State of the art on advanced visualisation methods D7.2

Project Number: FP6 - 045389
Deliverable id: D7.2
Deliverable name: State of the art on advanced visualisation methods
Date: 24 september 2007

COVER AND CONTROL PAGE OF DOCUMENT	
Project Acronym:	VITALAS
Project Full Name:	Video & image indexing and Retrieval in the Large Scale
Document id:	D7.2
Document name:	State of the art on advanced visualisation methods
Document type (PU, INT, RE)	PU
Version:	2.0
Date:	24 september 2007
Authors: Organisation: Email Address:	Anne Verroust-Blondet INRIA Anne.Verroust@inria.fr

Document type PU = public, INT = internal, RE = restricted

ABSTRACT: the deliverable presents an overview of the existing visualisation techniques which help the user to navigate and compute queries inside large domain-independent multimedia databases. In the VITALAS project, many documents are composed of different media: image, audio, video and text. The aim of this deliverable is to identify the families of visual tools that would be useful to study and develop in VITALAS WP7.

KEYWORD LIST: information visualisation, 2D embedding techniques, multi dimensional scaling, video content summary

MODIFICATION CONTROL			
Version	Date	Status	Authors
1.0	10 august 2007	Draft	Anne Verroust-Blondet
2.0	24 september 2007	Final	Anne Verroust-Blondet

List of Contributors
- Hervé Goëau, INA
- Jérôme Thièvre, INA
- Anne Verroust-Blondet, INRIA
- Marie-Luce Viaud, INA

Contents

1 Introduction

We will focus here our interest on visualisation and interaction methods devoted to navigation and queries on large domain-independent multimedia databases. In the context of VITALAS, these databases may contain images or audio-visual documents which can be annotated.

The aim of a visualisation tool, in such a case, is to provide the user a visual representation of the whole database, or of a part of the database, adapted both to the specificity of the data (image, audio, video) and to the user's goal (navigation, query, annotation, etc.). This representation is generally based on the features automatically extracted from the data.

Browsing strategies for searching have been studied in information searching [Bat02] and in the WWW context [CP95] (where three browsing strategies are listed: *search browsing*, directed search where the goal is know; *general purpose browsing*, consulting sources that have a high likelihood of items of interest and *serendipitous browsing*, which is purely random). These three browsing strategies also seem to be useful for multimedia databases. In fact, a study of the end-user needs in digital newspaper photo archives [MS00] and a more recent one on video [MS06] showed that browsing is an essential strategy in searching photos or videos. Moreover, user experiments made in [RBSW01, Rod02, LXT+04, CC05] have revealed that similarity based organisation facilitates browsing and searching in digital databases.

When the results of queries are presented in an array, it can be difficult for the user to obtain rapidly an overview of them. Thus it is useful to introduce alternative visual representations to facilitate annotation and navigation processes. These representations are based on the type of data and use the features extracted from the documents and the similarity matrices computed during the indexing process (i.e. work-package WP2 in VITALAS project).

Since a few years, interactive systems offering visual tools devoted to navigation and annotation of digital archives (images and-or videos) have been developed. We present an overview of the main systems in the next section.

Most of these approaches are based on 2D representations of the multimedia content issued from information visualisation methods. They are adapted to browse inside image archives and propose solutions helping the user to obtain a global view of a large collection of images or to navigate inside the database. We describe the main methods in section 3.

The specific case of video is examined in section 4 where we present existing methods enabling the user to browse inside videos.

Finally, we conclude in section 5.

2 Visualisation prototypes

Except the Autonomy products, the systems presented in this section are dedicated to image or audio-visual archives and propose visual tools helping navigation or annotation. When the archives contain video sequences, the systems presented here visualise only images extracted from the videos. Additional methods proposing solutions to represent video content are described in section 4.

2.1 Autonomy

Autonomy (cf. http://www.autonomy.com/) is a commercial product generating visualisation applications based on the process of automatic clustering. The clusters are generally built using textual information (as in the example of Figure 1), but this visualisation tool can be adapted to multimedia content. The 2D cluster map gives an overview of the database content and the spectrograph displays the relationship between clusters on successive periods and sets of data.

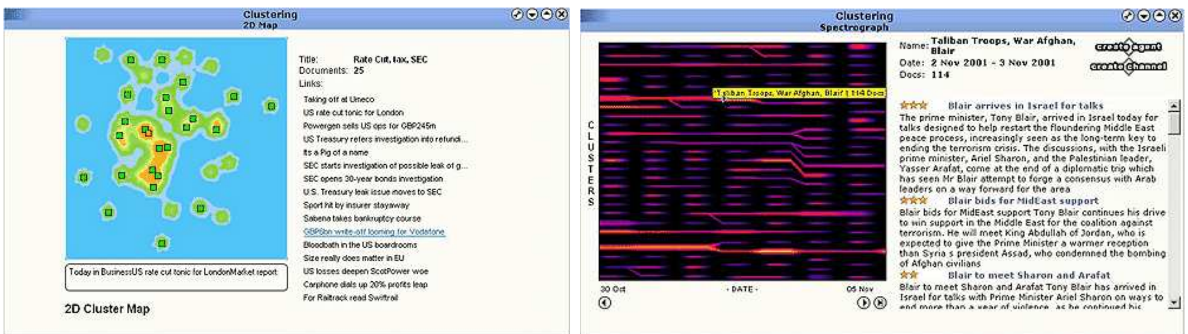

Figure 1: Left: Autonomy 2D cluster map. Right: Autonomy Spectrograph.

2.2 PhotoFinder

PhotoFinder prototype [KS00], developed by the University of Maryland Human Computer Interaction Laboratory, is a system that helps non-technical users searching and browsing personal photo collections. Direct annotation enables personal names to be easily placed on a photo as in Figure 2. Contrary to the other systems presented in this section, PhotoFinder does not calculate a 2D cartography of the collection based on the visual similarity of the photos. In fact, the axes of the thumbnail browser associated to the scatter plot display for the thumbnail pictures are defined by the user: he chooses the attributes associated to the x and y axes (for example the picture rating, the number of people present in the images, etc.) among the attributes attached to the photos.
This prototype is the ancestor of PhotoMesa Image Browser developed by Windsor Interfaces Inc. (see http://www.cs.umd.edu/hcil/photomesa/).

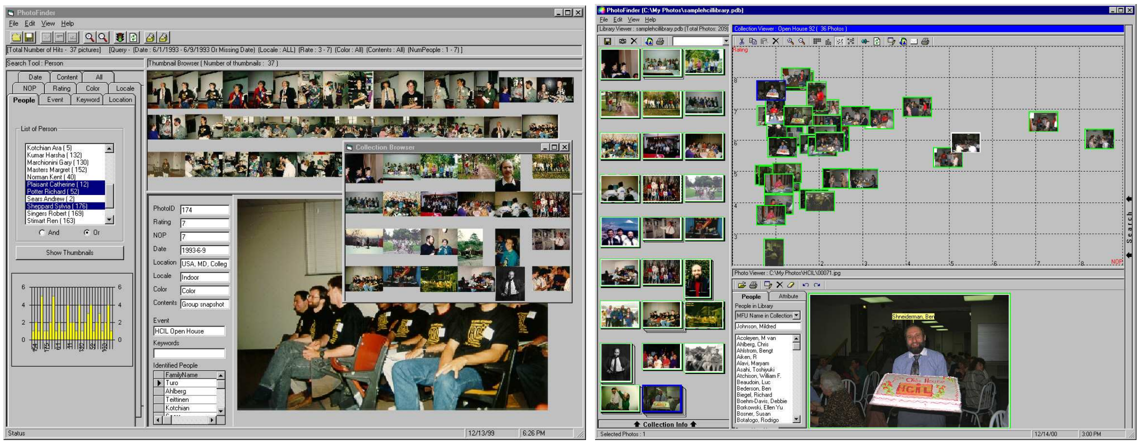

Figure 2: PhotoFinder system (with search tool, thumbnail browser and detail viewer windows on the left and with direct annotation on the right).

Figure 3: AETOS combined with ANSE to browse inside a video

2.3 AETOS

The AETOS system developed at the Multimedia and Information Systems Team of the Imperial College of London [HR04b, HR04a] is a graph-based system that allows the user to browse in large image collections. In their system, an arc is established between two images if there are one or more features for which the images are nearest neighbours (see the NN^k networks description in section 3). The arcs are weighed by the proportion of the features for which the nearest neighbour relationship holds. In [HHM+04, Pic04] the video summarisation process ANSES [HPRY03, PR03] is mixed with the AETOS system (cf. Figure 3).

