Interactive Components For Visual Exploration of Multimedia Archives

ML Viaud, J. Thièvre, H. Goëau, A. Saulnier, O. Buisson
Research Department
INA, 4 avenue de l'europe
943600 Bry/Marne,
33 (0)1 49 83 20 00,

mlviaud@ina.fr

ABSTRACT

With the increase of online resources, one main challenge for multimedia content providers is to provide efficient and user friendly tools for a deep and shallow navigation adapted to large scale audiovisual content. This paper describes a generic framework to build visual interactive applications the objectives of which are to enhance the understanding and to allow easy access to multimedia resources and management. Visual Maps are built on multi-modal similarity matrices computed from automatically extracted descriptors and use graph clustering and layout methods. Active relevance feedback methods are applied to allow users to control the maps evolution according to their needs. The First results of users' evaluation are presented for one of our tools.

Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]: Content analysis and Indexing, *indexing methods* Information Search and retrieval *Clustering, Information filtering, Relevance feedback*, Digital Libraries *Collection*

I.4 [Image processing and computer vision] feature measurement, *feature representation*

H.5 [Information interfaces and presentation] User Interfaces, *Evaluation/methodology, Graphical user interfaces*

General Terms

Algorithms, Documentation, Theory, Experimentation, Human Factors.

Keywords

Interactivity, cross modal descriptors, active learning, information visualization, index structuring, scalability, multimedia archives.

1. INTRODUCTION

The role of the French Audiovisual Institute (INA) is to store and preserve the French audiovisual heritage, to assure its exploitation and to make it more readily available. Moreover, under the terms of the French law of 20 June 1992, INA became the French audiovisual legal deposit, and is responsible for collecting and preserving radio and television broadcast archives and audiovisual documents. To achieve this mission, 150 INA's media librarians annotate the audiovisual documents received daily, at different levels of precision. The archives contain 1 500 000 pictures and 1 300 000 hours of video (400 000 digitized) and more that one million of documentary notes. Today, INA records continuously 51 TV and 20 radio channels. In 2010, 100 TV and 40 radio channels will be collected. The entire collection of INA's archives is available for professionals or researchers by subscription. Since April 2006, INA's public web site provides access to more than 20000 hours of Radio and TV, representing about 85000 extracts or programmes.

Like most media content owners, INA has followed the evolution of Network, Media Asset Management, format and storage technologies. However, if the whole process of archiving has improved considerably in the last few years, new challenges are appearing as a counterpart to these evolutions. New constraints and paradigms are emerging from the exponential growth of resources, the new business and institutional opportunities and user practices.

On the one hand, INA as a national archive institute has the duty to guarantee reliable access to its resources. In the last few years, user practices have evolved with online services. However, INA's first user targets are movies and TV producers, broadcasters, film directors or archivists who are

using archives for new productions. They are very compelling about the quality of the data access. In fact, they may very well know the resources they want and their characteristics (author or producer names, or date of broadcasting….). They ask for efficient search processes based on semantics. To achieve such objectives, INA has to maintain a high quality coherent annotation process the goal of which is to associate with each resource a set of descriptive textual data. In our context, like for libraries, inaccessible resources are lost resources. Two phenomena can make a resource inaccessible. If the annotation is not precise enough, the results set is too large to be parsed and archivists are facing "*documentary noise*". Conversely, if the annotation is too precise, the corresponding document will not be retrieved easily by query: this phenomenon is called "*documentary silence*". Moreover, natural language is ambiguous and evolves over time. Two strategies ensure the coherence and the relevance of the annotation process: the use of a thesaurus, which normalizes text fields, and the structuring of the annotations. If full text search represents most of the queries in the initial stage, the first evolution that INA had to carry out on its professional web site to satisfy user demands was to provide advanced search forms to allow query by fields.

On the other hand, the general public creates new paradigms for resource access. Discovering TV archives becomes a recreative process: search strategies are totally different, mainly based on the same rules as web search. Users will follow advice from friends, will have *butterfly type navigation* with in the resources or will try loose queries…So, open attractive interfaces and navigation modes are required.

