# DAG-BASED VISUAL INTERFACES FOR NAVIGATION IN INDEXED VIDEO CONTENT[1]

*Maylis DELEST, Anthony DON, Jenny BENOIS-PINEAU*

LaBRI CNRS UMR 5800/Université Bordeaux 1
{jenny.benois,maylis,don}@labri.fr

**ABSTRACT**

Indexing and segmenting of video content by motion, color and texture has been intensively explored leading to a usual representation in a storyboard. In this paper, a novel method of visualization of video content is proposed.

First of all, the content is segmented into shots, and then a spatio-temporal color signature of shots, based on color distribution in the frames, is proposed. This spatio-temporal color signature serves as a basis for graph clustering and graph visualization tools. The latter, united in a general huge graphs visualization tool, Tulip, supply an exciting graph-based navigation interface for multimedia content. The results obtained on feature documentaries are promising.

**Keywords:** Video indexing, color descriptor, graph clustering, navigation.

## 1. INTRODUCTION

Structuring video documents into logical video segments such as shots, scenes, objects and indexing them by characterizing their motion, color or texture is a key feature of the MPEG7 [10] standard. Numerous descriptors can be used and several techniques have already been proposed to describe video content by means of low-level features. Then, queries can be realized on indexed video in order to re-build or explore it.

For this purpose, storyboards or linear navigation interfaces [5] are widely used to visualize the content of video documents, but they do not show the global structure of a video document nor allow an easy identification of similar shots that are distant in time. Based on structure of the document, the navigation interface has to ensure intuitive and attractive tools for browsing through video. It is natural to expect such tools will give to the user capability to access specific segments of video content and also show the structural dependencies of these segments (shots) across the document.

"Image maps" [13] are an interesting approach for navigation and retrieval applications in databases of still images. It consists of a 3D browsing environment that allows users to navigate in clusters of color-indexed images that are organized in a hierarchical way. Between each level of clusters, user satisfaction is taken into account to refine subsequent clusters' construction. Although using video data and an underlying graph-structure, our method is based on the latter concepts and provides a 3D view of video-shots clusters at different levels of granularity.

After video segments have been characterized by low-level descriptors, an aggregation of them into higher order structural content units, often named "scenes" becomes possible. In this paper, we limit ourselves to a

[1] This work was supported by the Conseil Régional d'Aquitaine and the STIC Department of the "Centre National de la Recherche Scientifique".

particular descriptor, namely spatio-temporal color signature (SCS) firstly introduced in [6] and focus on the opportunity of graph-based visual interfaces for clustering shots into scenes.

Graph-based interfaces represent today a very active research field. The main reason is surely in the fact that from all data structures one can easily extract a graph representation. It has been proved that our visual perception is very efficient and that when algorithm cannot connect objects, the eye can if a good representation is given to the user. Our approach is clearly in the field of information visualization and more precisely in the Schneiderman's paradigm [16]: "Overview first, zoom and filter, then details on demand". First attempts to use graph theory formalisms for video structuring and navigation were done in pioneer work [19]. Nevertheless, due to their planar structure the proposed story graphs have not found their route to users. In this paper we try to go further and propose to use huge graph navigation techniques for structuring and navigation in video documents.

In a video document, where each shot is characterized by a vector or scalar descriptor, a distance between shot descriptors can be computed. Based on SCS-distance matrix for shots, we consider a video document as a weighted complete graph for which the distance matrix is in fact an incidence matrix. The vertices of the graph are the shots and the edges between two vertices are weighted by the corresponding entry in the SCS-distance matrix.

The main problem is now to find a method for selecting a subset of relevant distances (edges) for grouping shots into scenes. This is done by statistical methods. Then, Strahler numbers [4] on graphs can be used to build clusters of similar shots according to Strahler values of vertices.

We also experiment a 3D navigation in the clustered graph, using Tulip [3], a software tool for visualization of huge graphs. The first results given here show that this kind of exploration for video can be a nice alternative to storyboards.

This paper is organized as follows. Section 2 describes the spatial similarity measure between shots based on discrete projection transform and vector quantization. In section 3, we present Tulip, a software tool that allows visualization and manipulation of huge graphs. Section 4 explains how a distance-matrix can be processed to obtain a weighted oriented graph. Sections 5 and 6 describe application of structural graph clustering methods to grouping shots in a video document. Results are discussed and perspectives of this work are outlined in Section 7.

## 2. VIDEO STRUCTURE INDEXING

### 2.1. Video document model

A video document can be decomposed into a hierarchy of scenes and shots. A shot represents a continuous camera work. Shots with same semantics of their content could be grouped into scenes. We will not discuss here the complex problem of scene definition, which strongly depends on the type of video document. We limit ourselves to the assumption, that the visual semantics could be sufficiently well represented by low-level descriptors. Thus, a specific distribution of color in the image plan surely allows the separation of interior and exterior views, a properly measured intensity of motion activity allows distinguishing between strong action and nearly static narrative content and so on.

In the rest of this paper we suppose that the initial segmentation of a video document in the set of consecutive shots has been done. For this purpose we can use one of the methods we previously developed [6].

### 2.2. Scene grouping

Scene grouping consists of re-building of semantics of a video document from low-level temporal segmentation and description: shots and their descriptors. The main idea here is to consider that shots with similar descriptor values belong to the same scene. Color and texture descriptors such as those described in MPEG7, Part 3, could be candidate descriptors. Therefore, a distance measure between shots based on descriptor-based

distance must be set-up. In this paper, we use the results of our previous work defining the spatio-temporal color signature of shots [6] and use it as shots' descriptor.

### 2.3. Spatio-temporal signature of shots and related similarity measure

Spatio-temporal color signature of a video shot is a specific descriptor, which enables measuring similarity between shots. The signature we proposed in [6] is based on a 1D discrete projective transform called "X-Ray image". The X-Ray image of a video frame was first introduced in Tonomura's VideoMap [18]. It is done by "projecting" an image $(I(x,y))_{1\le x\le W,\,1\le y\le H}$ along straight lines either vertically or horizontally. This projection consists in integration of image signal along these straight lines and in this sense, represents a discrete analogue of Radon transform used in tomography.

Each projection is a vector with each element, or bin, being the mean of pixel values along the given direction. The value of an X-Ray vector element is:

$$I_x[I](y) = \frac{1}{W}\sum_{k=1}^{k=\mathrm{W}} I(k,y) \tag{1}$$

for a horizontal projection, or:

$$I_y[I](x) = \frac{1}{H}\sum_{k=1}^{k=\mathrm{H}} I(x,k) \tag{2}$$

for a vertical projection.