2.4 FutureViewer

Campanella et al. [CLM05a, CLM05b] propose a visual environment called FutureViewer to explore and annotate audio-visual sequences. The user can explore graphically how the basic segments

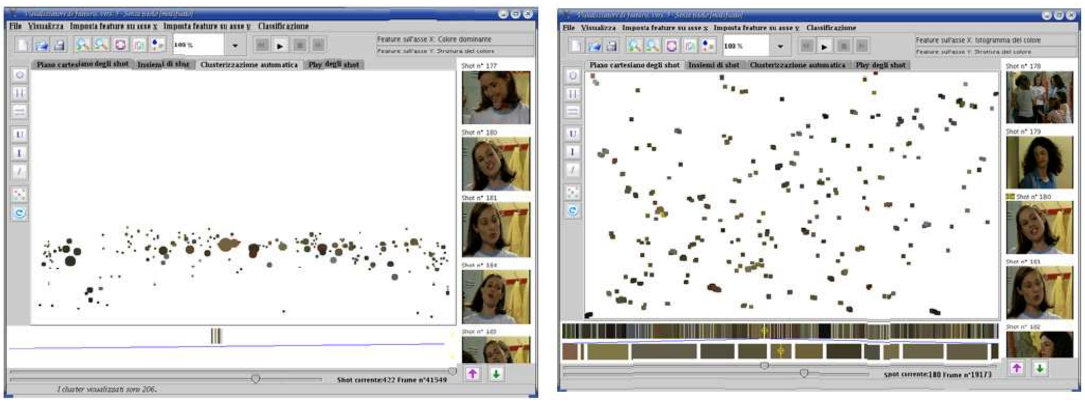

Figure 4: FutureViewer: clusters visualisation with Principal Component analysis (left) and visualisation of the feature space (right).

of a video sequence are distributed in the feature space and can interactively visualise clusters and annotate them. More precisely, each audiovisual document is considered as a sequence of shots. MPEG7 features are then extracted from each shot. The feature space is displayed in a 2D Cartesian plane, where each axis corresponds to one feature type selected by the user and each shot is represented by a little square filled by the dominant colour of the shot (see the right part of Figure 4). Shots are clustered w.r.t. their temporal distance and the MPEG7 features selected by the user. The calculated clusters are visualised mapping their centroids from the n-dimensional space to a 2D plane with the Principal Component analysis (cf. left of Figure 4).

2.5 MediaMill

In the Intelligent Sensory Information Systems team of the University of Amsterdam, the MediaMill video search engine [SWKS06, WSK+06, WSdR+07] proposes four browsing tools:

- The Galaxy Browser, which is issued from Nguyen and Worring's work [NW05a, NW05b, Ngu06, NW07a, NW07b] and uses 2D similarity-based visualisation of image frames where the user can annotate a large collection of images (cf. Figure 5) . Mixing relevance feedback and 2D visualisation this system can improve the efficiency of the relevance feedback process. In fact, by using relevance feedback given by the user, the dissimilarity space can be updated. In [Ngu06, NWS06], a 2D manipulation space in which the user interacts with the images is created .

Figure 5: on the left, the Galaxy browser. The upper-left corner shows the whole collection as a point set with red points representing the currently displayed set. The bottom-left corner shows an enlarged version of the currently selected thumbnail. The main screen shows the representative set as images. The green rectangle is an example of a user selection of a group of images by dragging the mouse. On the right, the sphere browser.

- The Sphere Browser (right of Figure 5) demonstrates a novel interface for searching through semantic space using conceptual similarity. This is done by (off-line) clustering shots with a similar conceptual index together into threads. The Sphere Browser shows the time-line of the current video on the horizontal axis, and for each shot from the video it displays the relevant threads on the vertical axis.

Figure 6: From left to right: the cross browser and the rotor browser.

- The Cross Browser (left of Figure 6) exploits the fact that ranking is a linear ordering. The horizontal axis is used for the visualisation of the time-line of the video program from which a key frame is selected, while the vertical one is associated to a selected concept. This makes sense as frequently other items in the same broadcast are relevant to a query also.

- The Rotor Browser (right of Figure 6), which can be interpreted as an extension of the Cross Browser to more than two axes: here the current active shot is visualised and the semantic thread in which this shot occurs (time, similar speech or visual content, etc.).

3 2D visualisation of image archives

The methods described in this section must facilitate the user's interaction when browsing and navigating through the database. Thus two different problems have to be solved:

- how the database content can be structured and represented on a 2D map so that the user gets quickly an overview ?
- what kind of representation will help the user to navigate through the database, starting from a given image ?

Let us consider the first problem. In [Ngu06, NW07a] three requirements for a generic system visualising a large visual collection are defined:
R1: Overview requirement: the visualisation should give a faithful overview of the distribution of images in the collection,
R2: Structure preservation requirement: the relations between images should be preserved in the projection of the information space and the visualisation space,
R3: Visibility requirement: all displayed images should be visible to the extent that the user can understand the content of each image.
These requirements are valid for the visual documents and also for the textual annotations associated as soon as the notion of similarity between two documents can be established.
Thus, if we use the similarity matrices associated to the objects, the "ideal" is to build a 2D representation of the objects such that the similarity distances are preserved. The similarity matrices may concern either the visual features, the textual annotation or the audio features of the multimedia content, as it is the case in [CKGB02, PDW03, NR07].

When the purpose is to navigate through the database, the *overview requirement R1* may be not necessary and the layout shall concern only a neighbourhood of an object or of a small set of objects and a local view of the dataset will be built (cf. section 3.2).
When the 2D layout produces a representation of the whole collection, then a global view of the dataset is computed (cf. section 3.1).

3.1 Global view of the dataset

In most of the cases, we have a collection X of N objects $\{x_1, ..., x_N\}$ and a $N \times N$ similarity matrix $\Delta = (\delta_{ij})$ where δ_{ij} is the similarity distance between objects x_i and x_j. The purpose of a 2D embedding method is to associate to each object x_i a vector $\vec{y}_i$ in the 2D space, such that the neighbours of x_i are in the neighbourhood of $\vec{y}_i$ in the 2D space and vice versa. In the following d_{ij} will denote the distance between $\vec{y}_i$ and $\vec{y}_j$.

3.1.1 Approaches based on projections

If the feature points $(p_1, ...p_N)$ associated to the similarity matrix Δ of the N objects are given in a high dimensional space, the 2D embedding problem can be solved by finding a 2D space such that the distance between points is preserved in the process of projection on this subspace. Two methods are based on this principle:

FastMap [FL95]. In this case, the points are recursively projected to the hyperplanes perpendicular to an orthogonal set of 2 lines passing through the most dissimilar objects called pivot objects chosen by an heuristic.

PCA, used in [CLM05a, CLM05b, MTT01b, MTT01a]. Considering $Y = (p_1, ...p_N)$ as a matrix, the eigenvectors and the eigenvalues of $Y'Y$ are computed and the points $(p_1, ..., p_n)$ are projected on the subspace defined by the two greater eigenvalues of $Y'Y$.

3.1.2 Approaches based on graphs

Graph visualisation being a huge domain, we will focus our interest in systems adapted to the visualisation of audio-visual data. In these approaches, each image is represented by a vertex in a graph (directed or not). The arcs represent the neighbouring relations between images and they are used for navigation. The 2D visualisation is built from the graph layout.

Graph modelling may be performed on any type of data containing objects and relations between objects. The characteristics of "what should be perceived" specify the type of layout and the graphic properties for a given visualisation. In our case, nodes may be images, reportings, video sequences or documentary notes, and links or edges represent audio, temporal, visual or textual neighbouring relations. Edges may be valued (Boolean or numerical), directed or not, simple or multiple.

A lot of work have been done on graph visualisation (see for example the paper of Herman et al. [HMM00]) and on graph layout (orthogonal, force directed layouts, etc., see [BETT99, KW01]). The main objectives for global graph layout are to preserve node's neighbourhood in the 2D representation and to make emerge data structure.

For retrieval and analysis purposes, two types of layout are interesting: tree layouts which develop relational view from a focus, and global graph layouts, which present an overview of the nodes distribution.

