SPATIAL VIDEO
Exploring Space Using Multiple Digital Videos

Edmundo M. N. Nobre[1] and Antonio S. Câmara[1]

[1]Environmental Systems Analysis Group, New University of Lisbon,
2825 Monte de Caparica, PORTUGAL
{en, asc}@mail.fct.unl.pt

Abstract. This paper presents and discusses a new approach to collect, organize and explore multiple digital videos on a spatial basis. We present a methodology based on direct video frame indexation to the real space represented in the video images. This approach facilitates the exploration of space through multiple videos. An illustrative application of this technology for multimedia spatial information systems is provided. Future developments related to the three-dimensional explorations of space using digital video are also discussed.

1 Introduction

Digital video captures both the spatial and temporal dimensions of macro-phenomena such as a forest fire or micro-phenomena such as the evolution of biological films.

Video capture is usually a linear process that involves a pointing, rotation or moving operation with a camera. To completely capture on video an open space, a 360 degrees rotation capture operation may be performed with satisfactory results if the area is small. Larger areas with heterogeneous topography will require more than one capture. Those captures may be probably performed with different techniques. As a result, the handling of the different video segments becomes a complex task.

Natural areas are irregular, with rivers, mountains and other natural or artificial obstacles that always occlude some relevant location. How can we see what is behind a hill?

The systematic collection of geo-referenced video segments solves this problem. Land is a mosaic of pieces, like a puzzle, where each piece represents some particular corner, an indoor scene or a large landscape. Each one of them can be documented with a different image or a video. Together they allow a complete visual perception of the space [1, 2].

This paper introduces an approach where pre-defined capture processes and subsequent spatial indexing of video are used to facilitate the exploration of a real world scene. An application to the development of a spatial multimedia information system is provided for illustrative purposes. Future developments related to the use of digital video in the exploration of three-dimensional representations are also discussed.

2 Introducing Spatial Video Systems

Modeling and composing complex synthetic 3D worlds have been explored and applied in fields ranging from scientific interactive visualization to gaming. Most of these techniques typically involve the construction of complex 3D models, which try to reproduce, as closest as possible, the reality they intent to re-create.

Digital video is a more realistic representation of the natural world. Video is less expensive and easier to produce, and incorporates both spatial and temporal dimension. Nevertheless, videos have been looked mainly as rigid data content linear structures with restrictive browsing capabilities (typically, forward or backward). Consequently, the use of video in the development of applications such as the mentioned above is still limited.

Methods to organize, process and visualize video information have been proposed by several authors [7, 8, 9, 10]. In the present approach, we propose an interactive way to explore video, which looks, not only to the video content itself but also to the context of the capture process.

The term "interactive video" is used in the sense that we can compose and travel along different videos and images in a natural way. The interaction is not with the video itself as a simple multimedia object as it is usually done. A real image or a schematic map may be used in the background for geographical reference.

In our system space may be understood as a mosaic of pieces, like in a puzzle, where each peace is represented by a video. These videos are flexible multimedia objects that can be shaped on space to build the needed pieces to fulfill the desirable area (Fig.1).

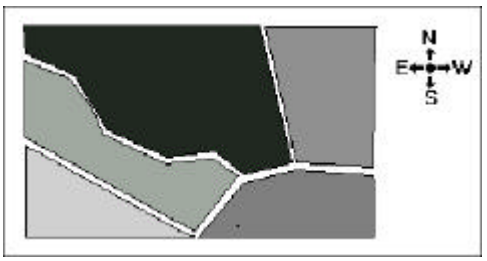

Fig.1. Space as a puzzle where each peace is a spatial video object.

A still image or a video frame collects visual information on a specific geographical area. These images can be naturally indexed to that area in a graphical way by drawing a representation of its spatial distribution (covering area) on a background map or image with a known geographical description.

Once we have enough video captures to cover the surrounding space, it is possible to use the video-space content as the background for several different visual experiments. We can use it as an exploratory tool to search, travel along, or visualize elements present on those videos. We can use it to compose geo-referenced outputs from a spatial simulation process (for example, to visualize a forest fire evolution projected on the video images of the forest). We can also use it to augment reality by superimposing virtual objects, like buildings, dams, or any other 3D structure on the video images.

3 Implementation

There are two main inter-related stages to develop spatial video information systems such as those proposed herein: The *Data Structuring Stage* and the *Exploration Stage.*

3.1 The Data Structuring Stage

The implementation process is based on a basic assumption: an image, whether alone or extracted from a video file, represents spatial data that can be geographically referenced.