The spatial distribution of the image intensity in the direction of the axis orthogonal to the projection vector is preserved. This is an interesting property of this transform compared to usual histograms. Applying (1) and (2) to a color image, we will form a set of 3-component color bins $\vec{I}_x[I](y)$, respectively $\vec{I}_y[I](x)$, which represents a reduced (to a column, respectively, to a line) version of a complete color video frame $\vec{I}(x,y)$.

As we mentioned before, the color is a salient feature to express video semantics. This is why several methods have been already developed for grouping shots into scenes by color similarity [1, 14, 19]. These methods go from histogram-based distances to more complex models.

In our case, all the frames of a shot are transformed into 1D signal using one direction of projection, then the global color descriptor of this set of projections have to be defined to characterize the shot.

A coding–based approach was proposed in [1] to measure a similarity between shots based on vector quantization. Here, each frame is segmented into fixed-size blocks. Then, for a given shot, all blocks of all its frames are considered as vectors in the $N^2x3$ space. Here, $N$ is a linear size of a block. A codebook for each shot, which optimally encodes the blocks by a lowest number of vectors, can be built by VQ (Vector Quantization) technique. In [6], we showed that the same approach could be applied to the X-Ray vectors of each frame of a shot. Here, the projection $\vec{I}_x[I]$ (1) in color space can be segmented into vectors of size $Nx3$. Then, a codebook (we call it dictionary $Dico_i$) can be built for the X-Ray projection of a whole video shot. *We call a codebook $Dico_i$ of a shot $S_i$ its spatio-temporal signature.*

One of the most popular methods for building of a codebook with VQ is the Split-LBG approach [9], which is a generalized K-means clustering [15]. In particular, it was used in [1]. We also applied the same method for building our spatio-temporal signatures.

In case if the resolution of video frames is reduced, for instance only DC frames from MPEG-encoded [11] video, which are 8 times smaller than the full-resolution frames, are used, then we can consider a degenerated vector of dimension 3 restricted to only one color bin $\vec{I}_x[I](m), m = 0,..,H$. The signature thus defined is in fact spatio-temporal in the sense that the X-Ray projection preserves spatial color-distribution in a video shot globally. To illustrate this, X-Ray projections $\vec{I}_x[I]$ of 8 times reduced video frames (DC resolution), cumulated along the time in one single image, are shown in Figure 1. It can be seen that on one

hand the cumulated X-Rays preserve homogeneity inside a shot (see figure 1.a). On the other hand, it can very well express camera motion [12] inside a shot while keeping the global color homogeneity. Thus, in figure 1.b, the slope of lines changes with changes in camera work.

a)

b)

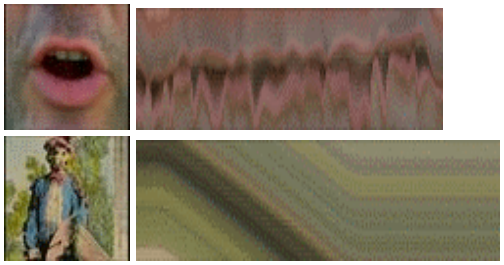

Figure 1. Key-frames and X-Ray images of shots a) 18 and b) 19 from "La joueuse de Tympanon", SFRS®.

As the spatio-temporal color signature of shots is defined, we will now introduce similarity measure based on it.

Let $\vec{I}_x[I]$ be an X-Ray projection of a color frame $\vec{I}(x,y)$. Let $\vec{I}_x[I]\,(m)=(u_1,u_2,u_3)^T$ be its value at bin *m.* The bin is thus a vector with color coordinates $u_1$, $u_2$ and $u_3$ in a chosen color system. Let $\vec{V}=(v_1,v_2,v_3)^T$ be some vector in the same color system. Then,

$$d(\vec{I}_x[I]\,(m),\vec{V}) = \sum_{k=1}^{k=3}(v_k-u_k)^2$$

is the distance between a bin of an X-Ray projection and a color-vector.

Let $S_i$ be a shot considered as a set of bins, constituting all the X-Ray projections, then $Dico_i = \{\vec{V}_{i,j}\}$, $j=1,..,J,$ is the code-book associated to shot $S_i$ with $J$ elements, in $Nx3$ space, computed with Split-LBG.

Let $S_i$ be a shot, $Dico_i$ be its associated codebook and $\vec{I}_x[I_{i,p}]$ be the X-Ray projection of the $p^{th}$ frame of $S_i$. A vector $\vec{V}_{i,j_m}$ "encodes" the bin $\vec{I}_x[I_{i,p}]\,(m)$ if $d(\vec{I}_x[I_{i,p}]\,(m),\ \vec{V}_{i,j_m})$ is minimal for all $\vec{V}_{i,j}$ from $Dico_i$.

Let $S_i$ be a shot, $Dico_i$ be its associated codebook, $\vec{I}_x[I_{i,p}]\,(m)$ be the X-Ray projection of $p^{th}$ frame of $S_i$ at bin $m$ and $\vec{V}_{i,j_m}$ be the element of $Dico_i$ that encodes $\vec{I}_x[I_{i,p}]\,(m)$. Then,

$$D_i(S_i)=\sum_{p=1}^{P_i}\sum_{m=1}^{M} d(\vec{I}_x[I_{i,p}](m),\vec{V}_{i,j_m})$$

is the "proper" coding distortion when one uses a codebook designed for the shot $S_i$ to represent a X-Ray projection of this shot. Here, $M$ is the total number of bins in the projection, which corresponds to the vertical size of video frames. $P_i$ is the total number of video frames in the shot $S_i$. Let $S_j$ be a shot and $Dico_i$ be a codebook for the shot $S_i$. Then, $D_i(S_j)$ is the "cross" coding distortion when one uses a codebook designed for the shot $S_i$ to "encode" a shot $S_j$.

Based on coding distortion, a similarity measure $\Delta(S_i,S_j)$ between shots $S_i$ and $S_j$ can be computed as proposed in [1] :

$$\Delta(S_i,S_j)=|D_j(S_i)-D_i(S_i)|+|D_i(S_j)-D_j(S_j)| \qquad (3)$$

It can be easily checked that $\Delta$ always satisfies two of three axioms of a distance. Thus, from (3) it follows

i. $\Delta(S_i, S_i) = 0$

ii. $\Delta(S_i, S_j) = \Delta(S_j, S_i)$.

