch et al),(Waseda et al),(Xiao et al), SVI_LAMIRA (Hamroun et al) tools and proposed SVI_SIRI.

et al.), (Lin et al.), (ElAlami et al), (Murala et al.), (Yildizer et al.), (Kundu et al), (Dubey et al), (Aun et al), (Dubey et al), (Ashraf et al), (Walia et al.), (Zeng et al.), (Zhou et al.), (Shrivastava et al.) and (Xiao et al.).

4.4 Step 2: Our System versus Detection of Concepts Systems

Figure 13 represents the weighting level of the 30 concepts. These weighting levels are variable according to the systems. We have compared our SVI_LAMIRA (Hamroun et al) system with (Elleuch et al),(Waseda et al) and (Dubey et al).
For some concepts such as Bicycling, Singing, Telephones and Car Racing, SVI_LAMIRA obtained the best results. Thus, the use of a semantic interpretation process improves the detection of semantic concepts by using our DCM method (Hamroun et al)

4.5 STEP3: Our System versus Height Level Systems

In the 2nd step of the assessment, we compare our system with the most pertinent semantic retrieval systems. The following figure 14 shows some query outcomes.

Based on the histogram shown on figure 14, we note that accuracy values corresponding to the sports and vegetation concepts are equal to 1. If compared to accuracy values corresponding to the works proposed by (Hamroun et al), we can see a significant improvement. We observe that all the accuracy values exceed 0.85, which means that the improvement encompasses all the concepts. It is broadly clear that the suggested Interactive Search technique improves the system performance.

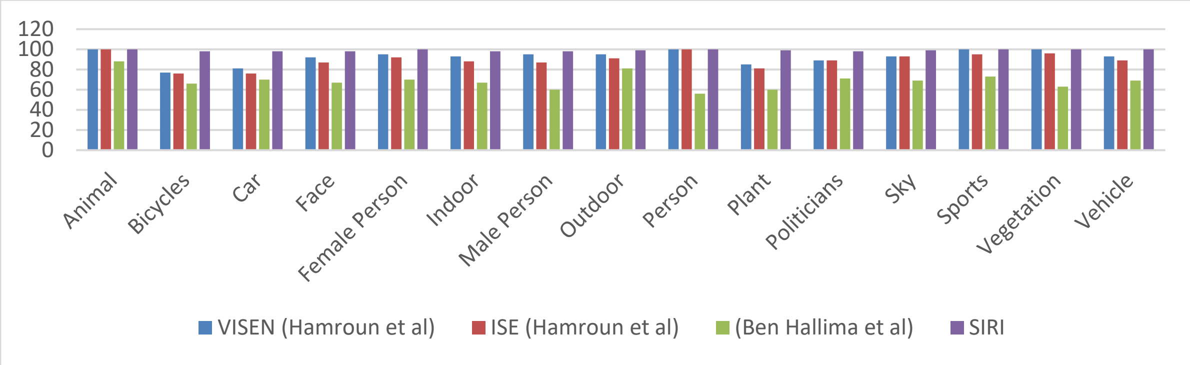

Figure 14: Precision value.

5 CONCLUSION

In this paper, we propose a semantic indexing model implemented through a new framework called SIRI, which represents an extension of VISEN based on a new indexing method that combines low-level and high-level descriptors. On the basis of this combination, SIRI becomes capable of responding to any type of user. Specialists such as medical doctors who search for particular areas in the videos, can retrieve relevant results with image queries. The system response is based on low-level descriptors. However, if we are in the case of a non-specialist user looking for general information on medicine, the research can be based on other forms of queries and some of the semantic descriptors. In future works we will try to improve the request response time by using new algorithms in machine learning

REFERENCES

Kundu, Malay, K., Manish, C., and Samuel, R. (2015). A Dubey, Shiv Ram, Satish Kumar Singh, and Rajat Kumar Singh: 'Local neighbourhood-based robust color occurrence descriptor for colour image retrieval', IET Image Processing, 2015,9, (7), pp. 578–586

Wang, Xiang-Yang, Yong-Jian Yu, and Hong-Ying Yang: 'An effective image retrieval scheme using color, texture and shape features', Computer Standards & Interfaces, 2011,33, (1), pp.59–68

Pass, Greg, and Ramin Zabih: 'Histogram refinement for content-based image retrieval', Pro-ceedings 3rd IEEE Workshop on Applications of Computer Vision (WACV'96), Sarasota, FL, 1996, pp. 96–102

Singha, M., Hemachandran, K. and Paul, A.: 'Content-based image retrieval using the com-bination of the fast wavelet transformation and the colour histogram', IET Image Processing, 2012,6, (9), pp. 1221–1226

Chun, Young Deok, Nam Chul Kim, and Ick Hoon Jang: 'Content-based image retrieval using multiresolution color and texture features', IEEE Transactions on Multimedia, 2008,10, (6), pp. 1073–1084

Hamroun. M, Lajmi. S, Nicolas. H, Amous,VISEN: A Video Interactive Retrieval Engine Based on Semantic Network in large video collections, International Database Engineering & Applications Symposium. IDEAS 2019: 25:1-25:10.