The Background Image

A background image (typically a map or an aerial photo) describing the space content is used to facilitate the video spatial indexing process. This background image is also relevant to the *Exploration Stage*.

Composing Spatial Videos

To introduce a new video in the system, a line is drawn on the background surface (Fig. 2). Each drawn line represents an image (if a single image is selected) or a stream of images (when the selected file is a video file). For video files, this line also represents the path along which the video was captured (Fig. 2. (1)), and reproduces the video image stream.

In the implemented prototype, videos and images can be previously picked from a selection box and visualized using a display video window (see Fig. 6).

The Data Structure

Each video-frame image captures a section of a geographical space depending on the camera properties (zoom, lens) and the cameraperson abilities.

Dividing the line in a number of sections equal to the number of frames of the video, each section will represent one frame (Fig. 2. (3)). Selecting one of those sections on the background map will retrieve the corresponding video frame. The content of that frame can then be visualized on the video display window.

In ideal video capture conditions the central axis of each frame will coincide with a different section on the drawn line. These conditions depend on the camera's displacement. The capture should be done in a perpendicular way to the prevailing movement direction.

Automatic capture procedures following principles like the ones present in the moviemaps system [5, 6] would be ideal. Moviemps allows a complete frame flow control once videos are captured along pre-defined routes and filmed with a stop-

frame camera triggered by distance rather than by time. With this system, frames become automatically equally distributed over space and quite stables in terms of pointing orientation. A GPS system for automatic video-frame spatial indexation would also improve the data accuracy for spatial distribution.

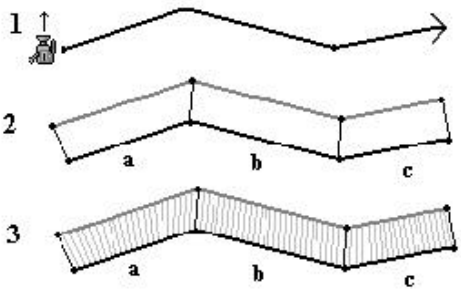

Fig. 2. (1) The video captures Path. (2) The main capture sections. (3) Automatic distribution of the video frames along the space.

Shaping the Video on Space: The Video Covered Area

While the background surface, where video and image-lines are displayed, should represent a top-bottom view, indexed videos usually represent horizontal captures. In a horizontal capture, the area covered by each image goes from the point where the capture camera is until the end line of the landscape or until some obstacle is reached (a mountain or a building). Applying this principle to each frame, we get the covered area for that video.

To define the covered area of a video we use a set of control points. Each point represents a different video-frame (control images). Manipulating those points it is possible to define the extension of the covered area along the video path (Fig.3). The covered area should go straight on from the capture path line until some occluding obstacle is reached (Fig. 4). Occluded areas should be captured in another video from the opposite side of the obstacles, until the entire scenario is covered.

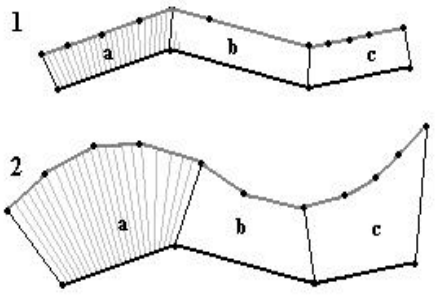

Fig. 3. Using the video line control-points to define the video coverage area.

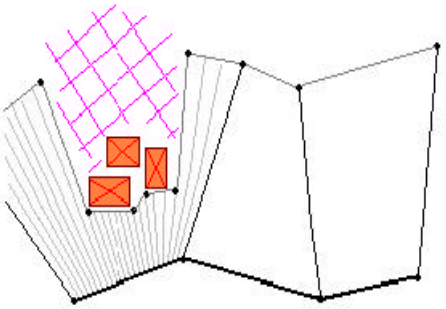

Fig. 4. The image covers exactly the area from the capture point to the existing buildings in front.

Spatial Video-Frame Corrections

Once ideal conditions are difficult to reproduce and sophisticated equipment is not always available, some manual corrections can be performed along the camera path line. These corrections are useful whenever captures are hand-made. In these instances, camera displacements from the perpendicular may occur often.

To perform these corrections, other control points can be introduced and manipulated in order to correct the frame distribution along the line.