As for the triangle inequality, it is satisfied in a trivial case when a proper coding distortion for a triplet of shots $S_i$, $S_j$, $S_k$ satisfies: $D_i(S_i) = D_j(S_j) = D_k(S_k)$.The "cross" coding distortion is also the same when $D_i(S_j) = D_j(S_i) = D_k(S_i) = D_i(S_k) = D_j(S_k) = D_k(S_j)$ and for any pair of shots, the proper coding distortion is less than the

"cross" coding distortion. In general case, the validity of triangle inequality for $\Delta$ is not guaranteed. Thus, $\Delta$ is a pseudo-distance.

### 2.4. Results

We computed the spatio-temporal signatures described above for three video documents in CIF resolution at PAL/Secam rate (25 fps) MPEG1 coded and decoded. They are described in table 1.

| Sequence Name | Size in frames | Size in shots |
|---|---|---|
| « La joueuse de Tympanon », SFRS® | 40501 | 200 |
| « Aquaculture en méditerranée», SFRS® | 9200 | 41 |
| « Le temps des lagunes », SFRS® | 19400 | 87 |

Table 1. Indexed video documents.

The color space used was YUV. In order to reduce computational complexity, we used DC images for each color component giving a resolution of 44x30 pixels, each pixel representing the mean value of an 8x8 pixels block in the original frame. One X-Ray image $\vec{I}_x[I]$ was computed on each frame. Each bin of $\vec{I}_x[I]$ corresponded to the mean of pixel values along the line associated to the bin (1). Codebooks for each shot were computed using Split-LBG, with a total of 16 elements per codebook. Computing $\Delta$ for all pairs of shots, we will get a "distance matrix"

$$(\Delta(S_i,S_j))_{1\le i\le n-1, i+1\le j\le n}$$

where $n$ is the number of shots in the video sequence.

It is clear that such a matrix can be considered as the incidence matrix of a weighted complete graph. In the following, we describe methods in order to cluster and navigate in this graph.

## 3. NAVIGATING IN HUGE GRAPHS

The main problem in handling huge graphs is the number of elements (edges and vertices) but also the management of the whole data structure (cost of time and memory). Some operations such as displaying labels, displaying clusters or managing colors can lead to an unusable solution.

In this work, we use the Tulip software[2][3]. The main features relevant in this software are:

- a powerful kernel due to a minimization of used memory in the implementation of graph data structures,
- extensibility by plug-ins without recompiling,
- possibility of mapping texture and colors on edges and nodes without loosing "performance",
- easy management of clusters.

The figure 2 shows a shot of the main screen of Tulip.

A direct effect of application of the visualization techniques to the complete graph representing a video document is the possibility to "look-and-feel" the general esthetics of a movie. Thus, associating the most representative color of X-Ray projections of frames in a shot to the corresponding node in the complete graph and proposing a simple linear color interpolation for coloring of edges based on the distance between nodes, we can get an idea of the whole color of the graph. Figure 3 shows the complete graph for the documentary "La joueuse de Tympanon", SFRS®, mapped on a disk. Here the shot-nodes are situated on the borders and edges

[2] Download at www.labri.fr/~auber/projects/tulip

are inside. It can be felt that the dominant color of the indexed movie is brown with some light spots of white and blue.

If the grouping of shots is to be performed with graph clustering methods, then minimal weighted spanning trees can be used. A fruitful strategy here would consist in pre-filtering the graph in order to suppress useless edges. Furthermore, the edges can be oriented in order to get more information on the distance-based relation of nodes.

One of our main reasons for doing this is that when a graph is oriented, several algorithms, implemented in Tulip, are proved to be efficient for clustering or drawing.

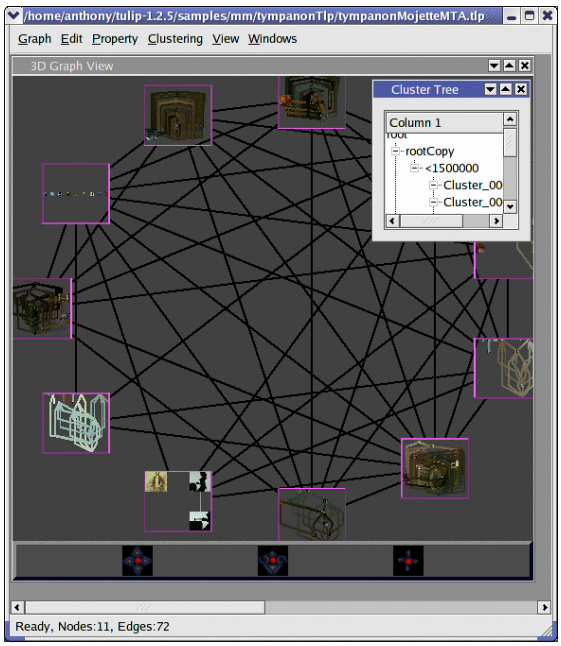

Figure 2. A view of Tulip software.

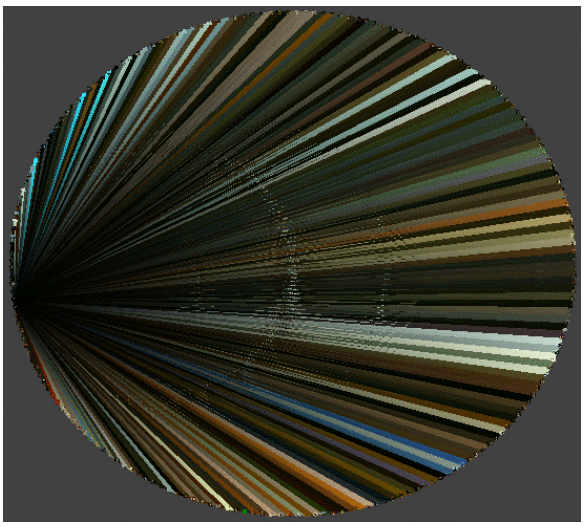
Figure 3. Complete graph of “La joueuse de Tympanon”.

## 4. FROM A DISTANCE MATRIX TO A WEIGHTED ORIENTED GRAPH

In this section we will describe the filtering of the weighted complete graph $G=<V,E>$. It consists in suppressing "weak" edges. These edges have a strong weight $\Delta$, which means that the associated shot-nodes are dissimilar. Thus, we have to explore the distance function $\Delta$ defined on $E$.