As a national archive institute, INA is more and more often asked to give analyses on its resources. For example, INA may receive requests of different kinds, related to judicial topics (which programme a channel was broadcasting on, on a given date?) or media analysis (what is the general trend for prime time programmes? or the duration of TV speeches for each presidential candidate). Moreover, in the last few years, on-line policy and new business opportunities have also brought new trends in the archive process. For example, short programme excerpts would be more adapted to multi support broadcasting; fast reaction to news events is needed as for press agencies. These evolutions have led to a new structuring of the corpora: TV programmes are segmented and pre-organized in specific collections, such as "songs" or "news events" corpora.

All these new trends, combined with the growth of input documents and constant human resources, make the use of new technologies an inevitable evolution. Needs of new tools are clearly expressed by archivists to assist them in the creation of video and radio excerpts, to facilitate content annotations tasks, to allow new types of search and navigation and finally, to monitor and analyse sets of data. Our applications are mainly semi-automatic to assist archivists in their tasks without loss of control and quality. Then, they can focus on tasks with high added-value: the production of high-level semantic annotations.

2. DATA & FRAMEWORK DESCRIPTION

In the following, we will call multimedia document or entity a media resource (still image, video or radio segment) associated with its textual description. Textual descriptions may present different levels of precision. The shortest notice contains only few factual fields like "title", "channel"," date" or "time", while the most descriptive ones, such as TV news programmes, are described with precise textual content: keywords, people and location names, and free text summaries.

Our framework contains different types of modules:

- Several *indexation modules* analyse each resource to extract cross modal low-level features. These processes are executed off-line and specific files containing the low-level global or local descriptors are generated and associated to the original resource files. Mono-modal or cross-modal distance matrices are computed from the descriptors.

- An *index structuring and search* module allows time efficient searches in huge sets of indexes (1 billion 20D-descriptors). In fact, features extraction generates a huge amount of indexes. There is a trade-off between the potential quality of the retrieval and the quantity of indexes which describe the resources.

- A Graphic User Interface module, which provides users with interactive visualisations to get an overview, to analyse or navigate in sets of structured or non-structured items. The size for the document set is for the moment limited to 35000 visualized entities.

- An active learning module provides resource structuring functionalities according to user action: The originality of our approach is that the system gives diagnosis in terms of positiveness, rejection and ambiguity, allowing users to better understand and control their actions.

The main requirement for the framework was to support collaborative and iterative developments and remote deployment of resources and services. We have chosen Web services technology which allows both of these capabilities: several teams may work on different components of the global system at the same time and the efforts for the integration steps are minimized. Such a framework implies common specification of resources and

services. Typically, these specifications are written in XML, with the advantage of being independent from any programming language. This independence allows the integration of partners' components that are not designed to work in our framework. Another advantage of web services oriented framework is the possibility for each team to preserve their core components hidden from other teams.

The integration of several distributed components in a common application consists in our case in building a web user interface which consumes various services and create a specific representation of the results.

2.1 Low level features extraction and similarity matrix generation modules

We have developed several modules of extraction of global visual low level features such as histograms in different colour spaces (RGB, LUV and HSV), gradient orientation histogram, corner detection or motion estimation. It is possible to compute various distances on colour histograms such as Bhattacharyya, classical Euclidean and Manhattan distances… Some are more noise robust, others are faster to compute, some are closer to human perception, and at least their effectiveness depends on data distribution. In an archive context, video formats are numerous, potential deteriorations may occur, and so colour quality or resolution may influence the distance choice.

In [1], we proposed a new description of video sequences, which is based on the dynamic content. The visual features are composed of local jets around Harris points and Symmetry points. For the indexing stage, the features are computed in each frame and tracked to build the trajectories. However, the trajectory is not used directly as a local feature but simply used to compute a bounding box saved as metadata. The feature itself is based on the local visual content only, by averaging the local jets along the trajectory. During the search stage the visual features are computed only in key-frames and the trajectories are not constructed. Then, a specific registration algorithm allows matching the points' positions of the queries with the indexed trajectories. This asymmetric strategy drastically limits the computational costs of the search process. It also provides accurate and flexible tuning during the search process: the rate of extracted local features can be changed on-line, and then the granularity of the targeted video segments can be adapted to the application.

However, we are also using efficient results from other teams for audio, text and visual indexes. In the context of The European project VITALAS, visual indexes are built by IMEDIA/INRIA [2, 3] team. audio indexes are generated by Fraunhofer [4].