When pointing to an object that is referenced in the background map and that is inside some video covered area, the corresponding frame for that point is automatically selected and displayed on the video display window. If the content of any selected video frame reports the same information that is referenced in the background control map, then we reach to the correct video frame distribution along the line (Fig. 5). If not, we have to readjust or introduce some new control points until we get a satisfactory correspondence. This test should be performed on several different points along the camera path until we get a satisfactory video frame distribution with coherent image content correspondence with the content of the background map.

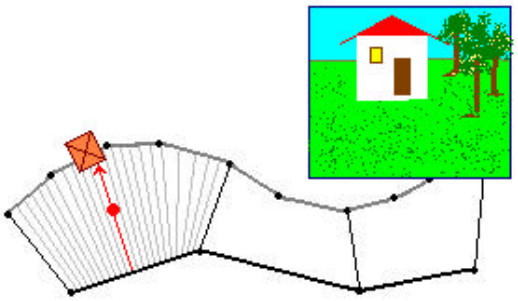

Fig. 5. Achieving the correct correspondence between the selected point in the background and the retrieved video frame for that point.

Automatic Capture Procedures

Ideal capture situations are those where human intervention is minimized, especially if that process allows automatic frame indexation and spatial distribution along the cameras capture path.

The ideal situation can be achieved in a continuo top-view capture from a camera on a plane and connected to a GPS devise. With the GPS data we get the exact position for each frame along the capture path. We can also clearly now the exact image covering range by reading the corresponding height coordinate from the GPS record. Fig.6 shows the projection of the frame distribution diagram resulting from a camera attach to a plane that is flying at different altitudes.

This diagram is slightly different from those represented before. Hear the covered image extends from the central point (the projection of the capture point on the path

line) to both sides. Nevertheless, the image axis is still represented by the perpendicular line to the cameras movement along the capture path.

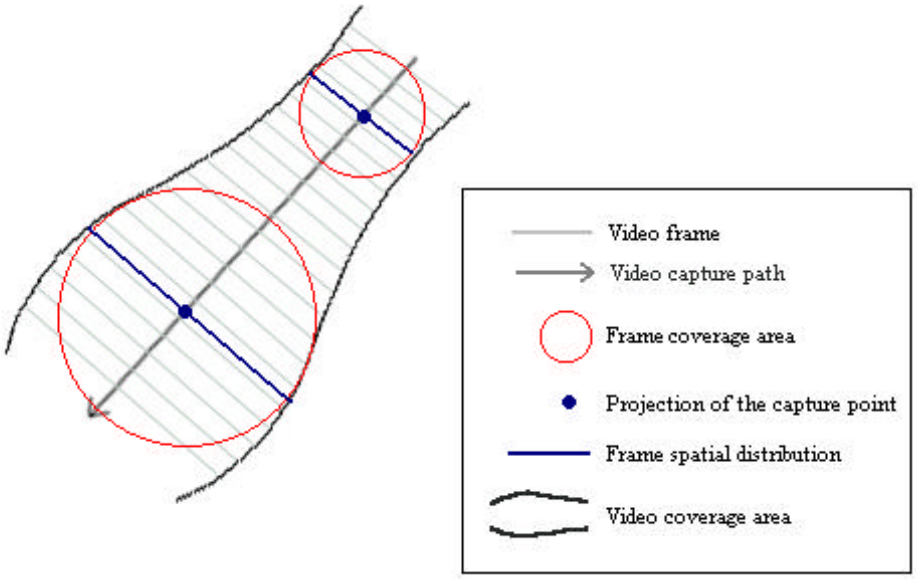

Fig. 6. – The video coverage area along the camera's path in a top-view captures.

3.2 The Exploration Stage

The data exploration stage is where users explore and interact with the visual data. With the presented structure, the data retrieval process is basically a natural consequence of the video spatial organization performed on the first stage.

Exploring Space

Once all videos have been spatially distributed over the background map, we just have to drag the mouse over the map to explore the space. For each selected point, the system will automatically retrieve the image or the video-frame indexed to that point and shows it on the video display window. To do that, the system will first detect to which video belongs the covered area we are pointing to. Once the video is found, the system uses the relative position of each selected point inside the video coverage area to retrieve the frame where those points belong.

If the aim is not a simple exploratory operation but the composition and visualization of some results from a spatial simulation process, the exploratory process is similar to the one described above, except for the point selection process. Now, the selected point is retrieved, not by a mouse 'click' on the map, but, for example, from some selected simulation object which position depends on the simulation evolution.

