# VISUAL INTERFACE FOR THE CONCEPT DISTRIBUTION ANALYSIS IN VIDEO SEARCH RESULTS
## *Multi-Concept Relevance Map (MCRM)*

Alain Simac-Lejeune
*Litii, Alpespace, 15 rue Saint Exupéry, 73800 Frànçin, France*
*alain.simaclejeune@litii.com*

Keywords: Information Retrieval, Visualization, Search Results, Visual Interface.

Abstract: Video media are now emerging with the democratization of broadband, capture tools and content delivery services. Despite the current performance of the 'traditional' search engines, the video search engines are challenged by the semantic difference between requests made verbatim and the data as video. Thus, most engines are not based on content but on the tags associated with documents. The approach presented in this paper is based on extraction of high-level semantics information and offers a visualization of the results based on the content, representativeness, temporal location and redundancy, which ensures the users a high level of relevance of their research results.

## 1 INTRODUCTION

In Internet, multimedia content are more and more visualized with nearly 10 million of sequence read per day to YouTube for example. So, tools must be developed in order to index and visualize sequences available for a given query. YouTube offers such 45TB sequence indexed (Saxena, M. et al., 2008). However, only a handful of sequences correspond to actual demand for a given query. Several proposals have already been made visualizations such as (Luo, H. et al., 2007 and 2008) or (Ji Soo Yi, J. et al., 2005). But, generally, the proposed visualization systems cannot reliably select the most appropriate sequence or can forgeta sequence or relevant or moreover having spent some time to see the full results. The first part of this paper presents the research systems of information and current methods for indexing and tagging. The second part proposes a visualization method for assessing the relevance of a result based on properties of temporal location of a concept and methods of relevance representation. Finally, we conclude on the prospects and future work.

## 2 BACKGROUND

This section presents the current methods for indexing video and the methods of visualization of the proposed results.

### 2.1 Information Retrieval

The aim of information retrieval is to help users to access efficiently to large collections of objects in order to satisfy their information needs (Cha, M. et al., 2007).

The current method of research is based on search of keywords in a database of tags (keywords associated with a number of documents). The results are directly related to the query but also the tags. Moreover, these methods are problematic for the creation of a classification (based only on the tags) and especially the question of the relevance of a result. In these systems, the relevance of a document is often associated with the number of people who have consulted it, the notation made by users (with which rules?) and the age of the document, or rather to its novelty but rarely to the quality of semantic content from the request.

### 2.2 Indexation and Tagging

#### 2.2.1 Tagging

The method currently most used in search engines is the tagging (Cheng, X., 2007) assigning one or more tags in a sequence. The assignment of the tag is carried out manually or by an expert (in systems dedicated to a specific application in most cases) or by any user (mostly). There is no guarantee on the quality of the tag to the content. This is the main reason of the low correlation between results and the given query.

The problem is also a lack of rule in the creation of tags (Keelan, J. et al., 2007). For example, 'swimming' may be associated with swimming lessons, swimming events or any other activity that is linked, but also to documents unrelated but which has the tag in their list of keywords because a user has added.

### 2.2.2 Indexation

The automatically or semi-automatic indexing method for data assigns automatically tags to each sequence and can be done in two ways: by expertise in defining the rules or by learning a number of examples matching a tags (say positive) and contrary examples (called negative). The main drawback of this method is the relatively small coverage of tags. Thus, only very specific systems can currently use this method.

## 2.3 Concept Definition

In (Simac-Lejeune, A. et al., 2010), we proposed a model representation of concepts in the form of spatio-temporal block. This model is built starting from primitives extracted directly from the image by increasing the level of semantic combination. The construction is performed by a system of interaction between the system and the user in the form of questions and answers. Then the system allows the detection of spatio-temporal concepts, allowing a representation in the form of Tilebars.

## 2.4 Results Visualization

Results from search engines commonly used like YouTube or Dailymotion all present the same screen: a key frame and duration (Figure 2). It is therefore difficult to check the accuracy of the application content highly dependent on the quality of the selected picture for the call. The number of visits is often put forward for information of relevance.

# 3 VISUALIZATION

This section presents a proposed solution to the problem described previously on the difficulty to present search results of video documents to users.