Radial tree layout [YFDH01] is used to produce interactive focus-based views of a graph. Continuous animations are generated based on polar interpolation of the nodes coordinates to ensure the preservation of the user mental map during focus change, cf. Figure 7. This kind of layout is well-suited for local exploration of graphs.

Force-directed algorithms have been widely studied [FR91, Ead84, BH86, KK89] for graph drawing. One of the main issues in graph visualisation is the size of the graph. In fact, the algorithmic complexity on graph structures is more often on $O(N^2 log(N))$. A large part of the research works in the last past years have been focused on reducing the cost of the layout, introducing multilevel approaches to generate graph layout [HK02, Wal03, HJ04, KCH03].

A major drawback of classical force models is their inability to produce informative layout for very dense graphs. [Noa04] introduces a new force model called LinLog, especially designed for computing layouts that reveals cluster in dense graphs (cf. Figure 8). The vertices attractive energy is linear while the repulsive one is logarithmic. The energy for a layout p is then defined as:

$$U(p) = \sum_{u,v \in E} \|p(u) - p(v)\| - \sum_{u,v \in V^2} - \ln(\|p(u) - p(v)\|)$$

This model is well designed for graphs with uniform edge degree. [Noa06] extends this approach

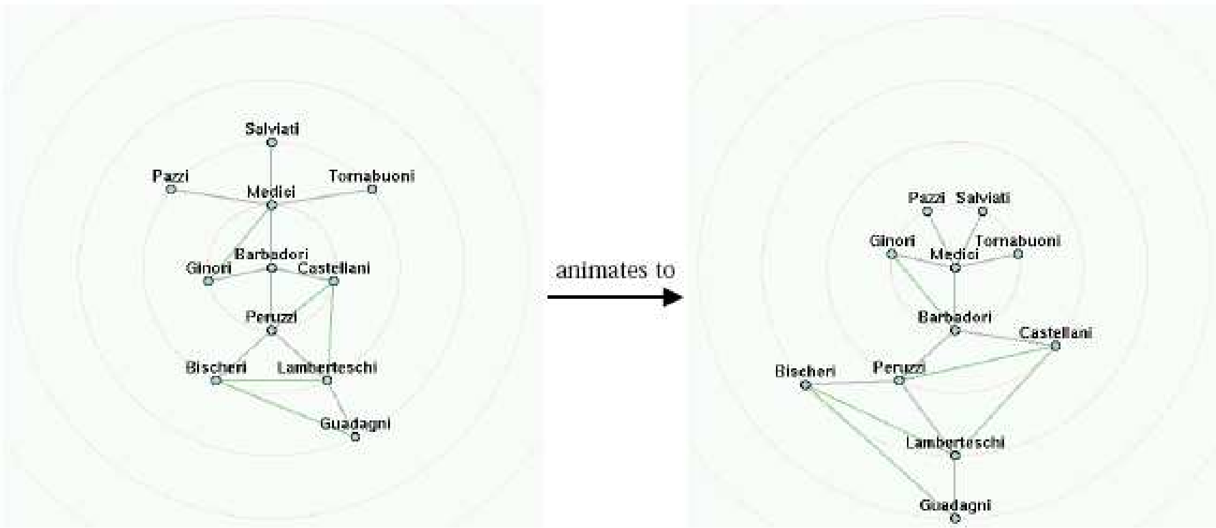

Figure 7: Animated radial tree layout of a social graph.

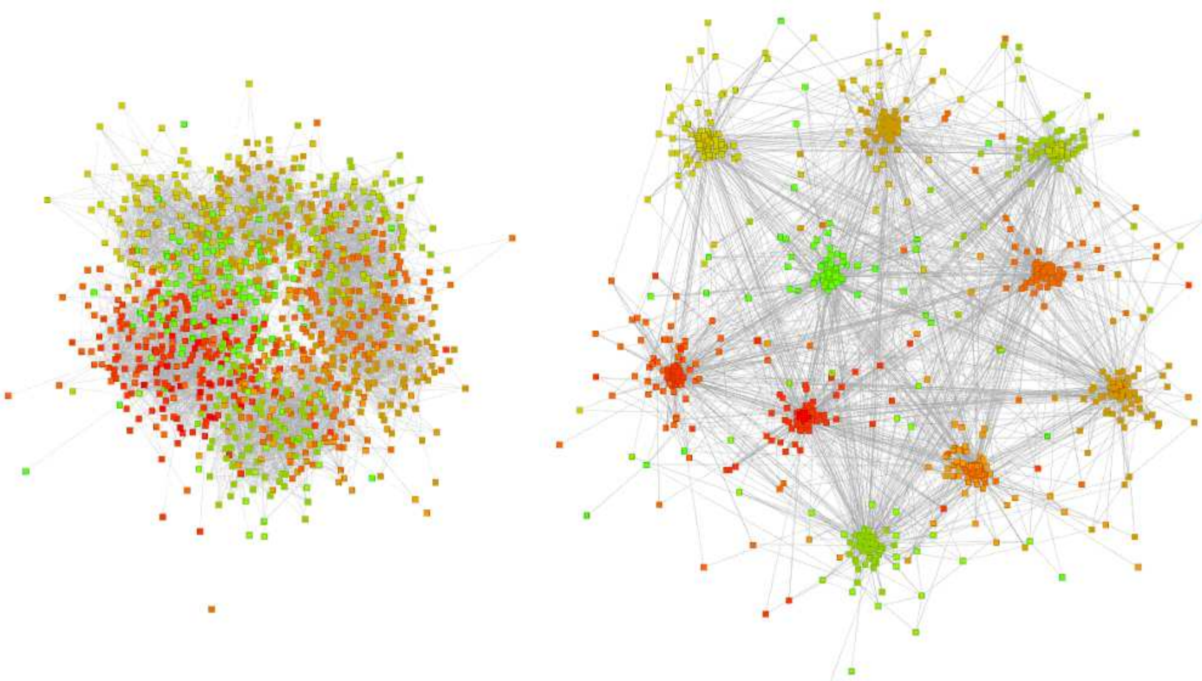

Figure 8: Fruchterman-Reingold and Edge LinLog layouts.

to overcome this limitation with the edge-repulsion LinLog energy model:

$$U(p) = \sum_{u,v \in E} \|p(u) - p(v)\| - \sum_{u,v \in V^2} -deg(u)\ deg(v) \ln(\|p(u) - p(v)\|)$$

Interactive focus-based exploration of graph may be a solution to large graph visualisation.

[vHvW04] replaces the linear attractive force by an r-polynomial force to avoid local minima, r is decreasing with the system energy to reach a linear force. Data clustering produces a hierarchy of partitions, called dendograms, used to generate visual aggregates. A fisheye-like lens renders continuously the clusters at different levels of details, keeping the whole graph context (see Figure 9).

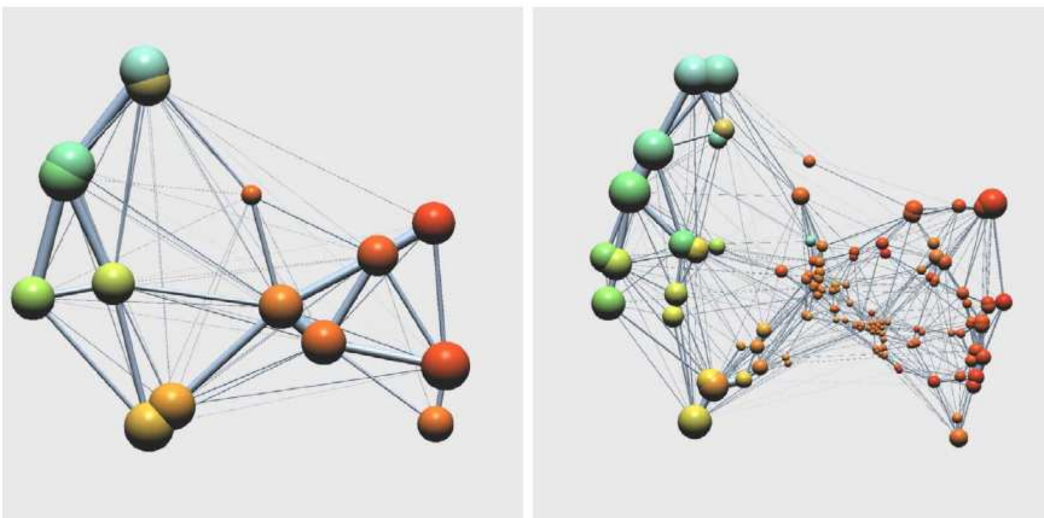

Figure 9: Visualisation of small-world graph.