Overlapping Videos

When traveling on the "information surface" (the background maps with video-lines drawn on it), we may cross over points where several visual records have been introduced – basically, this means that the same point was captured by different videos. The system keeps the reference of which video has been tracked before reaching an area covered by overlapping videos.. If the tracked video is one of the overlapping ones, the system just keeps on it. In this way, it is possible to following a video documented track, even if it is constantly crossed with other existing visual information for the same points. This property allows a smooth traveling experience even in systems with large amounts of overlapping data.

Traveling on Space and Time

It is easy to imagine a collection of pictures or videos from the same place but collected at different hours, days or years. The slider on the Spatial Analysis on Multiple Video System prototype may be used to travel on videos and images that have the same space reference but correspond to different time captures. This capability also allows selecting a time interval and traveling in space only through videos that fit that interval.

4 Browsing on Interactive Digital Video: An Illustrative Example

A computer prototype applying this methodology was implemented and tested on a visual database application as a tool for video and image spatial exploration. Fig.7 reproduces the interface of that prototype where five main components may be identified. These components are the *video selection box*, the *video display box*, the *drawing toolbox*, the *drawing surface* and the *command bar.*

A background image (for geographical reference) is imported using the *Import image* option from the bottom's *command bar.* Videos are selected, one at a time, from the *video box* and, using the *draw* option from the *toolbox*, the corresponding line is drawn on the *drawing surface* following some specific path on the background image. The area covered by each video can now be adjusted to the background image by moving the control points that are displayed with the drown video-line (see Fig. 8). The stronger lines are the capture paths and the related thin lines are the correspondent control point lines. Each color corresponds to an individual video capture. It is possible to explore the area using the *explore* option. Pressing the left button of the mouse and moving it on the drawing surface does that. As we move the mouse, the image corresponding to each position is calculated and displayed on the *video display box*.

The proposed structure allows navigation facilities through different video files. In this example, if each side of a road is captured on a different video, users can travel directly from the video frame that covers the right side view to the video frame corresponding to the left side view just by crossing the road with the mouse.

For exploratory proposes the actual prototype version supports AVI and MPEG video files as well single BMP image files. For composition procedures of objects or simulation results on video images, AVI video files are required.

A continuous mouse movement on the background map retrieves a sequence of images from several different sources (still images and video images). These images can be composed in a new output video representing the selected exploratory path.

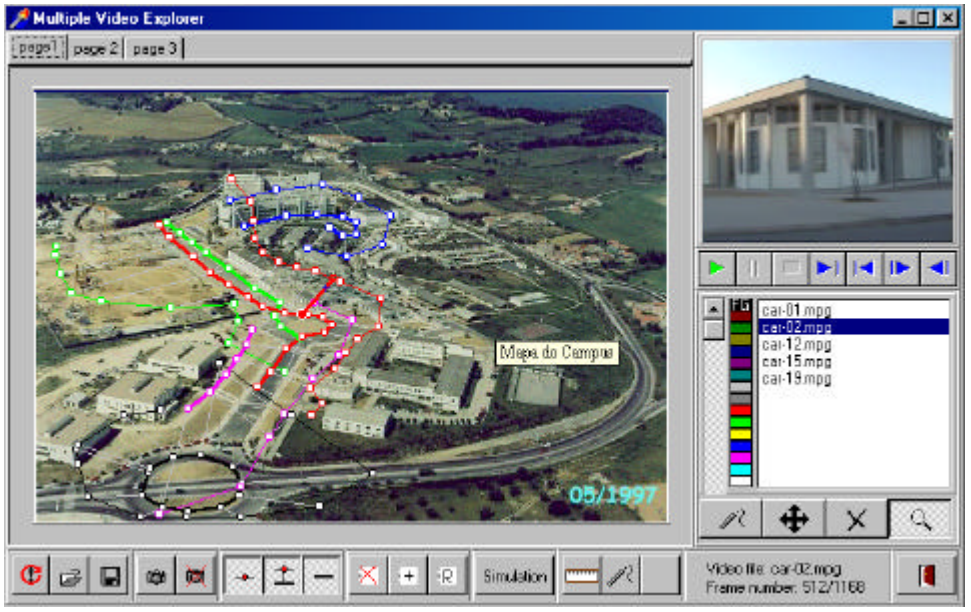

Fig. 7. Example of a spatial multiple video composition covering the College of Sciences and Technology of the New University of Lisbon, Portugal.