## 3.1 Proposed Method

The proposed method is based on two main ideas. The first is that the temporality of the concept in a desired sequence is important in terms of duration and location. The second is that you must submit documents by providing a test of relevance for each results based on content analysis in relation to the request.

In the method, the available information is both based on the semantic content (this concept) and the temporal information (presence, position). The approach is synthesized in two hypothesis:

- The **time of occurrence or frequency of occurrence** of a concept in a document provides information on the relevance of the document from the concept. For example, if we find in a document half the time the language (textual or visual) of love, it is likely that the paper focuses on this theme. However, the duration of a document also has an impact because the time of occurrence is highly dependent on the total length;
- The **relationship between** the documents found and the components of a request must be presented in a compact, consistent and accurate so that the user immediately identifies why a result has been proposed.

The representation is intended to indicate simultaneously: **length** of a document; **distribution** of occurrence or concepts; and **relevance** of the document in relation to the request.

## 3.2 Color Map: A Relevance Indicator

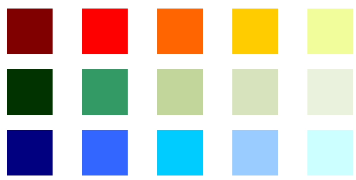

Figure 1: Magnification of Overview Map.

The HotMap color scale visually represents how often a term appears in each search result.

Similarly to what has been proposed by (Hoeber, 2006) with HotMap, color is used to provide two pieces of information. The color gives information on the presence of concept. Each concept is associated with a range of color and a single (Figure 1). The intensity of the color of a given concept provides information on the level of relevance.

This representation allows to determine several concepts (one color by concept) and the degree of relevance found for the segment (color intensity).

For example, using the map (Figure 2), and seeking 'running' the concept, we obtain the following results: **Red** for sequences with the actually concept 'running'; **Orange** for sequences of walking, a concept close to ‘running’; and **Yellow** for sequences whose movement is considered to be approaching ‘running’ with such bike sequences.

### 3.3 Distribution Indicator

Like what Hearst (1995) offered with Tilebars for the representation of the distribution of information in the texts, the temporal distribution of concepts in the videos can be represented as a space within a geometric shape such as a rectangle. A rectangle is a video and its length represents the duration of the presence of the concept in the total length of the document. In this representation, the colored segments indicate the presence of a concept. This method allows to assign a value distribution for each concept and visualization of relationships between concepts in the same sequence.

### 3.4 Duration Indicator

To represent the length, we propose to use just the horizontal size of the bar representing the concept. In the case of multiple documents with different sizes and if the party making the concept is small, it will zoom and if you have several parts with the concept, there will be a bar cutting across all parties and zoom in on one of them to report an approximation.

### 3.5 Relevance Value Processing

From all available information (time, position, distribution, concepts, etc.), we can establish a value of relevance of the classification of documents for a given query. The calculation of relevance is made for each concept. This calculation is relevant in several stages.

The first step (Equation 1) is to establish the level of presence of a concept in a sequence. We can define the level of presence as the number of image with the concept:

$$presence = e^{Mimages} \tag{1}$$

From the level of presence, we can then determine the level of relevance (Equation 2) as the ratio between the presence and size of the video:

$$relevance = {}^{presence}/_{e^{Nimages}} \tag{2}$$

Thus, a concept present in the whole image has a coefficient of relevance 1. The last step is the calculation of the level of relevance of the sequence for a given query (Equation 3) and is defined as the weighted sum of relevance of each concept in relation to the number of concept:

$$level = \frac{\sum_{i=0}^{Nimages} relevance_i * quantity_i}{Nconcept * 10} \tag{3}$$

## 4 EVALUATION

This section presents the evaluation of the proposition.

### 4.1 Video and Concept Database

In order to characterize the performance of the proposed approaches, we used the sequences of a database of existing videos, that all are available in a base consisting by the University of Florida (University of Central Florida): UCF Sports Action 50 (http://server.cs.ucf.edu/~vision/data/). It is composed of 5000 sequences collected on YouTube and spread over 50 categories of 100 sequences per category (listed 5). Each sequence lasts four seconds.