First of all, we compute a histogram of the values of the edges using a step of $s=50\ 000$, that has been experimentally defined analyzing typical range of $\Delta$-function on the used corpus of video documents. Thus, we will have histogram intervals $l=0,..,L$. Let $k$ be the total number of edges of $G$. Let $h_s(x)$ be the number of edges $e$ such that

$$s*l < \Delta(e) \leq s*(l+1).$$

We define

$$\eta_s(l) = \frac{h_s(l)}{k}$$

that is the distribution function and

$$\xi_s(l) = \sum_{j=0}^{j=l} \eta_s(j)$$

that is the characteristic function of the distribution. We have displayed in the figure 4, the $\xi$ functions for the two documentaries from our corpus. One can see that the two distributions ("Tympanon" and "Aquaculture") look similar.

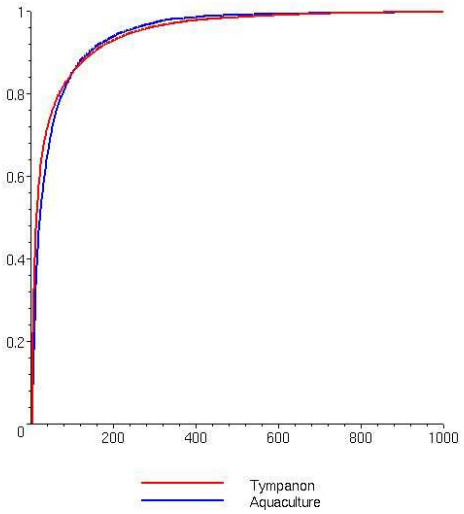


Figure 4. Characteristic functions.

In the figure 5, it can be seen that the distribution themselves reach a maximum. Let $\mu_s$ such that $\eta_s(\mu_s)$ is maximum. For a given weighted complete graph, consider the set of edges $e$ such that

$$\Delta(e) > \mu_s * s\ .$$

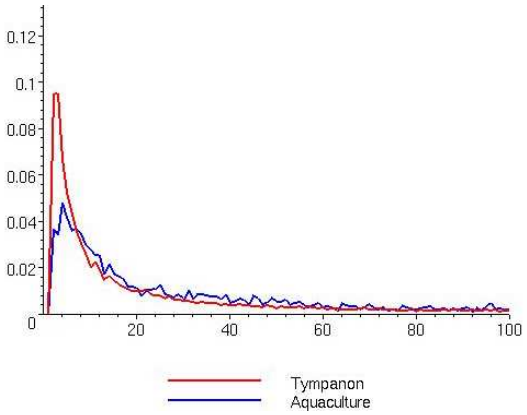

Figure 5. Distributions functions.

Each edge of this set has a high valuation thus the shots associated through these edges are not similar and moreover it does not contribute to the shape of the characteristic function. Thus these edges are deleted in the initial graph *G*. Let *G'* be the resulting graph.

Using the natural order of the vertices (rank of the shots in the movie), one gets a directed acyclic graph (for short, a DAG). Let *(i,j)* be an edge, then if $i<j$, *(i,j)* is an edge of the DAG and *(j,i)* is not. This order outlines the time sequences. In a certain way, it is a view of the original graph from the vertex 1. Note that the vertex *n* is a well, where *n* is the number of vertices (shots) of the graph.

In what follows, we construct a family of DAGs, $\{D_s(i), i\in[1,n]\}$. Let $D_s(i)=(V,E_i)$ then $D_s(i+1)=(V,E_{i+1})$ with

$$E_{i+1} = \{(j,k)\in E_i : j \neq i\} \cup \{(k,i) : (i,k)\in E_i\}.$$

An example is displayed in figure 6. Here, the initial non-oriented graph is shown at the top, the family of $n=4$ graphs is shown below. Note that $D_s(n+1)=D_s(1)$.

In fact, each $D_s(i)$ has its shot-node $i$ being the root of a DAG and its shot-node $i-1$ being a well. This is done to allow the comparison of the Strahler parameter, presented in next section, on the root of each DAG, i.e. at each shot.

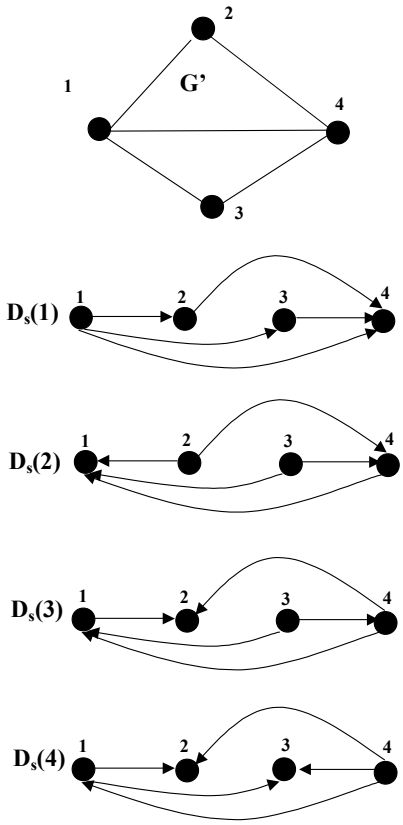


Figure 6. A set of DAGs built from the non-oriented graph *G'*.

## 5. APPLICATION OF STRAHLER CHARACTERISATION TO FILTERED GRAPHS OF SHOTS

In this paper, we have only studied the effect of the Strahler parameter [2] that consists in weighting vertices guided by the DAG. This weight takes into account the ramifications of the DAG. The nodes that correspond to similar shots according to $\Delta$ induce similar DAGs. The Strahler values at the root of the DAG associated to such nodes will be similar because they reflect the ramifications of the DAG. Therefore, we use a clustering method that puts nodes of the filtered shot graph with similar Strahler values in the same cluster.