2.2 Index structuring module

INA has proposed new methods of probabilistic similarity search based on distortion. In these techniques, the probabilistic model does not model the distribution of the features in the dataset, but rather the distribution of neighbours relevant for a given query. Whereas clustering based techniques are probabilistic versions of K nearest neighbour queries, distortion-based techniques are more related to range queries, since the model is agnostic to the dataset content.

In [5], Joly et al. defined distortion-based probabilistic queries relying on the distribution of the relevant similar features for finding a transformed image or video. Finding that the probability of a tolerated transformation decreases when the "amplitude" of the transformation increases, Joly et al. proposed to model the effect of tolerated transformations on a signature by an isotropic multidimensional Gaussian probability density function, and to perform probabilistic retrieval based on this model transformed document. They use a space partition index structure based on the Hilbert space filling curve. Probabilistic retrieval then consists in selecting a minimum number of cells such that their cumulated probability (following the model) is above a fixed threshold. They show in [6] that the search cost of the technique is sub-linear in database size up to a given size and report some real-world experiments on very large video databases including several tens of thousands of hours of video (more than 1 billion 20-dimensional features).

In [7], Poullot et al. proposed some improvements of the technique (ZPSS, for Z-curve Probabilistic Similarity Search). To speed up the computation of the keys (cell addresses) they used a Z-grid that appeared to be more efficient than the Hilbert curve. To improve the balancing between the populations of the cells, they also defined an adaptive version of the Z-grid taking into account the distribution of the dataset along each component. The probabilistic retrieval is also improved by ranking the component-wise exploration of the feature space according to decreasing uniformity of the components. This allows improving significantly the hierarchical pruning of the cells. An experiment is performed on a huge database including more than 5 billion 20-dimensional features.

Distortion-based probabilistic techniques were successfully applied to huge datasets including several billions signatures and have experimentally proved to be sub-linear in database size.

2.3 Graph model for layout and clustering module

The visual representation is based on graph models [8]. The modelling with graph structures has several advantages:

• First, the paradigm used for graph layout matches our objective in terms of representation: the proximity between entities and groups of entities in the 2D representation should reflect to the best their proximities in the N-dimensional descriptor space.

• The mathematical duality between graphs and matrices makes graphs efficient for expressing distance matrices.

• Graph filtering and layout have been already largely studied and drawing techniques exist, more or less adapted to our data [9].

• Passage from valued to symbolic may be easily done on the model. Links removal may promote topology emergence in the data set.

Two approaches are possible: Multi-Dimensional Scaling (MDS) approaches usually applied for complete graphs with valued edges, and topological placements for non-valued links. Our approach uses a customized energy force model algorithm [10] [11] to achieve the layout. Within this model, we consider a repulsive force between nodes, and a spring force between connected nodes. Each edge is seen as a spring characterized by its resting length and its stiffness coefficient. The resting length of each edge is linearly correlated to its distance attribute. For our applications, the goal of the visualisation is to identify neighbourhoods and aggregates. MDS is not very efficient in layout quality and computing time because we have to consider the complete graph. Then, we have implemented generic filtering methods of nodes and edges based on their inner or topological attributes (centrality, degree, hub/authority values…). For graphs based on similarity matrices, the highest similarities of each node appears to be the most significant to elaborate the layout. Then, we generate a graph issued from knn filtered matrices. In such cases, the radius of a cluster and the distance between two clusters are related to the inverse of the edge density (normalized edge-cut) [12]. Filtering functionalities may be adjusted interactively to make the view more readable.

Finally, we use a standard agglomerative hierarchical clustering algorithm to identify and label clusters in the rendering space. To obtain clusters of arbitrary shapes, we choose a linkage metrics based on the minimum distance between objects. These distances can be parameterized in the interface, allowing the user to *control* the view.