This database was supplemented by external database composed of 100 sequences, illustrated the 50 actions of UCF-SA, but from YouTube.

### 4.2 Search Visualization Evaluation

As we seen, our model allows together the visualization of the temporal distribution of concepts in a video, the distribution of different concepts in a document, the level of relevance of the concepts found in relation to research, document size and then the overall level of relevance of the document from the search.

Table 1 shows the results of relevance for three specific concepts and the average relevance (50 concepts). These results demonstrate that the system provides interesting performance as the results show relevance of the order of 85% on average when we are generally around 40% in searches from YouTube. The display allows quickly seen the relevant sequences. In our preliminary tests, no video viewed did not present the concept sought.

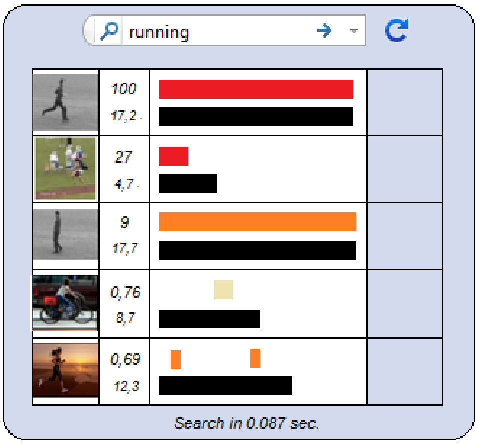


Figure 2: MCRM Video Search Engine on 'Running'.

Table 1: Results of Relevance for Search.

| Concept | Relevance of results |
|---|---|
| running | 88% |
| swimming | 78% |
| diving | 96% |
| average | 85% |

## 5 CONCLUSIONS

We have introduced a new tool to represent the search results of video, called Multi-Concept Map Relevance (MCRM), based on tools named HotMaps and TileBars, created for the representation of search results of textual records. This tool lets you to instantly see the relevance of a document without having to watch it.

However, this method is directly dependent on the way to index documents. It is necessary to have the location information in a temporal sequence of concepts, what is not available at present.

This representation would bring clarity and precision in research, but requires an adaptation of our current methods of indexing and requires either the removal of tagging (or replace it with a more precise marking and therefore more difficult to achieve).

Finally, this method will not replace current methods based on tags for data quality research but could be a very interesting engine dedicated to more specific applications (television, festivals) or the concept of numeric DVR.

## REFERENCES

Cha, M. et al. (2007). I tube, you tube, everybody tubes: analyzing the world's largest user generated content. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet measurement*, pp.1-14.

Cheng, X. (2007). Understanding the characteristics of internet short video sharing: YouTube as a case study. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet measurement*, pp.15-28.

Hearst, M. (1995). TileBars: visualization of term distribution information in full text information access. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp.59-66.

Hoeber, O. and Yang, X. (2006). A Comparative User Study of Web Search Interfaces: HotMap, Concept Highlighter, and Google. In *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, pp.866-874.

Ji Soo Yi, J. et al. (2005). Dust & Magnet: Multivariate Information Visualization using a Magnet Metaphor. *Information Visualization*, Vol. 4, No. 4, pp. 239-256.

Keelan, J. and al. (2007). YouTube as a source of information on immunization: a content analysis. In *JAMA*, pp.1424-1425.

Luo, H. et al. (2007). Mining Large-Scale News Video Databases for Supporting Knowledge Visualization and Intuitive Retrieval. *Proc. IEEE Symposium on Visual Analytics Science and Technology*, pp.107-114.

Luo, H. et al. (2008). Integrating Multi-Modal Content Analysis and Hyperbolic Visualization for Large-Scale News Videos Retrieval and Exploration. *Image Communication* 23(7), pp.538-553.

Roth, V. (1999). Content-based retrieval from digital video. *Image and Vision Computing*, pp.17-531.

Saxena, M. et al. (2008). Analyzing video services in Web 2.0: a global perspective. *Proceedings of the 18th International Workshop on Network and Operating System Support for Digital Audio and Video*, pp.39-44.

Simac-Lejeune, A. et al. (2010). Spatio-temporal block for video indexation assistance. *KDIR*, pp.475-480.