The Strahler number $\sigma_b$ has first been introduced for binary trees in a work about the morphological structure of rivers [17]. It consists of associating an integer value to each node of a binary tree. These values give quantitative information about the complexity of each sub-tree of the original tree. The algorithm is a recursive one. Let $s$ be a vertex of a tree then

- If $s$ is a leave then $\sigma_b(s)=1$,
- Else $s$ has two sub-trees of root $s_1$ and $s_2$ and

$$\sigma_b(s) = \begin{cases} \max(\sigma_b(s_1), \sigma_b(s_2)) \text{ if } \sigma_b(s_1) \neq \sigma_b(s_2) \\ \sigma_b(s_1) + 1 \text{ otherwise} \end{cases}$$

A generalization to general trees can be made using the nice interpretation by Ershov [8] who has proved that the Strahler number of the root of the binary tree incremented by one is exactly the minimal number of registers needed to compute an arithmetical expression given by the tree output by the syntactical analysis. Thus, one can define the Strahler numbers $\sigma_T$ on trees [2] by:

- if $s$ is a leaf then $\sigma_T(s)=1$ ,

- if $s$ has $k+1$ subtrees rooted $s_0, s_1, \ldots, s_k$ and ranked in decreasing order by their Strahler numbers then

$$\sigma_T(s) = \underset{0 \le i \le k}{Max}(\sigma_T(s_i) + i)$$

The extension to DAGs is straightforward. Let $D=<V,E>$ be a DAG. Each DAG has at least one root that is a vertex with no entry edge. Each DAG has at least one well. We define the Strahler numbers $\sigma_M$ on DAG by replacing in the previous definition "leaf" by "well" and "subtrees" by "subDAGs". An example is shown in figure 7. Here, Strahler numbers are displayed near nodes.

The generalization on arbitrary graphs is defined in [2] and was used in [7], but is not helpful here.

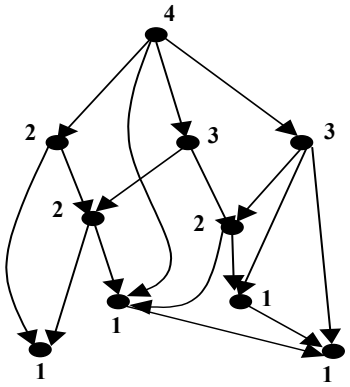


Figure 7. An example of the Strahler numbers on DAG.

In our application, we will consider that to each vertex $i$ in $G'$, we associate its Strahler number in the DAG $D_s(i)$ denoted by $\sigma_M(i)$. Thus, we obtain a graph whose vertexes are weighted. In the example of figure 6, we get

$$\sigma_M(1)=3,\ \sigma_M(2)=2,\ \sigma_M(3)=2,\ \sigma_M(4)=3.$$

Roughly, if one considers the graph G' associated to the $D_s(i)$, the vertices 1 and 4 have the same situation in regard of the vertices 2 and 3. They have the same Strahler value.

## 6. APPLICATION OF GRAPH-CLUSTERING METHODS.

After pre-filtering of the complete graph and computing Strahler numbers we have used a clustering method [4] based on convolution of Strahler metrics on the graphs. Then, clustering nodes with associated similar Strahler metrics allows us to take into account the induced structure of the graph. The clustering method allows graph splitting based on a distance threshold or on a target number of clusters. Note that this method could be applied recursively to the resulting sub-graphs. We refer the reader to [4] for the complete description of the method.

Clustering is fulfilled interactively, based on user-satisfaction of resulting clusters. It is made possible by Tulip, which allows an easy manipulation of graphs by providing numerous graph-drawing algorithms, an easy manipulation of clusters and a 3D view of graphs with zoom, rotate and translate functionalities. Clusters are displayed as a quotient graph, i.e. each cluster is a node in the graph of clusters. Users can assess the quality of clusters by viewing the key-frame that is mapped on each shot-node.

Using this process, we decided to split "La joueuse de Tympanon" into 11 clusters and "Aquaculture" into 5 clusters. Figure 9 gives a view of the clusters of the feature documentary "La joueuse de Tympanon" and figure 13 shows the clusters of "Aquaculture" in which one can navigate, using the mouse, seeing the image abstracting a shot.

Figures 10 and 11 show the navigation in a cluster. Figure 12 and 15 illustrate the performance of the clustering method and the underlined Strahler parameter: the majority of shots with similar color-content and

even scene layout were grouped together in one cluster. This can be explained both by the color signature, which reflects not only a color but also a texture in the projection domain, by the nature of the initial distance (similarity of projected frames using a coding distortion approach) and by the structuring capacity of the approach with regard to the initial distance matrix. In figure 11, several edges between thumbnail key-frames displayed show strong links between corresponding nodes, thus reflecting the similarity of shots even if they are not consequent in the movie (shots numbers are displayed in the white boxes in this figure).

Tulip draws automatically the sequences close one to each other because the underlying guide of the layout is a DAG. Giving a 3D overview of the whole DAG, Tulip also allows a planar zoom in clusters. Displaying the shot-nodes in such a way, with key-frames mapped on nodes, allows a fast recognition of clusters by human perception. Moreover, the fact that connectivity between nodes in a cluster is displayed reinforces the identification of clusters. Therefore, this kind of display is well adapted to the task of assessing the quality of clusters.

## 7. RESULTS AND FURTHER PERSPECTIVES

This paper describes our second attempt to apply huge graph visualisation methods to structuring and navigation in video content. Despite the fact that only one descriptor has been used in our experiments, and only one similarity measure based on coding distortion has been studied, the results obtained on real-world movies "La joueuse de Tympanon" (see figures 9, 10, 11 and 12) and "Aquaculture" (see figures 13, 14 and 15) are promising and also better than in [7]. In figures 11 and 13, the quotient graph of clusters is displayed. In figures 10 and 14, a zoom on two clusters is shown. A close view of clusters is displayed in figures 11, 12 and 15. The user gets an intuitive tool for navigating inside a video document and the clusters are really relevant.

We have also tested our method on another movie "Le temps des lagunes". This last movie supplies a good example of what our method cannot handle. Let us denote by $\delta$ the discrete derivative of the distribution function $\eta_s$:

$$\delta(l) = \eta_s(l+1) - \eta_s(l) .$$

The function $\delta$ is displayed for "Aquaculture", "La joueuse de Tympanon" and "Le temps des lagunes" in figure 8. It can be seen that this function is oscillating more for "Le temps des lagunes" than for the other two movies. In other words, the area of the hull of δ for "Le temps des lagunes" is larger than those for the two other movies. Thus choosing a maximum in order to suppress edges is not relevant for this kind of distribution.