2.4 Active learning with feedback diagnostics module

Our active learning module involves user's feedback and is based on two steps: a classifier and a sampling strategy which selects pertinent samples to improve the classification according to user actions [13]. One constraint

for this module was to provide users with diagnosis about the system decision in order to guide their actions. Then, we have designed jointly the classifier and sampling strategies in the Transferable Belief Model framework (TBM). Indeed, the TBM offers us an accurate and complete formal modelling of the knowledge and characterises the samples by positiveness like a classical classifier, but also by distance rejection and ambiguity. Distance rejection can be seen as a measure of visual content diversity and allows handling interactively multi-prototype classification: users can create or aggregate classes interactively. Moreover, the TBM framework models doubt and ambiguity which allow expressing hierarchical classifications. Sampling strategies may be designed directly and easily from the output characteristics of the TBM classifier and their impact has been theoretically evaluated on two criteria: the error rates and the user cognitive load according to the distribution of effort over time.

3. APPLICATIONS

3.1 Short description of the interfaces

The GUI provides a fully operational interface for users in two main modes:

▪ A Friendly exploration mode: the map has been previously built and the interface is used to search, navigate or analyse a given set of data.

▪ An Expert mode to create visual maps: Expert user may access several setting tabs such as the data table, available media similarities, layout parameters, clusters identification or filtering operations, search actions… He may customize and experiment settings to design his own profile adapted to his task and the set of resources he is looking at.
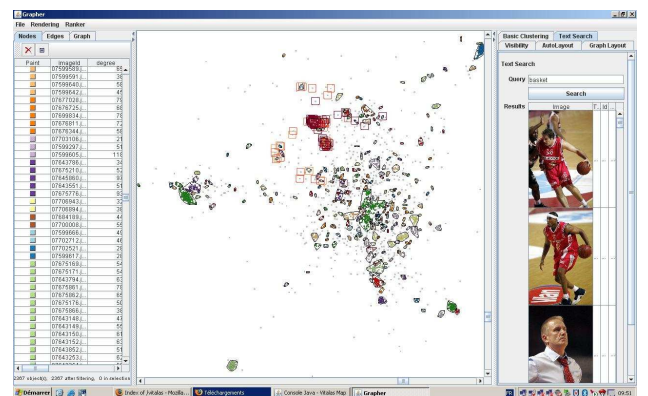

Fig 1: Cartography of results for the query "match" on a corpus of Belga press agency. The Map has been built on visual similarities computed by INRIA/IMEDIA. The left panel shows resources textual information, the central panel contains the graphic map; the right panel shows the current

search function on the word "basket". Results are marked by colored squares according to their rank.

The interface contains 3 main panels (see fig 1). The data table contains 3 types of data: graph data, node data or edge data. The table may display any pre-existing information or computed information for each entity. For a programme represented by a node, its filename, title, duration or keywords may be displayed, as well as computed data such as node degree (the set of programmes this programme is linked to), centrality…These tables are sortable and linked to the graphic view. The graphic panel contains the graph view. The function panel contains all the functionalities described above and as well as the matching of graphical attributes or search functionalities…

The active learning panel may be launched from the interface and proposes the user a list of the 10 images of highest value for the current sampling strategy (see fig 3 upper right corner). When running the semi-supervised decision process, the upper left image moves to its "destination class". The list of images to be classified is then updated by the system according to this decision and the current sampling strategy (most positive, ambiguous or rejected). The user can take the image to modify its destination, create a new class 'on the fly' or change the current strategy. The decision process may also be tuned and executed without supervision.

3.2 Analysis of TV news on a given period of time

TV news reports illustrate daily world events. They represent interesting items to archive because they are often used in productions or for research purposes in media, sociology or contemporary history. For these reasons, evening news documentary notices benefit from the highest level of description. TV reports are numerous: about 12 reports a day, 365 days a year. The Map shown in figure 2 has been computed from the corpus of evening news of the French public channel "France 2" between 1st July and 30th September 2000. Each item represented on the map is a single news report. Its size is correlated to its duration and its colour represents the category of event the report is associated with. INA has defined 16 "major categories" for news events such as *catastrophe, economy, education, environment, world...* Links between reports are built from textual and temporal proximities: if two reports deal with similar subjects and if they have been broadcasted in the same week, then they are linked. The graph layout is performed with the following additional rule: new subjects of the first and last day are constrained in x coordinate in order to report roughly the time axis on the x axis.
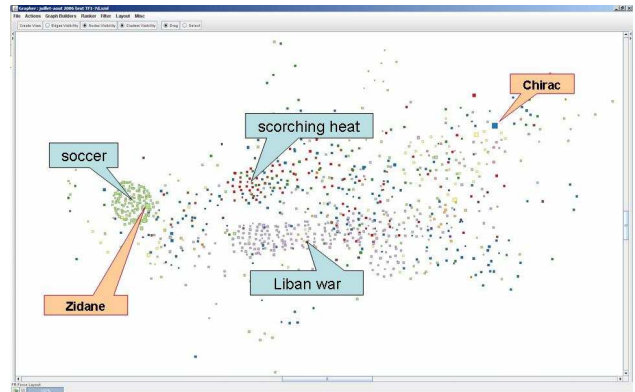