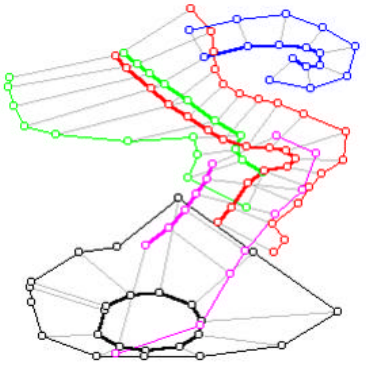

Fig. 8. Detail of the spatial distribution of video lines and their covering area shown on Fig. 7.

5 Guidelines to Future Work: 3D Modeling and Virtual Object Composition

Another application of this methodology is the ability to compose virtual objects on real video images. Having ensured the correct correspondence between the video-frame content and the schematic top-view 2D scenario, any virtual object placed on that scenario can be correctly projected on each video frame. Once we are able to freely travel along the video images trough the 2D representation of the area, we can explore virtual objects that have been placed on that scenario.

Capture-depth: the fourth dimension

Objects are usually geo-referenced using the three main coordinate axes (the X, Y and Z-axis). Time may be introduced as an additional coordinate to introduce temporal variations and allow time evolution. In the present system, to connect synthetic spatial data with real world images captured in a collection of digital videos, a new dimension is introduced. This dimension is the capture-depth of a point. The capture-depth of a point is the distance from the point's geographical position and the camera's position from where the video-frame that covers that area was captured.

This capture-depth information is crucial to perform the correct projection of a synthetic point in the corresponding video image. In the same way, by projecting all points describing any synthetic object in the current image, it is possible to compose complex 3D objects on the video stream to produce the desirable realistic output solution. When exploring the space, the capture-depth of each point has to be recalculated for each new position so we can explore different projections of the same object on images captured from different points of view (different images from the same video or from different videos).

Real environments (captured on the video images), whether they represent natural or humanized landscapes, usually present topographic irregularities that can vary from almost imperceptible variations to ones that are easily noticeable. For most situations, the introduction of height on the cameras capture path is fundamental to achieve the correct projection of virtual data in the video frames. The connection to a GIS or the use of a GPS may provide the additional topographic information required to obtain the correct camera's path.

The Viewing Pyramid: The perspective transformation

Any real world video frame picture is a projection of a 3D world into a 2D plane. Each point of the 2D plane corresponds to the precise mathematical projection of another point from the Real World environment. The formula for this projection depends on the prevailing viewing pyramid. The viewing pyramid defines the portion of eye-coordinate space, which the viewer can actually see [3, 4]. In a video capture situation it defines the portion of the camera-vision-coordinate space captured in a picture that depends mainly on the camera's capture conditions (zoom and lens).

Each set of camera conditions defines a different viewing pyramid. Keeping the same capture conditions, we get the same viewing pyramid for all the captured sequences.

According to Newman&Sproull [3] we can convert world coordinates into dimensionless fractions. As we probably do not know the camera conditions, a manual parameter to control those dimensionless fractions was implemented. The control ratio works as follows: an object is placed in the 2D background in a well known geographical position (for example, placing an object with the shape of a house in the top of the drawing where the house is represented). The system retrieves the image that covers the place and projects the object on it (in the central axis of the image). If the viewing pyramid is correct, the object will fit the house. If not, it will be displayed bigger or smaller than it should be. In this situation, we can move the manual control to up and down until the object fits the image of the house. When it fits, we reach the right value that characterizes this particular viewing pyramid.

Object Composition

If an object is placed in the 2D scenario in a point covered by some spatial video-line, the system will automatically retrieves the corresponding video frame that fits that point. It will also retrieve the distance between that point and the point from where the selected video frame was captured (the capture-depth). Knowing the capture-depth, the viewing pyramid (camera's capture conditions) and the frame related to the objects' position, it is possible to compose the object on the right position of that frame (which means, the right position on a real world picture) (Fig. 9).

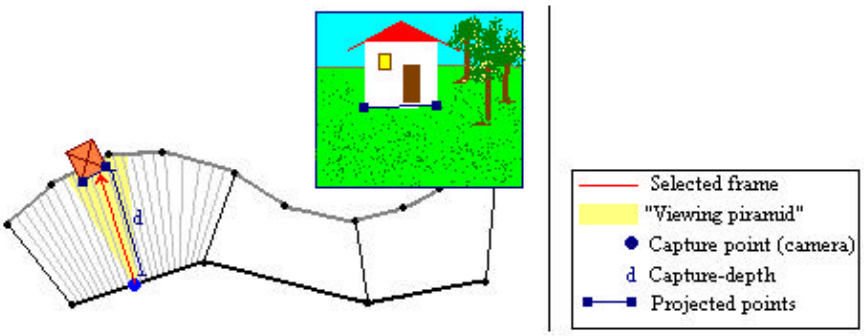

Fig. 9. Composing synthetic points on a video image.