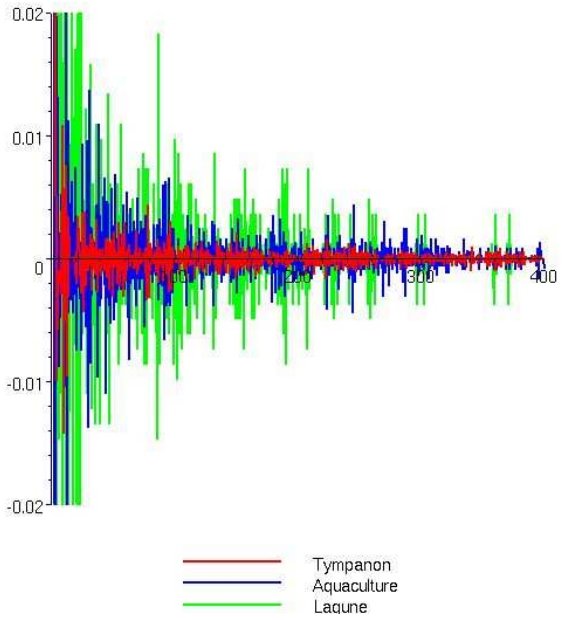

Figure 8. Function Δ for the three movies.

In the future, we plan to explore the use of these methods and the underlying Tulip software tool with a richer set of low-level descriptors of video content, such as motion and object-based descriptors. In this case different metrics should be studied and experiments on user satisfaction have to be conducted.

The choice of a representative key-frame for a shot or for a whole cluster is now also being studied.

Another open question remains on parameterisation of graph-clustering algorithms in supervised and unsupervised manner when applied to different types of video content. This is the subject of our current work.

## 8. REFERENCES


[1] N. Adami, A. Bugatti, R. Leonardi, P. Migliorati, L. Rossi, "Describing multimedia documents in natural and semantic-driven ordered hierarchies", Proc. ICCASP'2000, Istambul, Turkey, June 2000, IV:2023-2026.

[2] D. Auber, "Using Strahler numbers for real time visual exploration of huge graphs", International Conference on Computer Vision and Graphics, Zakopane, Poland, 25-28 September 2002, pp. 56-69.

[3] D. Auber, Tulip a huge graphs visualization framework, Graph Drawing Software, P. Mutzel, M. Jünger Eds., Springer Mathematics and Visualization series.

[4] D. Auber, Y. Chiricota, M. Delest, "Graph and software clustering from valuations", LaBRI 2003, submitted to IEEE Int. Conf. on Software Engineering and formal methods'2003.

[5] M. Barbieri, G. Mekenkamp, M.P. Ceccarelli, J. Nesvadba, "The color browser: a content driven linear video browsing tool", ICME 2001 (Int. Conf. on Multimedia and Expo), Tokyo, Japan, August 22-25 2001.

[6] J. Benois-Pineau, W. Dupuy, D. Barba, "Re-covering of visual scenarios in movies by motion analysis and grouping spatio-temporal colors signatures of video shots", Eusflat 2001 (Internationnal Conference on Fuzzy Logic and Technology), Leicester, Septembre, 2001.

[7] M. Delest, A. Don, J. Benois-Pineau, "Graph-based visual interfaces for navigation in indexed video content", Proc. CBMI'03, Rennes, France, 2003, pp. 49-55.

[8] A. P. Ershov, "On programming of arithmetic operations", Com. of the A.C.M, 1958, vol. 8, pp. 3-6.

[9] R.M. Gray, "Vector quantization", IEEE ASSP Magazine, 1984, Avril, pp. 4-28.

[10] ISO/IEC JTC 1/SC 29/WG 11/M6156, MPEG-7 Multimedia Description Schemes WD (Version 3.1), Beijing, July 2000.

[11] ISO/IEC 13818, MPEG-2, Information technology. Generic coding of moving pictures and associated audio, 1995.

[12] P. Joly, H.-K. Kim, "Efficient automatic analysis of camera work and micro-segmentation of video using spatio-temporal images", Signal Processing : Image communication, 1996, vol. 8, pp. 295-307.

[13] T. Meiers, T. Sikora, I. Keller, "3D browsing environment for MPEG-7 image databases", Proc. SPIE Storage and retrieval for media databases 2002, vol. 4676, pp. 324-335.

[14] J.-M. Odobez, D. Gatica-Perez, M. Guillemot, "Video shot clustering using spectral methods", Proc. CBMI'03, Rennes, France, 2003, pp. 94-102.

[15] J. Mac Queen, "Some methods for classifications and analysis of multivariate observations", Proc. 5th Berkeley Symp. Math. Stat. Prob., 1965/1966, pp. 281-297.

[16] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualization", IEEE Conf. on visual languages, Boulder, 1996, pp. 336-343.

[17] A. N. Strahler, "Hypsomic analysis of erosional topography," Bulletin Geological Society of America 1952, vol. 63, pp. 1117-1142.

[18] Y. Tonomura, A. Akutsu, K. Otsui, T. Sadakata, « VideoMap and VideoSpaceIcon : tools for anatomizing video content », Proc. InterChi'93, ACM, 1993, pp131-136.

[19] M. Yeng, B.-L. Yeo, "Time-constrained Clustering for Segmentation of Video into Story Units", Proc. ICPR'96, August 1996, vol.3, pp. 375–380.

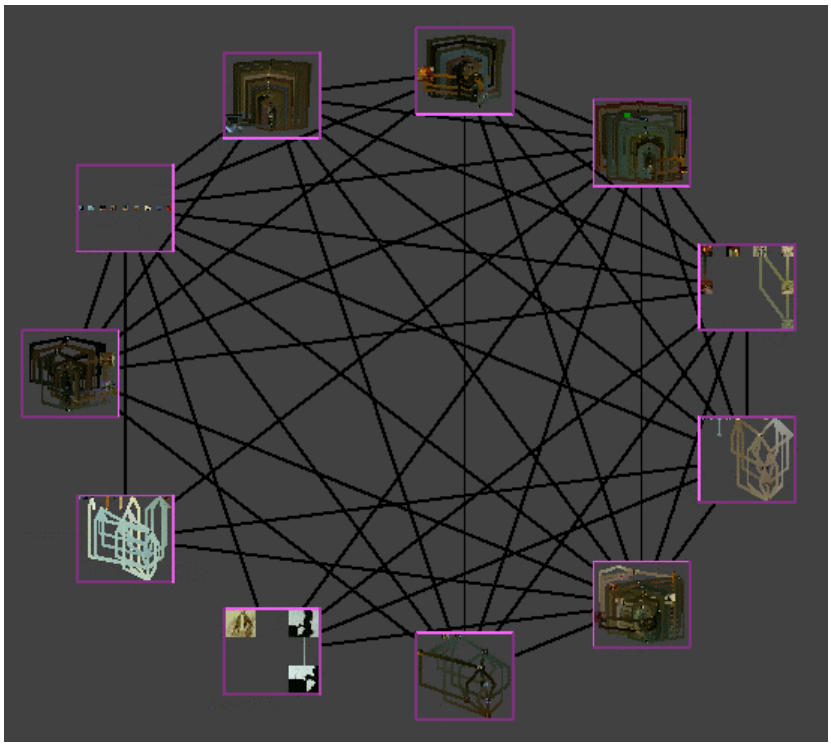

Figure 9. The 11 clusters of "La joueuse de Tympanon".