Fig2: map of news report from France 2 from 01-07-2000 to 30-09-2000. The spatial proximity of the reports matches their semantic proximity.

The map figure 2 allows users to easily analyse the main trends of the news during this period because the main events are visually emerging:

- The dense green circle on the left of the image represents reports about the world soccer cup.
- The elongated and more diffuse purple cloud gathers reports about the Lebanese war.
- The colored cluster is about the possibility of a heat wave. In 2003, France had a severe heat wave, which has killed many elderly people.
- Two items are really bigger that the others: a speech by the French president Jacques Chirac and an interview with Zidane explaining his headbutt during the soccer cup final.

3.3 News Visual Summary

Our motivation here is to produce visual summarized overviews of structured TV programmes. Currently, temporal media (radio and video) need time linear reading to be perceived. Before 1985, INA's archivists watched TV and annotated programmes "on the fly" in real time. Now, TV broadcast is stored and displayed on demand. Archivists are interested to get graphic views giving an "idea" of the content without watching it as a whole. The problem is difficult: audiovisual forms are usually complex to characterize and identify. However, TV broadcasting presents good properties and makes the challenge conceivable. In fact, part of TV broadcasting is very redundant; programmes form and channels' identity designs are usually very precise, frequent and relatively permanent. One of the first objectives of media content analysis is to locate temporal, visual or audio objects (such as settings, logos…). The following step is to cluster and label the objects manually or by propagation according to chosen proximity criteria. To test these hypotheses, we have

developed an application to create a "news summary Map" from any TV news resource.

Key frames are selected at a predefined constant frequency then computed to extract low-level features and create indexes. Matrices of similarities between key frames are computed from low-level indexes. These similarity matrices are transformed into distance matrices and finally exported in a complete graph representation: a node represents a key frame and its descriptors, and an edge between two frames symbolizes a link multi-valued by all the distances computed.

We will call *backbone* the set of frames making up the unifying thread of the video narration. In the case of TV news, the structure of the programme is very simple: the anchorman sequences constitute the backbone of the programme. Video incrustations may appear behind the anchorman leading to the generation of many dense, small and near clusters.

The module of active learning computes diagnosis on the selected key frames. Two classes are proposed to the user: the class with the highest time distribution and a rejection class. The initial class may be modified by the user or inherited from the collection level. In the specific case of TV news, the decision process classifies the frames without errors. However, the choice of the backbone may be more delicate for other programmes such as sport or variety programmes. In these cases, the user may create several classes composing the back bone.

Then, a new graph is built with two temporal proximity constraints on edges: a chronological sorting is applied for key frames of the backbone class, and the basic temporality is preserved for other frames.

Finally, a topological custom algorithm based on the graph layout module develops the time axis horizontally and spreads out thread loops on both sides of the backbone [11, 12]. The final visualisation emphasizes the video structure and gives a quick overview of its contents. The layout has to be read from the left to the right, in the chronological order as in figure 3.

A light evaluation has been done with 6 users to get the first feedback on the readability of the representation. 5 tasks of search or analysis have been tested with our representation compared to a classical grid of images extracted from the video at regular time. Tasks had to be performed in 3 minutes. Results shows that users achieve the tasks in less time with our interface and in general prefer our representation.
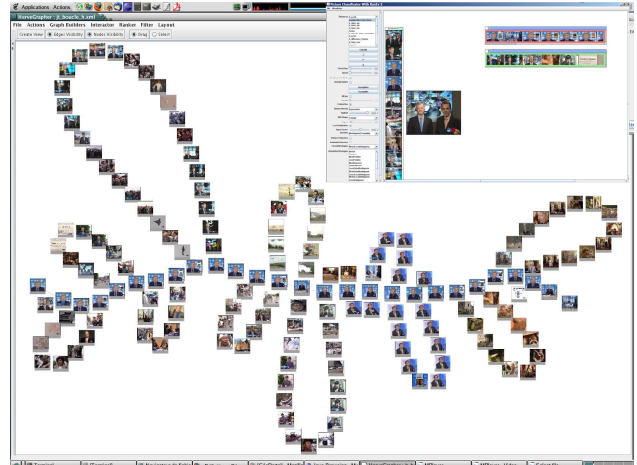