If the object is a 3D model composed by a set of points, each one of them will present a different capture-depth. Applying the viewing pyramid calculated for that video, it is easy to project each object point in the selected image (the wire-frame). Thus, the object points will be projected in the image incorporating the required depth and the perspective corrections.

Once the object's wire-frame is projected on the image, the realistic representation of the object on the video image depends only on a set of operations to apply the correct textures, shading and other image effects procedures. These procedures well known computer graphic applications and will not be discussed here. The composition of different objects placed on the background scenario and projected in the same image is treated as discussed before. When dealing with different objects, knowing the exact position where they have been placed in the background scenario, it is also possible to simulate the occlusion of some objects by other objects and hidden-surface elimination.

Moving the object, or the viewing-point, a new frame will be selected as well as a new capture-depth for the new viewpoint and, consequently, a new perspective of the object will be composed on the selected frame.

6 Main Limitations

The main limitations of this system are related to the techniques involved in the video capture processes, particularly with respect to homogeneity and continuity. These are the most difficult issues, mainly if it is not possible to control the capture process.

The system will assume that all the video has been captured on a single shot and with a uniform movement camera. This means that its possible to access a frame representing some particular spatial position just knowing the corresponding fraction of the drawn line representing the video on that place. With a non-uniform movement of the camera during the capture (i.e. the cameramen moving faster, slower or paused), some delays or advances can occur and the selected image that was supposed to represent a particular point-of-view may correspond to another slightly different one. A possible solution to this problem is to pre-decompose each video into smaller uniform clips. A better solution is to use a stop-frame camera triggered by distance like in the Movimaps system.

Another video capture problem is related to the cameraperson's movements. Even if he or she moves at a constant speed, some rotation movement may be executed (as if it was moving in a car but moving the camera from one side to another). Again, this may result in no coordination between the geographical position and the resulting retrieved image for that point. Nevertheless, this problem can be easily solved using the system tools by compensating the camera's "abnormal" movement with the control points of the video-drowned line. This is a manual process that can be done using the video display to control and confirm the line adjustment.

7 Conclusions

The system proposed herein was developed to organize and display geo-referenced videos that cover an area of interest. Appropriate capture, retrieval and exploration processes are recommended. An illustrative example of a spatial video browser is presented. Future developments related to 3D modeling and composition of objects on spatial video area also discussed.

References

1. E.Nobre A.S.Câmara: Interactive Ecological Movies. (Internal Report), GASA, DCEA, FCT, UNL, Monte de Caparica, Portugal (1997).
2. E.Nobre A.S.Câmara: Programming by Reproduction. (Internal Report), GASA, DCEA, FCT, UNL, Monte de Caparica, Portugal (1998).
3. William M. Newman, Robert F. Sproull: Principles of Interactive Computer Graphics. McGraw-Hill, New York (1979).
4. J. Foley, A. van Dam, S. Feiner, J. Hughes: Computer Graphics: Principles and Practice. Second Edition in C, Addison-Wesley, Reading, MA (1995).
5. Lippman, A.: Movie-Maps: An Application of the Optical Videodisc to Computer Graphics. Proc. ACM SIGGRAPH (1980).
6. Naimark, M.: A 3D Moviemap and a 3D Panorama. SPIE Proceedings, Vol. 3012, San Jose (1997).
7. A.Gupta, R.Jain,: Visual Information Retrieval. Communications of the ACM, vol. 40, Nº5 (May 1997) pp.71-79.
8. R.Lienhart, S.Pfeiffer, W.Effelsberg: Video Abstracting. Communications of the ACM, vol. 40, Nº12, (December 1997)
9. B.-L. Yeo, M.M. Yeung,: Retrieving and Visualizing Video. Communications of the ACM /vol. 40, Nº12 (December 1997)
10. C.H. Chen, L.F. Pau and P.S.P. Wang, (ed.): Video Content Analysis and Retrieval. Hongjiang Zhang Handbook on Pattern Recognition and Computer Vision (1997)