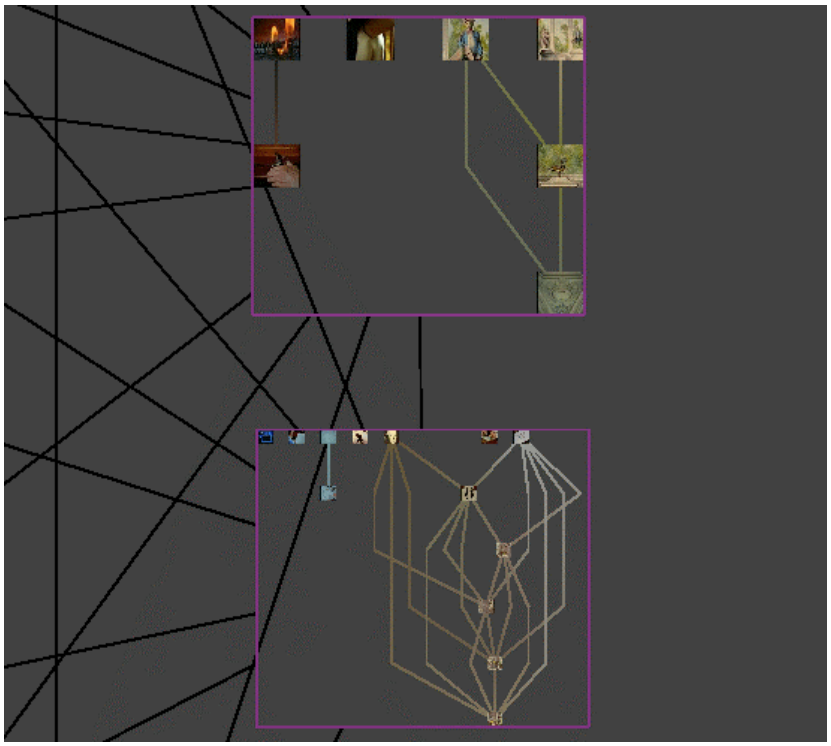

Figure 10. Zoom on clusters of «La joueuse de Tympanon ».

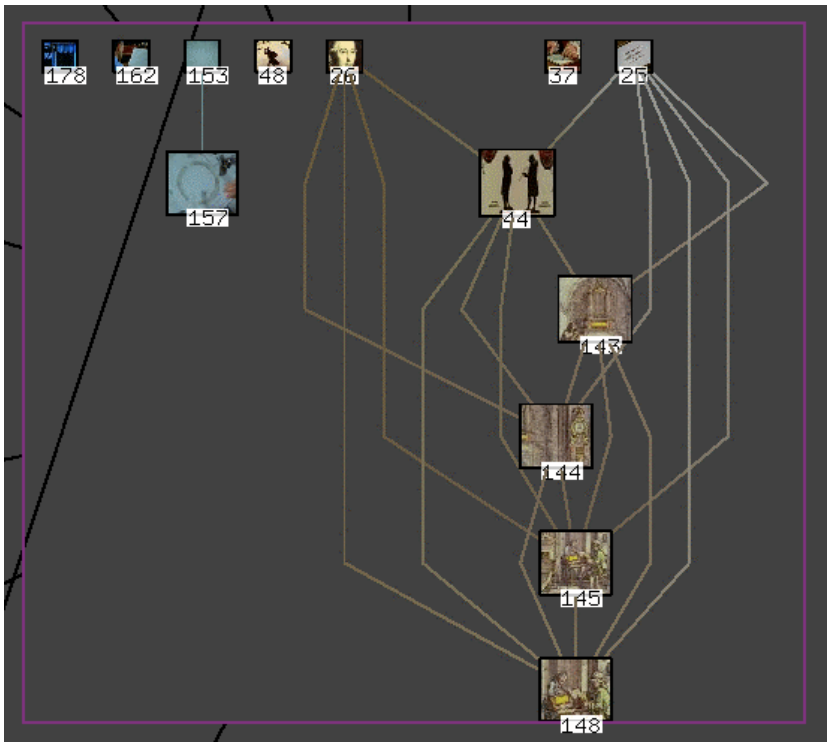

Figure 11. Navigating in a cluster of "La joueuse de Tympanon".

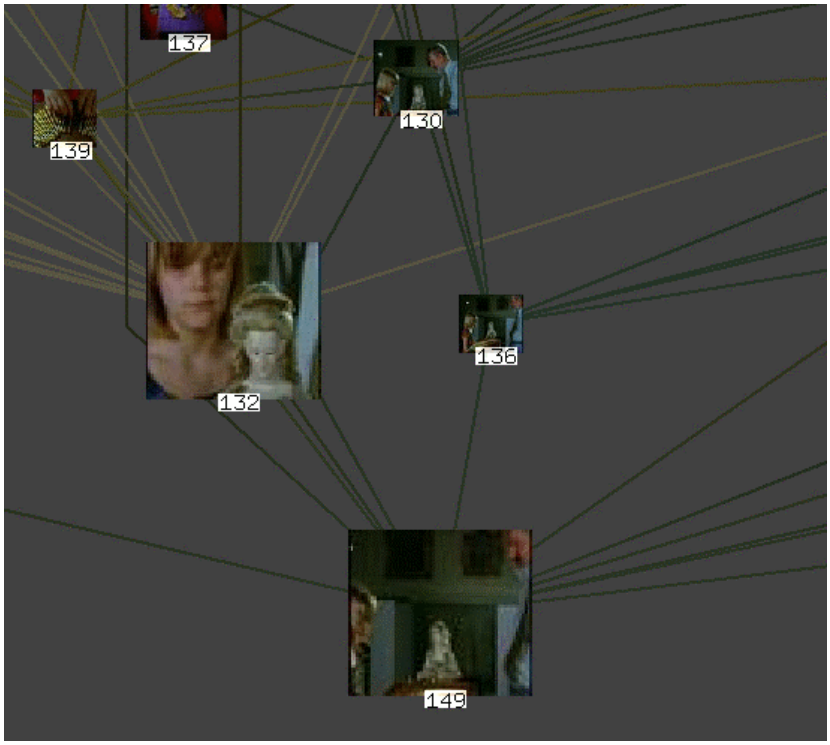

Figure 12. A Cluster of "La joueuse de Tympanon".