Fig 3: Visual Summary of France 2 TV evening news. In the upper right corner, a view of the decision interface to identify clusters.

3.4 Interactive Exploration from example

The objectives of this tool are to facilitate the annotation process of INA's picture bank and to develop and test new content search processes. In fact, digitization of INA's pictures began two years ago, and the annotation process is just starting.
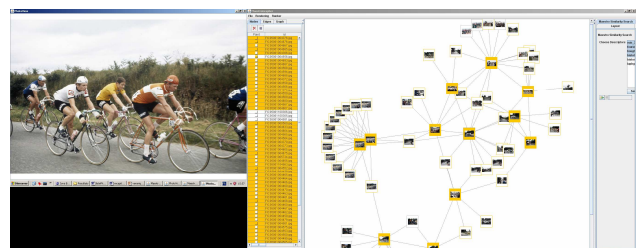

Fig 4: GraphoExp interface: exploration of resources based on resource similarity. On this example, links are built on visual similarities such as color, texture, orientation. (cycling: Tour de France 1972, 23/07/1972)

GraphoExp is an interactive visual module which allows the user to explore and develop similar entities from a given example. In this example, the corpus is a set of 170000 images corresponding to the set of digitized pictures of INA resources. The similarity criterion is based on global visual descriptors. Textual descriptors have not been used because only 5000 images are textually described annotations.

The interface contains the three panels described above. A fourth panel has been added on the left to show the image pointed at by the mouse. Different functionalities are proposed to the user. First he can develop the neighbourhood of an image with a double left click and select pertinent entities with a single left click or delete non pertinent ones with a single right click. Secondly, he can

navigate with classical functions such as translation, zoom on one image or a part of the graph, or move to a complete vision of the graph. This interface allows the user to progressively control its results. According to the goals, the number of images to find or given time constraints, different strategies may be employed. Some users will methodically develop and suppress nodes to get a "clean visualisation" and limit the graph size. On the contrary, others will rapidly develop a maximum number of nodes and select pertinent ones. The graph layout conveys information on the distribution of the images (see fig 6) and gives feedback on which image to explore: pictures belonging to dense aggregates will have fewer chances to develop new types of neighbourhoods. In a more advanced usage, users may change the visual similarity criteria to progressively specialize their demand according to the current results. For example, they could search images with similar texture and orientation and then switch to similar colour criterion if more pertinent (see fig 5). Users may active relevance feedback functionality based on TBM active learning.

Fig 5: The Shadoks, popular animation serie of the 70's, are well retrieved by color criteria. (Les Shadoks, 2nd serie 01/05/1970)

4. FIRST RESULTS OF EVALUATIONS AND NEXT STEPS

The objective of this evaluation was to analyse in real context the impact of GraphoExp, our navigation tool which has been described above without relevance feedback functionality, in comparison with a classical list interface. We have made a comparative ergonomic experimentation between two interfaces of image search by visual similarity criteria[1]: GraphoExp and Totem, the in-house annotation application currently used by archivists. As many evaluation criteria exist, we have chosen the usability criteria defined in the ISO 9241-11 norm completed with some criteria extracted from guidelines [14, 15, 16]. As we have the opportunity to make tests with professional media librarians, we chose to use empirical methods. The realisation of this user test evaluation relies

---

[1] For this experiment, the visual descriptors and similarities are generated by the system Maestro developed at INRIA-IMEDIA [ref]

on the following steps:

- Definition of the test hypothesis: 2D representation based on proximity offers a better navigation than list display for visual proximity search tasks. This kind of interface may be helpful for archivists.
- Creation of a scenario inspired by the tasks currently carried out by librarians: users had to find the largest set of visually similar images in less than 3 minutes. 11 searches had to be performed on both interfaces.
- Definition of the measures used for *usability* criteria (ISO 9241-11). *Effectiveness* is given by the number of good results and errors. *Efficiency* is expressed by the ratio of (number of selected images / number of displayed images) and the duration; and *satisfaction* is based on the positive and negative comments and answers.
- Specification of a list of heuristics (from existing ones) to collect feedback on learning, memorisation…
- Creation of groups of ten users with "equivalent" profiles + two people for a pre-test;
- Elaboration of the corpus: two sets of 11 images to search. The two groups should present similar difficulties.
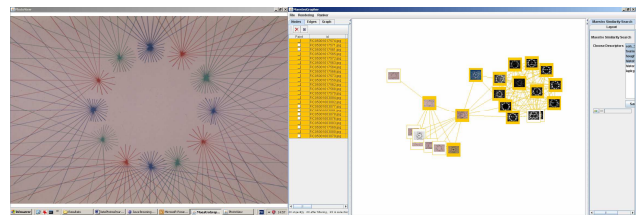

Fig 6: The dense aggregate on the upper right part of the map is composed of linked images (complete sub graph). Mires Eurovision TV 11/04/1968

A preliminary test was conducted to verify if the test was valid and if the context situation had no unexpected impact on the results. Then, each user completed the test with the following evaluation schema for both interfaces, alternating both test corpora:

1. presentation of the evaluation,
2. test completion alternating both test corpora with logs and expert observation.
3. questionnaire,
4. open interview.

The analysis of the results is based on statistical computation on logs and interpretation on observations and questionnaires. About the effectiveness: there are no clear differences between both interfaces for the number of correct results: it depends on the image characteristics and the user… On average, one group of users made more errors with the list interface (2,7 %) than with GraphoExp

(0,86 %). For the efficiency criteria: the ratio (number of images selected / number of displayed images) was clearly higher for GraphoExp for both groups and more requests were carried out in less than 3mn. Concerning satisfaction, some indicators show a preference for GraphoExp: the main advantage noticed for legibility criteria was that more information about the query state was given thanks to the presence of links: the similarity is higher in very linked groups. In general, users prefer to be active: they like to choose the image to expand and the very slow motion of the graph view which maintains attention.
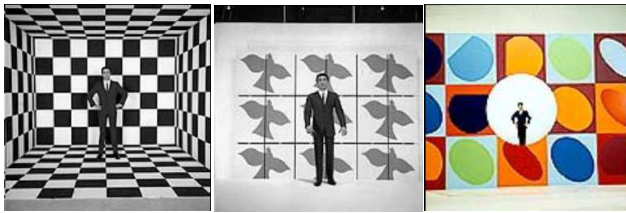

Fig 7: Example of images retrieved by visual similarity on the query "Becaud on a checkerboard". Show Becaud 01/01/1968, Director JC Averty.

Some additional remarks: Most of the users who used to play video games were definitely in favour of GraphoExp while the others were not so comfortable with mouse actions (zoom, translation, selection). We asked our pre-test archivists to perform the same search on our classical textual query system and without any time constraints. The results obtained show that visual search by similarity and textual search were complementary, even for content bases that have high quality text descriptions. In fact, for some queries such as the old "TV test patterns" or "Becaud on a checkerboard" (see figure 7), users with GraphoExp interface will find on average respectively 23 and 10 images and Totem 22 and 9 images, while the classical textual search gives only 15 and 9 results. On the other hand, on some other examples, visual similarity search appears too restrictive. Interviews show also that archivists were more able to find pictures or programmes they already know. GraphoExp appears for them a good way to explore the new resources. Finally, this evaluation also allows the elaboration of a list of propositions to improve GraphoExp interface for current use in the production process: image zoom, rejection folder…
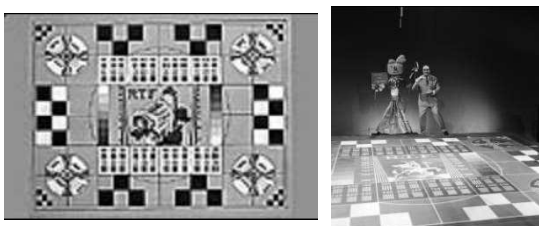

Fig 8: examples of images matching the "TV Test pattern" query. The image on the right has been retrieved by textual search. Mire RTF de diffusion avec Caméraman incrusté 24/03/1953.