[AMA05] is a system which builds a hierarchical description of a graph according to the topological features it contains (tree, small word, complete subgraph, biconnected components, near mesh). Then, a multiple level drawing algorithm integrates tree, circular and force directed layout. This allows to draw each identified subgraph with the most adapted layout without runtime penalty (cf. Figure 10). An interactive version has been described in [AMA07].

One aspect which may be challenging for users is that visualisations produced by force directed layout are not deterministic. Global data distribution is preserved but spatial location may vary. This behaviour may have a negative impact on the ability of the user to build a mental map of the data. [FT07] presents an algorithm for drawing a sequence of graphs online which maintains the global structure of the graph. The algorithm allows arbitrary modifications between consecutive layouts. The algorithm works online and uses various execution culling methods in order to reduce the layout time and handle large dynamic graphs.

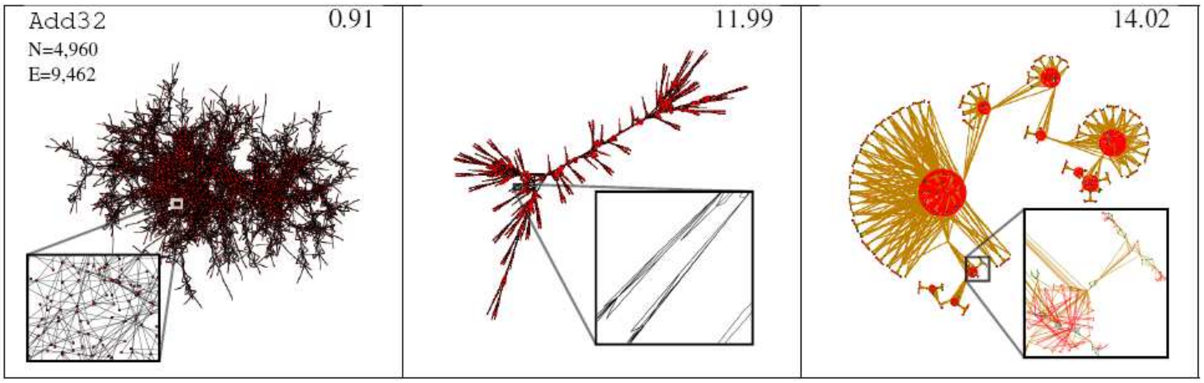

Figure 10: Grip, FM3, and TopoLayout of the same graph.

3.1.3 Multi Dimensional scaling and derivative methods

Multi Dimensional Scaling (MDS) [Sam69, BP06] method is based on the iterative optimisation of a stress function that measures the difference between the 2D Euclidean distances d_{ij} and the original dissimilarities δ_{ij} between pairs of objects. This method is very expensive ($O(n^3)$ or $O(n^4)$). Then some other approaches have been introduced to reduce this computational cost.

Isomap [TdSL00, SMR06] takes as input the similarity distances δ_{ij} between all pairs x_i, x_j of N objects. The algorithm first constructs a neighbourhood graph, having objects x_i as nodes and by connecting two nodes x_i and x_j is they are closer than ϵ or if x_i is one of the K nearest neighbours of x_j. The edge lengths are set to δ_{ij}. Using this graph, the shortest paths $d_G(i,j)$ between two nodes x_i and x_j in the graph are computed. Let $D_G = [d_G(i,j)]$. Then a classical MDS is applied on D_G, constructing an embedding of the data in the $d-$dimensional space.

Approaches based on Chalmers'method [MC03, MC04, JM04]
They mix graph layout with spring model [Cha96] and MDS.
In [MC03],

1. They take $\sqrt{N}$ sample elements and they compute their layout by an iterative process using the spring model

$$Stress = \frac{\sum_{i<j}(\delta_{ij} - d_{ij})^2}{\sum_{i<j} d_{ij}^2}$$

on a subgraph of the complete graph (each node is linked with a subset of the $\sqrt{N}-1$ nodes formed by its V nearest neighbours and a sample of S other nodes), δ_{ij} being the similarity distance between x_i and x_j and d_{ij} the layout distance.

2. Each element x_k not yet positioned is associated to one of the $\sqrt{N}$ samples: it is the nearest element x_i of x_k among the $\sqrt{N}$ samples and x_i is x_k's parent. Then a subset S_k of size $N^{\frac{1}{4}}$ of the $\sqrt{N}$ sample elements, containing x_i is formed. x_k is placed near the position of x_i by an iterative process in such a way that

$$\Sigma_{j \in S_k} |d_{kj} - \delta_{kj}|$$

is minimised.
The complexity of their method is reduced to $O(N^{\frac{5}{4}})$ in [MC04].
Jourdan and Melançon [JM04] extend Morrison and Chalmers' approach [MC03] and obtain a $O(Nlog(N))$ method, improving parent-finding strategy inside step 2 of the previous approach.

LMDS (local multidimensional scaling) [CB06] only uses local information from user-chosen neighbourhoods and uses ideas from the area of graph layout. A common paradigm in graph layout is to achieve desirable drawings of graphs by minimising energy functions that balance attractive forces between near points and repulsive forces between non-near points against each other. The forces consist in a parameterised family of stress or energy functions inspired by Box-Cox power transformations.

Visumap [Li04] is based on the Relational Perspective Map (RPM) method. Given a distance matrix δ_{ij} called the relational distance between N objects, the goal of the RPM algorithm is to map the objects x_i into a two dimensional map, called relational perspective map, so that Euclidean distances d_{ij} between the image points visually approaches δ_{ij} as much as possible. The

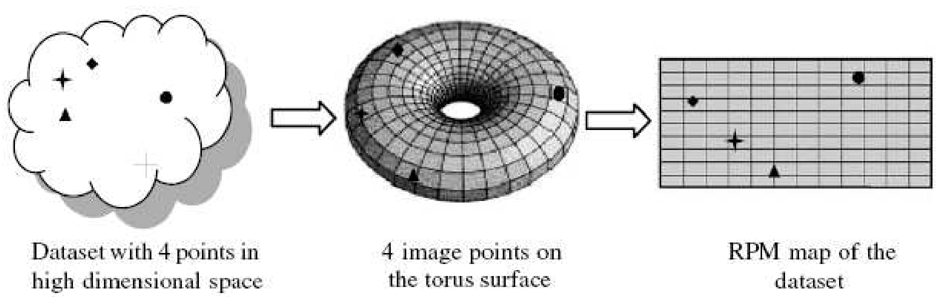

Figure 11: The RPM method.

RPM algorithm first maps data points to the surface of a torus, then unfolds the torus surface by a vertical and a horizontal cut (cf. Figure 11). An energy function E_p is introduced:

$$E_p = \sum_{i<j} \frac{\delta_{ij}}{pd_{ij}} \text{ and } E_0 = \sum_{i<j} \delta_{ij} ln(d_{ij})$$

where $-1 \leq p \leq 1$ is a parameter called the rigidity and d_{ij} is the geodesic block distance between two image points on the torus surface. The image points on the torus are considered as a force directed multi-particle system with mutual repulsive forces between them. The repulsive force between two points is proportional to their relational distance. The energy E_p is considered as a kind of total potential energy which is minimised by simulating the dynamic system directed by the repulsive forces.

I-PACK [Don06, DH06] is a deterministic layout algorithm based on a hierarchical decomposition of the data set. Given a set of points $X = \{x_1, ..., x_n\}$, they built a multi-level sampling of X, $S_h \subseteq S_{h-1} \subseteq ... \subseteq S_0 = X$ and an associated tree $\mathcal{T}$ such that:

1. $\forall x_i \in S_m$ with $m < h$, $\exists x_j \in S_{m+1}$ s.t. $\delta_{ij} \leq d_{m+1}$
2. $\forall x_i, x_j \in S_m$ with $x_i \neq x_j$, $\delta_{ij} \geq d_m$.

The 2D layout algorithm takes as input the tree $\mathcal{T}$ and assigns to each node of $\mathcal{T}$ a square containing its subtree (cf. Figure 12).