Hamroun. M, Lajmi. S, Nicolas. H, Amous, Descriptor optimization for Semantic Concept Detection Using Visual Content, International Journal of Strategic Information Technology and Applications IJSITA 10(1): 40-59 (2019).

N., Elleuch, A., Ben Ammar and A., M., Alimi. (2015). A generic framework for semantic video indexing based on visual concepts/contexts detection. In Mutimedia Tools and application.

S. Fadaei, R. Amirfattahi, et M. R. Ahmadzadeh, «New content-based image retrieval system based on optimised integration of DCD, wavelet and curvelet features», IET Image Processing, vol. 11, no 2, p. 89-98, 2017, doi: 10.1049/iet-ipr.2016.0542.

K. Schoemann, et al. 2010. Video browsing interfaces and applications: a review. SPIE Reviews.

M. Ben Halima, M. Hamroun, S. Moussa and A.M. Alimi, 2013 "An interactive engine for multilingual video browsing using semantic content", International Graphonomics Society Conference IGS, Nara Japan, pp 183-186.

R. Villa, N. Gildea, and J. M. Jose. 2008. Facetbrowser: a user interface for complex search tasks. In MM '08: Proceeding of the 16th ACM international conference on Multimedia, New York, NY, USA, pp 489–498.

M. Del Fabro, K. Schoeffmann, and L. Boeszoermenyi, 2010, Instant Video Browsing: A tool for fast Nonsequential Hierarchical Video Browsing. In Workshop of Intercative Multimedia Applications.

C.-H. Lin, R.-T. Chen, et Y.-K. Chan, « A smart content-based image retrieval system based on color and texture feature », Image and Vision Computing, vol. 27, no 6, p. 658-665, mai 2009, doi: 10.1016/j.imavis.2008.07.004.

M. E. ElAlami, « A novel image retrieval model based on the most relevant features », Knowledge-Based Systems, vol. 24, no 1, p. 23-32, févr. 2011, doi: 10.1016/j.knosys.2010.06.001.

S. Murala, R. P. Maheshwari, et R. Balasubramanian, « Local Tetra Patterns: A New Feature Descriptor for Content-Based Image Retrieval », IEEE Transactions on Image Processing, vol. 21, no 5, p. 2874-2886, mai 2012, doi: 10.1109/TIP.2012.2188809.

E. Yildizer, A. M. Balci, T. N. Jarada, et R. Alhajj, « Integrating wavelets with clustering and indexing for effective content-based image retrieval », Knowledge-Based Systems, vol. 31, p. 55-66, juill. 2012, doi: 10.1016/j.knosys.2012.01.013.

M. K. Kundu, M. Chowdhury, et S. Rota Bulò, « A graph-based relevance feedback mechanism in content-based image retrieval », Knowledge-Based Systems, vol. 73, p. 254-264, janv. 2015, doi: 10.1016/j.knosys.2014.10.009.

S. R. Dubey, S. K. Singh, et R. K. Singh, «Boosting local binary pattern with bag-of-filters for content based image retrieval», in 2015 IEEE UP Section Conference on Electrical Computer and Electronics (UPCON), 2015, p. 1-6, doi: 10.1109/UPCON.2015.7456703.

S. Lefèvre, C. L'Orphelin and N. Vincent : "Extraction multicritère de texte incrusté dans les séquences vidéo". Colloque International sur l'Ecrit et le Document (CIFED), 2004.

M. Worring, C. Snoek, O. de Rooji, G. P. Nguyen, R. Van Balen and D. Koelna: "Médiamill : Advanced browsing in news vidéo archives". CIVR, pages 533-536, 2006.

V. Stefanos, M. Anastasia, K. Paul, D. Anastasios, M. Vasileios and K. Ioannis: "VERGE: A Video Interactive Retriveal Engine", 2010.