5. CONCLUSION & PERSPECTIVES

To reach INA objectives in terms of evolutions, we have created links between different scientific fields (image description, index structuring, machine learning, information visualisation and IHM). These links enabled us to develop modules based on these different fields to achieve prototypes adapted to INA's needs (search and annotation help, analysis & monitoring). We have led evaluation on one search module (GraphoExp), with promising results.

We would like to integrate new or advanced components:

- new content features: visual local features, sound descriptors (MFCC and zero crossing rate),
- speech analysis module
- more generic index structuring modules
- new TBM active learning strategies

Finally we have to make evaluations on other prototypes in order to estimate the impact of cartographies for analysis and ground truth generation and the use of relevance feedback functionality for the search.

6. REFERENCES

[1] J. Law-To, O. Buisson, V. Gouet-Brunet and N. Boujemaa, Robust Voting Algorithm Based on Labels of Behavior for Video Copy Detection, in Proc. of the ACM Multimedia, 2006.

[2] N. Boujemaa and J. Fauqueur and M. Ferecatu and F. Fleuret and V. Gouet and B. Le Saux and H. Sahbi", IKONA for interactive specific and generic image retrieval", MMCBIR 2001, France", 2001.

[3] N. Grira, M. Crucianu, and N. Boujemaa. Fuzzy clustering with pairwise constraints for knowledge-driven image categorization. In IEEE Proceedings on Vision, Image and Signal Processing, 2006.

[4] M. Larson, S. Eickeler, J. Köhler. Supporting radio archive workflows with vocabulary independent spoken keyword search. Searching Spontaneous Conversational Speech Workshop, SIGIR 2007, Amsterdam

[5] A. Joly, C. Frélicot and O. Buisson, Feature statistical retrieval applied to content-based copy identification, in Proc. of Int. Conf. on Image Processing, 2004.

[6] A. Joly, O. Buisson and C. Frélicot, Content-Based Copy Retrieval Using Distorsion-based Probabilistic Similarity Search, in IEEE Transactions on Multimedia, 2007.

[7] S. Poullot, O. Buisson and M. Crucianu, Zgrid-based Probabilistic Retrieval for Scaling Up Content-Based Copy Detection, in Proc. of CIVR, 2007.

[8] Jérôme Thièvre PHD Thesis, INA/Univ of Montpellier, 2006

[9] I. Herman, G. Melançon, and M. S. Marshall. Graph visualization and navigation in information visualization: A survey. IEEE Transactions on Visualization and Computer Graphics, 6(1):24-43, 2000.

[10] A. Noack. An energy model for visual graph clustering. In G. Liotta, editor, 11th International Symposium on Graph Drawing (GD 2003), LNCS 2912, pages 425{436, Berlin, 2004. Springer-Verlag.

[11] T. M. J. Fruchterman and E. M. Rheingold, "Graph drawing by force directed placement," in Software - Practice and Experience, 1991, vol. 21, pp. 1129–1164.

[12] A. Noack. Energy-based clustering of graphs with non uniform degrees. In 14th International Symposium on Graph

[13] H. Goëau and O. Buisson and M-L. Viaud}, Image Collection Structuring Based On Evidential Active Learner, Sixth International Workshop on Content-Based Multimedia Indexing, London, UK, 2008

[14] H. Goëau and J. Thièvre and M-L. Viaud and D. Pellerin, Interactive Visualization Tool with Graphic Table of Video Contents, EEE International Conference on Multimedia and Expo, p 807-810, Beijing, China, 2007

[15] Nielsen J. Usability Engineering. Academic Press, Boston, 1993.

[16] Shackel, B. (1991). Usability – context, framework, design and evaluation. In Shackel, B. and Richardson, S. (eds.). Human Factors for Informatics Usability. Cambridge University Press, Cambridge, 21-38.

[17] Shneiderman, B. (1992). Designing the User Interface: Strategies for Effective Human-Computer Interaction. 2nd edition. Reading, MA: Addison Wesley.

Drawing (GD 2006), pages 309-320, Limerick, Ireland, 2006.