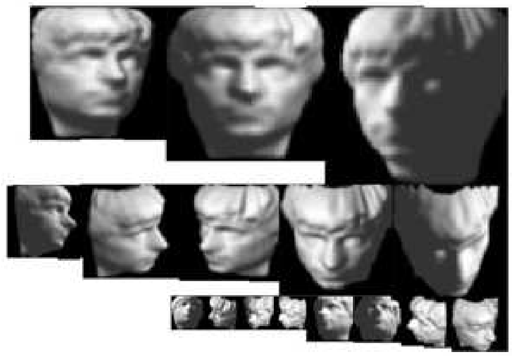

Figure 12: Images of level S_5.

3.1.4 Other methods

LLE (local linear embedding) [RS00] locally linear embedding (LLE), is an unsupervised learning algorithm that computes low-dimensional, neighbourhood-preserving embeddings of high-dimensional inputs.

Given N points $(p_1, ..., p_N)$ associated to the objects $(x_1, ..., x_N)$, the algorithm first computes the matrix of distances δ_{ij} and selects as neighbours of p_i its K nearest neighbours. Then it assigns weights w_{ij} to all points p_j, $j \neq i$, summarising the contribution of the jth data to the ith reconstruction. When x_j is not a neighbour of x_i, $w_{ij} = 0$. In the other cases, the optimal weights w_{ij} for the neighbours of x_i are computed by solving a least square problem (constraint $\sum_j w_{ij} = 1$). Then $\vec{y}_i$, the embedding coordinates of x_i, are computed by minimising the cost function:

$$\Phi(\vec{y}) = \sum_i |\vec{y}_i - \sum_j w_{ij}\vec{y}_j|^2$$

Self-organising maps (SOMs) are a data visualisation technique introduced by T. Kohonen [Koh95] which reduces the dimensions of data through the use of self-organising neural networks. The maps consist of regular lattices of neurons set in hexagonal or rectangular topology. As each element has a fixed number of neighbours, requirement R2 cannot be satisfied in all the cases. Deng et al.[DZP04, Den07] use this technique to produce a hierarchic content-based summarisation and comparison of an image collection (cf. Figure 13). Curvilinear Component Analysis [DH97] can be

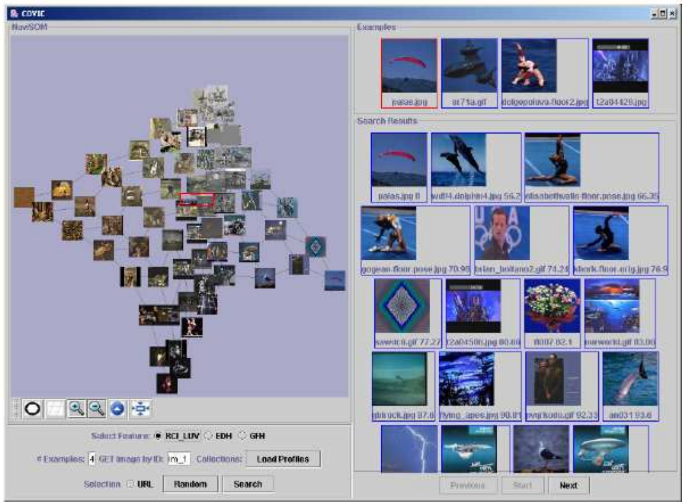

Figure 13: The 2D embedding of a 3D lattice built by SOM technique.

used instead of SOMs in this context.

Stochastic Neighbour Embedding [NW05a, NW05b, NW07b, NW07a, Ngu06] is a probabilistic projection method.
For each object x_i, the asymmetric probability

$$p_{ij} = \frac{exp(-\delta_{ij})}{\sum_{k \neq j} exp(-\delta_{ik})}$$

is the probability that x_i would pick x_j as neighbour, δ_{ij} being the distance between objects x_i and x_j in the high dimensional space.
In the 2-dimensional space, SNE initialises $\{\vec{y}_i\}$ at random positions and

$$q_{ij} = \frac{exp(-\|\vec{y}_i - \vec{y}_j\|^2)}{\sum_{k \neq j} exp(-\|\vec{y}_i - \vec{y}_k\|^2)}$$

is the probability that x_i would pick x_j as neighbour.
To preserve distances to nearby points as much as possible, a cost function C

$$C = \sum_i \sum_j p_{ij} log(\frac{p_{ij}}{q_{ij}})$$

measuring the distance between these two distributions is minimised using a gradient descent method.

Parametric embedding, PE [ISU$^+$07] is adapted to visualise classes. More precisely, it is a probabilistic method that embeds objects $X = \{x_1, ..., x_N\}$ and classes $C = \{c_1, ..., c_K\}$ simultaneously into a low-dimensional visualisation space. It takes as input a probabilistic distribution $(p(c_k \Delta x_n))_{k,n}$ for objects over classes (or for one set of points over another set). Assuming that each class c_k can be represented by a spherical Gaussian distribution in the embedding space, they attempt to fit that distribution in the embedding space by minimising a sum of Kullback-Leiber divergences.

3.2 Local view of the dataset

When navigating inside a dataset, the user does not necessarily need to visualise the whole data. Local approaches as NN^k networks and NAV-GRAPHE are proposed in this context.

3.2.1 NN^k networks

NN^k networks [HR04b, HR04a, HR05, Pic04] is introduced to build the neighbourhood of a given image, given the similarity distance matrices. In this case, the metric is parameterised in terms

Figure 14: Two images are connected if there exists one feature combination for which one image is ranked top when querying with the others.

of feature specific weights w_f associated with specific distance functions d_f: given an image q, its NN^k are all those images in a collection that are closest to it under at least one instantiation of a parameterised distance metric.

$$D(p, q) = \Sigma_{f=1}^{k} w_f d_f[p, q)$$

Once the NN^k images are computed, the layout of q and its neighbouring images is formed by putting q in the center of a circle and its neighbours on the circumference, as in Figure 14. this representation is adapted for browsing inside a database.

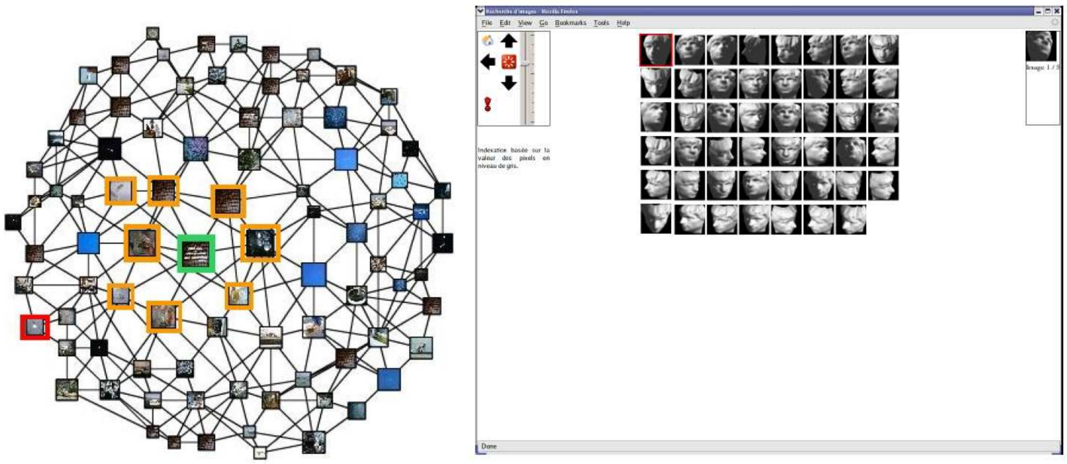

Figure 15: Left: a graph-based navigation tool; right: the hierarchical navigation interface.

3.2.2 NAV-GRAPHE

In [Don06] a hierarchical graph-based navigation process is presented. It is based on a greedy routing algorithm: given a starting node (in green in Figure 15) and its neighbours (in orange in Figure 15), the user chooses an image (in red in Figure 15) that will be used as starting node for another navigation step if needed. The hierarchical interface (cf. Figure 15) presents a partial view of the navigation graph to the user: only a sample of images is presented at each step of the retrieval task.