P. FAUDEMAY ans C. SEYRAT: " Intelligent delivery of personalised video programmes from a video database" International workshop on Database anx EXpert systems Applications, 172-177, 1997

L-I. Kuncheva, J-C. Bezdek, and R-P. Duin. Decision templates for multiple classier fusion: An experiemental comparaison.Pattern Recognition, 34 :299{ 314, 2001

A-K. Jain, R-P. Duin, and J. Mao. Combination of weak classiers. IEEE Trans. on Pattern Analysis and Machine Intelligence, 22(1), 2000

L-I. Kuncheva. Fuzzy versus nonfuzzy in combining classiers designed by bossting. IEEE Transactions on fuzzy systems, 11(6), 2003.

Waseda at TRECVID 2015: Semantic Indexing, Kazuya Ueki and Tetsunori Kobayashi, TREVVID 2015.

Qualcomm Research and University of Amsterdam at TRECVID 2015: Recognizing Concepts, Objects, and Events in Video, C.G.M. Snoek, S. Cappallo, D. Fontijne, D. Julian, D.C. Koelma, P. Mettes, K.E.A. van de Sande, A. Sarah, H. Stokman, R.B. Towal, TRECVID 2015.

S. R. Dubey, S. K. Singh, et R. K. Singh, « Rotation and scale invariant hybrid image descriptor and retrieval », Computers & Electrical Engineering, vol. 46, p. 288-302, août 2015, doi: 10.1016/j.compeleceng.2015.04.011.

Hamroun. M, Lajmi. S, Nicolas. H, Amous. I, ISE: Interactive Image Search Using Visual Content: In Proceedings of the 20th International Conference on Enterprise Information Systems (ICEIS 2018) - Volume 1, pages 253-261. ISBN: 978-989-758-298-1 (ICEIS 2018).

M. Hamroun, S. Lajmi, H. Nicolas and I. Amous, “An Interactive Video Browsing With VINAS System”, In Proceedings of the 15th ACS/IEEE International Conference on Computer Systems and Applications AICCSA 2018, Aqaba, Jordan.

N. Jhanwar, S. Chaudhuri, G. Seetharaman, et B. Zavidovique, « Content based image retrieval using motif cooccurrence matrix », Image and Vision Computing, vol. 22, no 14, p. 1211-1220, déc. 2004, doi: 10.1016/j.imavis.2004.03.026.

Y. Xiao, J. Wu, et J. Yuan, « mCENTRIST: A Multi-Channel Feature Generation Mechanism for Scene Categorization », IEEE Transactions on Image Processing, vol. 23, no 2, p. 823-836, févr. 2014, doi: 10.1109/TIP.2013.2295756.

A. Irtaza et al., « An Ensemble Based Evolutionary Approach to the Class Imbalance Problem with Applications in CBIR », Applied Sciences, vol. 8, no 4, p. 495, avr. 2018, doi: 10.3390/app8040495.

M. E. ElAlami, «A new matching strategy for content based image retrieval system», Applied Soft Computing, vol. 14, p. 407-418, janv. 2014, doi: 10.1016/j.asoc.2013.10.003.

R. Ashraf, K. Bashir, A. Irtaza, et M. T. Mahmood, «Content Based Image Retrieval Using Embedded Neural Networks with Bandletized Regions», Entropy, vol. 17, no 6, p. 3552-3580, juin 2015, doi: 10.3390/e17063552.

E. Walia et A. Pal, «Fusion framework for effective color image retrieval», Journal of Visual Communication and Image Representation, vol. 25, no 6, p. 1335-1348, août 2014, doi: 10.1016/j.jvcir.2014.05.005.

S. Zeng, R. Huang, H. Wang, et Z. Kang, «Image retrieval using spatiograms of colors quantized by Gaussian Mixture Models», Neurocomputing, vol. 171, p. 673-684, janv. 2016, doi: 10.1016/j.neucom.2015.07.008.

Y. Zhou, F.-Z. Zeng, H. Zhao, P. Murray, et J. Ren, «Hierarchical Visual Perception and Two-Dimensional Compressive Sensing for Effective Content-Based Color Image Retrieval», Cogn Comput, vol. 8, no 5, p. 877-889, oct. 2016, doi: 10.1007/s12559-016-9424-6.

N. Shrivastava et V. Tyagi, «An efficient technique for retrieval of color images in large databases», Computers & Electrical Engineering, vol. 46, p. 314-327, août 2015, doi: 10.1016/j.compeleceng.2014.11.009