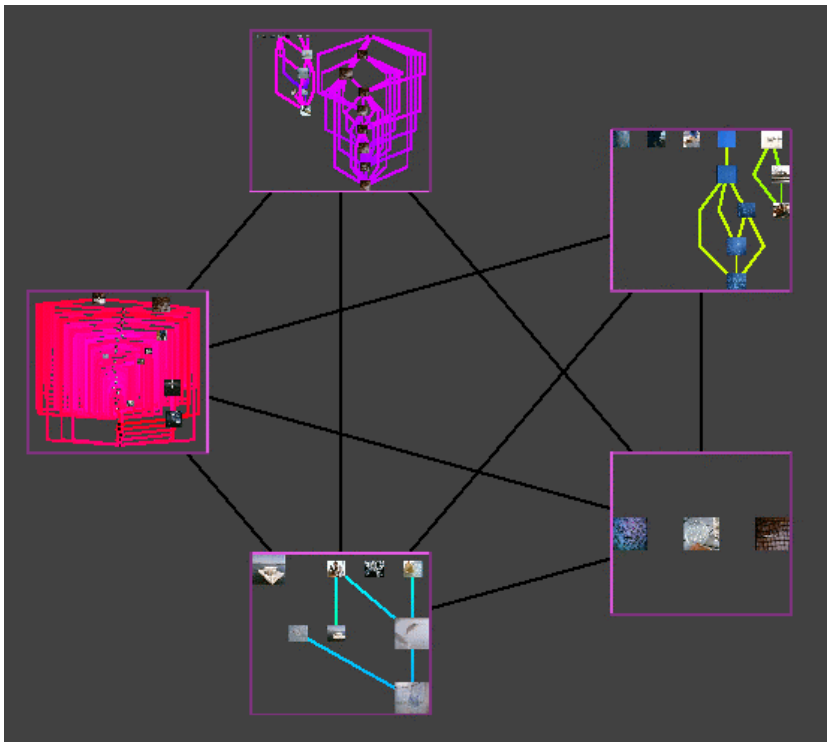

Figure 13. The 5 clusters of “Aquaculture”.

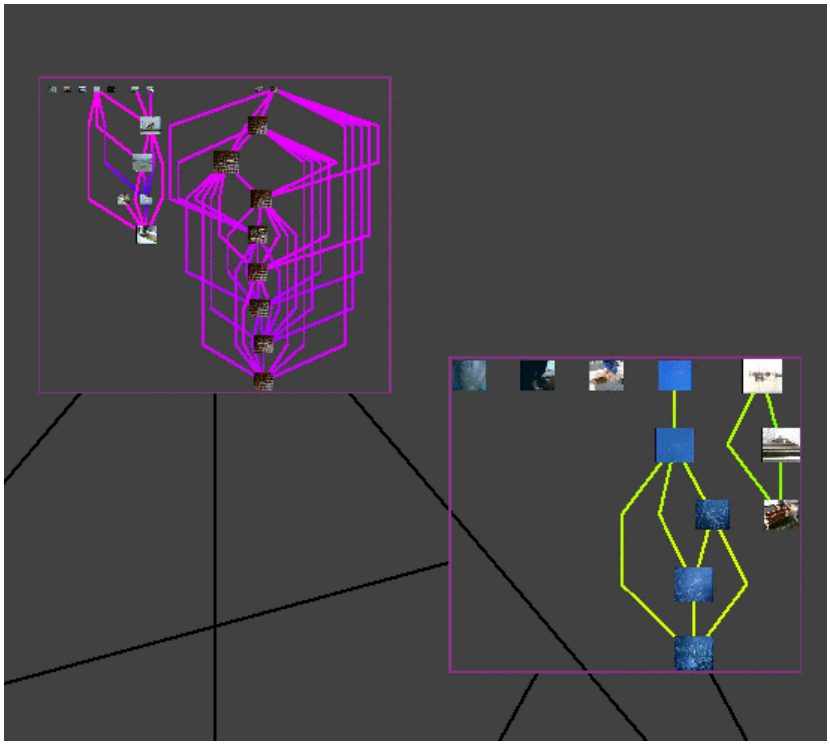

Figure 14. Zoom on clusters of “Aquaculture”.

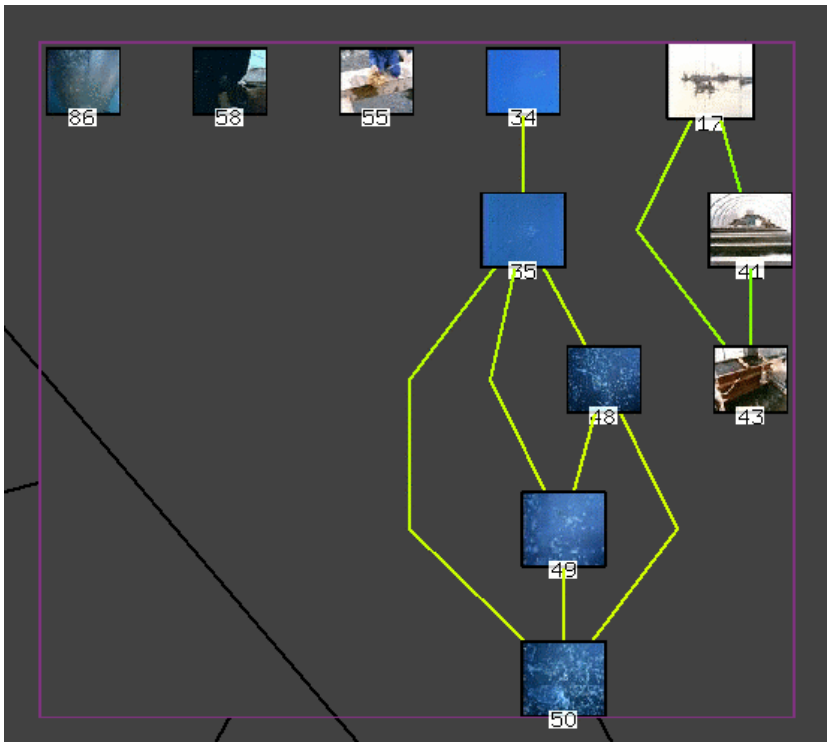

Figure 15. Navigating in a cluster of “Aquaculture”.