4 Visualisation of video content

Most of the approaches presented before propose visual tools adapted to browse inside image archives. The VITALAS project deals also with structured documents such as videos and proposing visual solutions to enable fast and intuitive browsing and access to video content is a challenge.
Various summarisation methods have been the topic of intensive research in the content-based video analysis community. Here, the challenge is to propose a short summary of a longer video document in order to propose a quick overview of the content. Therefore, different tools and visualisations has been explored to achieve these tasks or for showing the summary. These approaches can be split in two categories according to the way the video is considered:

- the dynamic video skimming, which is a shorter version of the original video made up of a set of continuous video clips, like a trailer.
- the static video summary, which is composed of a collection of salient images extracted or synthesised from the original video,

In the first case, tools consider video as a continuous sequence of images. In the second case, the methods focuses on shots segmentation and clustering tasks.

4.1 Continuous visualisation

The most common, oldest and easiest way to get a quick overview of a video content is the temporal slider. Mains problem are the lack of scalability to large document sizes and the unpleasant aspect because of the jerky visual feedback. In [HJ05] the Zoomslider interface solves the scaling problem by zooming into (or out of) the slider's scale. So, a user can visually skim through video at different granularity levels.

4.2 Shot visualisation and static video summary

For the approaches presented in this section, three problems have to be solved:

- which are the important segments composing the video sequence ? Is there an automatic or a semi-automatic way to compute their importance ?
- how the video summary is presented to the user ?
- how the user can browse inside the video to find a particular image or a video segment ?

In [Gir03] a visual video summary is comprised of images from the video that are sized according to their importance (determined by the length and uniqueness of the corresponding video segment). They call this representation "Manga" video summary (cf. Figure 16).
Considering that users do not pay attention to the entire key-frame, face detection and attention region detection techniques can be applied to extract specific regions or Region-Of-Interest (ROI) that users may be interested in, like in [JLZH07]. In [CGL04] key-frames with interesting area are

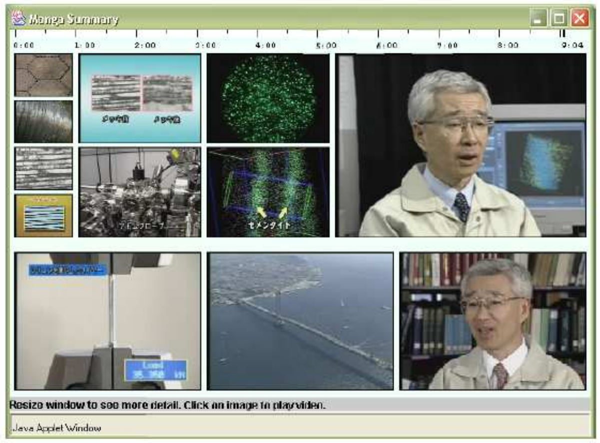

Figure 16: Manga Video Summary

packed and visualised like stained-glass with irregular shapes (cf. Figure 17). Here, a morphological grouping technique is described for finding the ROI. In [WMH07], the problem formulation is viewed as a minimisation of a transition measurement and a maximisation of a representativeness measurement. The representativeness is a combination of various measurements: colon contrast, blurring degree, saliency, temporal distribution.

Figure 17: Generic visual summary of a staff meeting video (left), and the stained-glass version (right).

In [TS06] the video summarisation is guided by the user's photo library. The video is first divided in short segments. These segments are judged "important" when the user's photo library contains many photos with similar content, and it is judged "unimportant" in the other cases. Summarisation is done by cutting off the unimportant segments and reconstructing the remaining segments into a coherent whole. Then the summary is presented to the user in an array.

In [CS06] a hierarchical representation of key frames of a video sequence for video summarisation is proposed. After removing the meaningless key frames, the algorithm groups key frames

Figure 18: A video collage of a home video. Salient region of interest are excerpted from representative key-frames. The temporal structure is preserved in the order left to right and top to down.

into visually and semantically homogeneous clusters to allow hierarchical summary representation. Then the user can browse the video content at different level of details starting from the default summary level build by the system and presented to the user in a tabular way.

The panoramic mosaics representation has been widely explored for security application and shot summarisation like in [TAT97]. In [GCSS06] the work presented attempts to extend the expressiveness in this kind of visualisation by adding a layout inspired from the visual language of storyboards (cf. Figure 19).

Figure 19: Schematic storyboard: the 3D arrow represents the motion activity of the scene.

For static shots from a CCTV camera in monitoring and surveillance systems, only events can be overlayed in one global summary image as in [PJ05] (cf. figure 20).
Introducing the time as a third dimension, the video sequence can be considered as a volumetric data. Then volume visualisation techniques can be used for summarising sequences, as in [DC03]: Figure 21 shows a visual representation for detecting events in one shot. The horseshoe-shaped is an interesting way to reduce the time line. The video visualisation pipeline is able to represent an entire video in 5 different views (cf. figure 22). The volumetric representation is also used in [FLM00] where one shot is represented as cube for a browsing application (cf Figure 23).

Figure 20: A static summary of half an hour of security video.

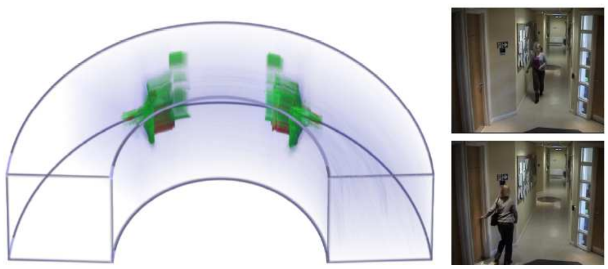

Figure 21: Volumetric shot visualisation horseshoe-shaped. The left part shows the levels and patterns of the activities recorded in the shot.

In [LFRS06] an interface with user interaction through a visualisation in a large-scale news video collection. The main objective is to demonstrate the power of integrating visualisation with automated analysis techniques. The visualisation can help users find news stories of interest in a short time. The interface shows a map of key-frames corresponding to relevant news stories automatically detected from a daily news programmes. The sizes of the key-frames are proportional to the a interestingness measurement. The animation between two successive daily news like in figure 24 shows the trend of topic change over time. By clicking on frames, user define a personalised attention model, which improve measurement of interestingness.

Figure 22: 5 visual representations of one news programme as volumetric objects.

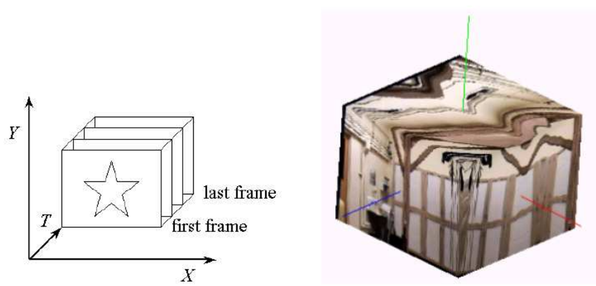

Figure 23: Video cube of 7 seconds of data.

Figure 24: Animated sequence between two key-frames map of successive daily news video. The maps show the news topics on given day, the animation represents the trend of topic change over time.

5 Conclusion

Most of the functionalities offered by the approaches presented previously seem to be adapted to the context of the VITALAS project, where the goal is to build a prototype system dedicated to intelligent access services to multimedia professional archives. In fact, existing approaches elaborate visual tools enabling the user to:

- get quickly an overview of the content of the database and annotate by groups neighbouring documents, thus to develop a 2D embedding approach that satisfies the three requirements expressed in section 3 and to allow the user to interact with it. When relevance feedback facilities are provided to the user, updates of the dissimilarity space can be taken into account in the visualisation modules (like in [Ngu06, NWS06]).
- navigate through the database starting with an element (approaches presented in section 3.2). Then it seems also interesting to develop an interactive approach, computing a local view of the dataset corresponding to the neighbourhood of the document and enabling the user to navigate smartly through the database.
- obtain a summary of the content of a video sequence and browse inside it (cf. section 4).

The graph-based approaches seem to be promising because they offer a unified and intuitive representation model for all the user's actions listed above. Moreover, the graph representation has proven to be effective in other applications, such as the WWW representation [Che97]. But it would also be interesting to offer the user alternative visualisations such as the Cross Browser and the Sphere Browser of the MediaMill search engine [SWKS06, WSK$^+$06].

Anyway, although [Chr06] underlined difficulties in evaluating information visualisation interfaces, our further researches and developments on visualisation for VITALAS project will be guided by the use-cases and the users needs specified in the WP1 deliverable D1.1 (Use cases and Users requirements) of the VITALAS project and will be evaluated by the VITALAS end-users.

References

[AMA05] D. Archambault, T. Munzner, and D. Auber. Topolayout: Multi-level graph layout by topological features. *IEEE Transactions on Visualization and Computer Graphics*, 13(2):305–317, 2005.

[AMA07] D. Archambault, T. Munzner, and D. Auber. Grouse : Feature-based, steerable graph hierarchy exploration. In *Eurographics / IEEE VGTC Symposium on Visualization (EuroVis 07)*, pages 67–74, 2007.

[Bat02] M.J. Bates. Speculations on browsing, directed searching, and linking in relation to the Bradford distribution. In *Fourth International Conference on Conceptions of Library and Information Science (CoLIS 4)*, pages 137–150, 2002.

[BETT99] G. Di Battista, P. Eades, R. Tamassia, and I.G. Tollis. *Graph Drawing; Algorithms for the Visualization of Graphs*. Prentice Hall, Upper Saddle River, NJ, USA, 1999.

[BH86] J. Barnes and P. Hut. A hierarchical O(n logn) force-calculation algorithm. *Nature*, 1986.

[BP06] U. Brandes and C. Pich. Eigensolver methods for progressive multidimensional scaling of large data. In *14th International Symposium on Graph Drawing*, Karlsruhe, Germany, September 2006.

[CB06] L. Chen and A. Buja. Local multidimensional scaling: a nonlinear dimension reduction method for data visualization. In *Joint Statistical Meetings, JSM 2006*, Seattle, USA, August 2006.

[CC05] M. G. Christel and R. M. Conescu. Addressing the challenge of visual information access from digital image and video libraries. In *5th ACM/IEEE-CS joint conference on Digital libraries (JCDL '05)*, pages 69–78, New York, NY, USA, 2005. ACM Press.

[CGL04] P. Chiu, A. Girgensohn, and Q. Liu. Stained-glass visualization for highly condensed video summaries. In *IEEE International Conference on Multimedia and Expo (ICME 2004)*, June 2004.

[Cha96] M. Chalmers. A linear iteration time layout algorithm for visualising high-dimensional data. In *7th conference on Visualization '(VIS '96)*. IEEE Computer Society Press, 1996.

[Che97] C. Chen. Structuring and visualising the www by generalised similarity analysis. In *Hypertext*, pages 177–186, 1997.

[Chr06] M. Christel. Evaluation and user studies with respect to video summarization and browsing. In *Multimedia Content Analysis, Management and Retrieval 2006, Proceedings of IS&T/SPIE Symposium on Electronic Imaging*, volume 6073, San Jose, CA, 2006.

[CKGB02] P. Cano, M. Kaltenbrunner, F. Gouyon, and E Battle. On the use of fastmap for audio retrieval and browsing. In *International Symposium on Music Information Retrieval (ISMIR '02)*, Paris, France, 2002.

[CLM05a] M. Campanella, R. Leonardi, and P. Migliorati. Future viewer: an efficient framework for navigating and classifying audiovisual documents. In *WIAMIS '05*, Montreax, Switzerland, April 2005.

[CLM05b] M. Campanella, R. Leonardi, and P. Migliorati. An intuitive graphic environment for navigation and classification of multimedia documents. In *ICME'05*, Amsterdam, The Netherlands, July 2005.

[CP95] L. D. Catledge and J. E. Pitkow. Characterizing browsing strategies in the world-wide web. In *Third International World-Wide Web conference on Technology, tools and applications*, pages 1065–1073, New York, NY, USA, 1995. Elsevier North-Holland, Inc.

[CS06] G. Ciocca and R. Schettini. Dynamic storyboards for video content summarization. In *8th ACM international workshop on Multimedia information retrieval (MIR '06)*, pages 259–268, New York, NY, USA, 2006. ACM Press.

[DC03] G. Daniel and M. Chen. Video visualization. In *14th IEEE Visualization 2003 (VIS'03)*, page 54, Washington, DC, USA, 2003. IEEE Computer Society.

[Den07] D. Deng. Content-based image collection summarization and comparison using self-organizing maps. *Pattern Recognition*, 40(2):718–727, February 2007.

[DH97] P. Demartines and J. Herault. Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets. *IEEE Transactions on Neural Networks*, 8(1):148–154, 1997.

[DH06] A. Don and N. Hanusse. A deterministic multidimensional scaling algorithm for data visualisation. In *10th International Conference on Information Visualisation (IV'06)*, pages 511–520, June 2006.

[Don06] A. Don. *Indexation et navigation dans les contenus visuels : approches basées sur les graphes*. PhD thesis, Université de Bordeaux, France, December 2006.

[DZP04] D. Deng, J. Zhang, and M. Purvis. Visualisation and comparison of image collections based on self-organised maps. In *Australasian Workshop on Data Mining and Web Intelligence (DMSW 2004)*, 2004.

[Ead84] P. Eades. A heuristic for graph drawing. *Congressus Numerantium*, 42:149–160, 1984.

[FL95] C. Faloutsos and K.-I. Lin. Fastmap: a fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets. In *ACM SIGMOD international conference on Management of data SIGMOD '95*, pages 163–174, New York, NY, USA, 1995. ACM Press.

[FLM00] S. S. Fels, E. Lee, and K. Mase. Techniques for interactive video cubism. In *Proceedings of ACM Multimedia*, pages 368–370, Oct 2000.

[FR91] T.M. J. Fruchterman and E.M. Reingold. Graph drawing by force-directed placement. *Software - Practice and Experience*, 21(11):1129–1164, 1991.

[FT07] Y. Frishman and A. Tal. Online dynamic graph drawing. In *IEEE VGTC Symposium on Visualization (EuroVis 07)*, 2007.

[GCSS06] D. B. Goldman, B. Curless, S. M. Seitz, and D. Salesin. Schematic storyboarding for video visualization and editing. *ACM Transactions on Graphics (Proc. SIGGRAPH 2006)*, 25(3), July 2006.

[Gir03] A. Girgensohn. A fast layout algorithm for visual video summaries. In *IEEE International Conference on Multimedia and Expo (ICME'03)*, pages 77–80, July 2003.

[HHM+04] D. Heesch, P. Howarth, J. Magalhães, A. May, M. Pickering, A. Yavlinsky, and S. M. Rüger. Video retrieval using search and browsing. In *TREC Video Retrieval Evaluation (TRECVID)*, Gaithersburg, USA, November 2004.

[HJ04] S. Hachul and M. Jünger. Drawing large graphs with a potential-field-based multilevel algorithm. In *Graph Drawing (GD'04)*, pages 285–295, 2004.

[HJ05] W. Hurst and P. Jarvers. Interactive, dynamic video browsing with the zoomslider interface. In *IEEE International Conference on Multimedia and Expo (ICME'05)*, 2005.

[HK02] D. Harel and Y. Koren. A fast multi-scale algorithm for drawing large graphs. *J. Graph Algorithms Appl.*, 6(3):179–202, 2002.

[HMM00] 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.

[HPRY03] D. Heesch, M. Pickering, S. M. Rüger, and A. Yavlinsky. Video retrieval using search and browsing with key frames. In *TREC Video Retrieval Evaluation (TRECVID)*, Gaithersburg, USA, November 2003.

[HR04a] D. Heesch and S. M. Rüger. NN*k* networks for content-based image retrieval. In *European conference on information retrieval (ECIR 2004)*, pages 253–266, Sunderland, England, 2004. Springer LNCS 2997.

[HR04b] D. Heesch and S. M. Rüger. Three interfaces for content-based access to image collections. In *International Conf on Image and Video Retrieval (CIVR '04)*, pages 491–499, Dublin, Ireland, July 2004. Springer LNCS 3115.

[HR05] D. Heesch and S. M. Rüger. Image browsing: Semantic analysis of NNk networks. In *International Conf on Image and Video Retrieval (CIVR '05)*, pages 609–618, Singapore, July 2005. Springer LNCS 3568.

[ISU+07] T. Iwata, K. Saito, N. Ueda, S. Stromsten, T. L. T. L. Griffiths, and J. B. Tenenbaum. Parametric embedding for class visualization. *Neural Computation*, 2007.

[JLZH07] S. Jiang, H. Liu, Z. Zhao, and Q. Huang. Generating video sequence from photo image for mobile screens by content analysis. In *IEEE International Conference on Multimedia and Expo, 2007. ICME 2007*, July 2007.

[JM04] F. Jourdan and G. Melancon. Multiscale hybrid MDS. In *International Conference on Information Visualisation (IV'04)*, pages 388–393, 2004.

[KCH03] Y. Koren, L. Carmel, and D. Harel. Drawing huge graphs by algebraic multigrid optimization. *Multiscale Modeling and Simulation*, 1(4):645–673, 2003.

[KK89] T. Kamada and S. Kawai. An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1):7–15, 1989.

[Koh95] T. Kohonen. *Self-Organizing Maps*. Series in Information Sciences, Vol. 30. Springer, Heidelberg, 1995.

[KS00] H. Kang and B. Shneiderman. Visualization methods for personal photo collections: Browsing and searching in the photofinder. In *IEEE International Conference on Multimedia and Expo (III)*, pages 1539–1542, 2000.

[KW01] M. Kaufmann and D. Wagner, editors. *Drawing Graphs: Methods and Models,*. LNCS 2025. Springer-Verlag, 2001.

[LFRS06] Hangzai Luo, Jianping Fan, William Ribarsky, and Shin'ichi Satoh. Exploring large-scale video news via interactive visualization. In *Visual Analytics Science And Technology, 2006 IEEE Symposium on*, pages 75–82, 2006.

[Li04] J. X. Li. Visualization of high dimensional data with relational perspective map. *Information Visualization*, 3(1):49–59, 2004.

[LXT+04] H. Liu, X. Xie, X. Tang, Z.-W. Li, and W.-Y. Ma. Effective browsing of web image search results. In *6th ACM SIGMM international workshop on Multimedia information retrieval (MIR '04)*, pages 84–90, New York, NY, USA, 2004. ACM Press.

[MC03] A. Morrison and M. Chalmers. Improving hybrid MDS with pivot-based searching. In *IEEE Symposium on Information Visualization*, 2003.

[MC04] A. Morrison and M. Chalmers. A pivot-based routine for improved parent-finding in hybrid MDS. *Information Visualisation*, 3(2):109–122, 2004.

[MS00] M. Markkula and E. Sormunen. End-user searching challenges indexing practices in the digital newspaper photo archive. *Information Retrieval*, 1(4):259–285, 2000.

[MS06] M. Markkula and E. Sormunen. Video needs at the different stages of television program making process. In *1st international conference on Information interaction in context (IIiX)*, pages 111–118, New York, NY, USA, 2006. ACM Press.

[MTT01a] B. Moghaddam, Q. Tian, and T.S.Huang. Spatial visualization for content-based image retrieval. In *IEEE International Conference on Multimedia and Expo, 2001. ICME 2001*, pages 162– 165, August 2001.

[MTT01b] B. Moghaddam, Q. Tian, and T.S.Huang. Visualization and layout for personal photo libraries. In *Content-Based Multimedia Indexing (CBMI'01)*, Brescia, Italy, September 2001.

[Ngu06] G.P. Nguyen. *Interactive Image Search using Similarity-Based Visualization.* PhD thesis, University of Amsterdam, December 2006.

[Noa04] 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.

[Noa06] A. Noack. Energy-based clustering of graphs with nonuniform degrees. In *14th International Symposium on Graph Drawing (GD 2006)*, pages 309–320, Limerick, Ireland, 2006.

[NR07] R. Neumayer and A. Rauber. Multi-modal music information retrieval - visualisation and evaluation of clusterings by both audio and lyrics. In *8th Conference Recherche d'Information Assistée par Ordinateur (RIAO'07)*, Pittsburgh, PA, USA, May 2007.

[NW05a] G.P. Nguyen and M. Worring. Scenario optimization for interactive category search. In *7th ACM SIGMM international workshop on Multimedia information retrieval (MIR'05)*, pages 159–166, New York, NY, USA, 2005. ACM Press.

[NW05b] G.P. Nguyen and M. Worring. Similarity based visualization of image collections. In *7th International Workshop of the EU Network of Excellence DELOS on AUDIO-VISUAL CONTENT AND INFORMATION VISUALIZATION IN DIGITAL LIBRARIES (AVIVDiLib'05)*, Cortona, Italy, May 2005.

[NW07a] G.P. Nguyen and M. Worring. Interactive access to large image collections using similarity based visualization. *Journal of Visual Languages and Computing*, 2007.

[NW07b] G.P. Nguyen and M. Worring. Optimization of interactive visual similarity based search. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, 2007.

[NWS06] G. P. Nguyen, M. Worring, and A. W.M. Smeulders. Similarity learning via dissimilarity space in cbir. In *8th ACM international workshop on Multimedia information retrieval (MIR'06)*, pages 107–116, New York, NY, USA, 2006. ACM Press.

[PDW03] E. Pampalk, S. Dixon, and G. Widmer. Exploring music collections by browsing different views. In *International conference on music information retrieval (ISMIR 2003)*, Washington, D.C., USA,, 2003.

[Pic04] M. J. Pickering. *Video retrieval and summarisation.* PhD thesis, Imperial College, London, UK, 2004.

[PJ05] Chris Pal and Nebojsa Jojic. Interactive montages of sprites for indexing and summarizing security video. In *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2*, page 1192, Washington, DC, USA, 2005. IEEE Computer Society.

[PR03] M. Pickering and S. Rüger. Anses: Summarisation of news video. In *International Conference on Image and Video Retrieval (CIVR 03)*, LNCS 2728, pages 425–434, Urbana-Champaign, July 2003. Springer-Verlag.

[RBSW01] K. Rodden, W. Basalaj, D. Sinclair, and K. Wood. Does organisation by similarity assist image browsing? In *ACM Conference on Human Factors in Computing Systems (ACM CHI 01)*, April 2001.

[Rod02] K. Rodden. *Evaluating Similarity-Based Visualisations as Interfaces for Image Browsing*. PhD thesis, University of Cambridge Computer Laboratory, Computer Laboratory, October 2002. Technical Report 543.

[RS00] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, December 2000.

[Sam69] J.W. Sammon. A non linear mapping for data structure analysis. *IEEE Transactions on Computers*, C-18(5):401–409, May 1969.

[SMR06] O. Samko, A.D. Marshall, and P.L. Rosin. Selection of the optimal parameter value for the isomap algorithm. *Pattern Recognition Letters*, 27(9):968–979, July 2006.

[SWKS06] C. Snoek, M. Worring, D. Koelma, and A.W.M. Smeulders. Learned lexicon-driven interactive video retrieval. In *International Conference on Image and Video Retrieval (CIVR 2006)*, volume LNCS 4071, pages 11–20, Tempe, AZ, USA, July 2006. Springer Berlin / Heidelberg.

[TAT97] Y. Taniguchi, A. Akutsu, and Y. Tonomura. Panoramaexcerpts: extracting and packing panoramas for video browsing. In *ACM MM*, pages 427–436, 1997.

[TdSL00] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, December 2000.

[TS06] Y. Takeuchi and M. Sugimoto. Video summarization using personal photo libraries. In *8th ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR'06)*, pages 213–222, New York, NY, USA, 2006. ACM Press.

[vHvW04] F. van Ham and J.J. van Wijk. Interactive visualization of small world graphs. In *IEEE Symposium on Information Visualization (INFOVIS'04)*, pages 199 – 206, 2004.

[Wal03] C. Walshaw. A multilevel algorithm for force-directed graph drawing. *J. Graph Algorithms Appl.*, 7(3):253–285, 2003.

[WMH07] T. Wang, T. Mei, and X.-S. Hua. Video collage: a novel presentation of video sequence. In *ICME'07*, Beijing, P. R China, July 2007.

[WSdR+07] M. Worring, C.G.M. Snoek, O. de Rooij, G.P. Nguyen, and W.M. Smeulders. The Mediamill semantic video search engine. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (invited paper)*, Honolulu, Hawaii, USA, 2007.

[WSK+06] M. Worring, C.G.M. Snoek, D.C. Koelma, G.P. Nguyen, and O. de Rooij. Lexicon-based browsers for searching in news video archives. In *International Conference on Pattern Recognition (ICPR 2006)*, pages 1256–1259, Los Alamitos, CA, USA, 2006. IEEE Computer Society.

[YFDH01] K.-P. Yee, D. Fisher, R. Dhamija, and M. Hearst. Animated exploration of dynamic graphs with radial layout. In *IEEE Symposium on Information Visualization 2001 (INFOVIS'01)*, pages 43–50, San Diego, USA, 2001.