VideoSOM: A SOM-Based Interface for Video Browsing

Thomas Bärecke[1], Ewa Kijak[1], Andreas Nürnberger[2], and Marcin Detyniecki[1]

[1] LIP6, Université Pierre et Marie Curie, Paris, France
thomas.baerecke@lip6.fr
[2] IWS, Otto-von-Guericke Universität, Magdeburg, Germany

Abstract. The VideoSOM sytem is a tool for content-based video navigation based on a growing self-organizing map. Our interface allows the user to browse the video content using simultaneously several perspectives, temporal as well as content-based representations of the video. Combined with the interaction possibilities between them this allows for efficient searching of relevant information in video content.

1 Introduction

The VideoSOM system performs structuring and visualization of video content [1]. It represents a video shot by a single keyframe and constructs higher level aggregates of shots. The user has the possibility to browse the content in several ways. The basic idea is to provide as much information as possible on a single screen, without overwhelming the user.

Before the information is visualized and thus the user can interact with the system, the following steps are performed. First, a shot boundary detection algorithm using a single threshold is applied. Then, each shot is described using its median frame as keyframe. Histograms are extracted for up to four different regions and merged together into a single vector. Finally, a growing self-organizing map algorithm [2,3], clusters the shots into groups ignoring the temporal aspect. The visualization is based on these groups and projects the temporal information on a time bar. Similar objects are linked with colours.

We combined elements providing information on three abstraction levels. First, there is an overview of the whole content provided by the self-organizing map window. On each cell, the most typical keyframe of the corresponding cluster is displayed. The second level consists of a combined content-based and time-based visualization. Furthermore, a list of shots is provided for each grid cell and a control derived from the time-bar control helps to identify content that is similar to the currently selected shot. We evalueted our system on news video from the TRECVID collection.

2 Walkthrough

This section is intended to introduce a chronological walk-through of the VideoSOM tool from the perspective of the user.

H. Sundaram et al. (Eds.): CIVR 2006, LNCS 4071, pp. 506–509, 2006.

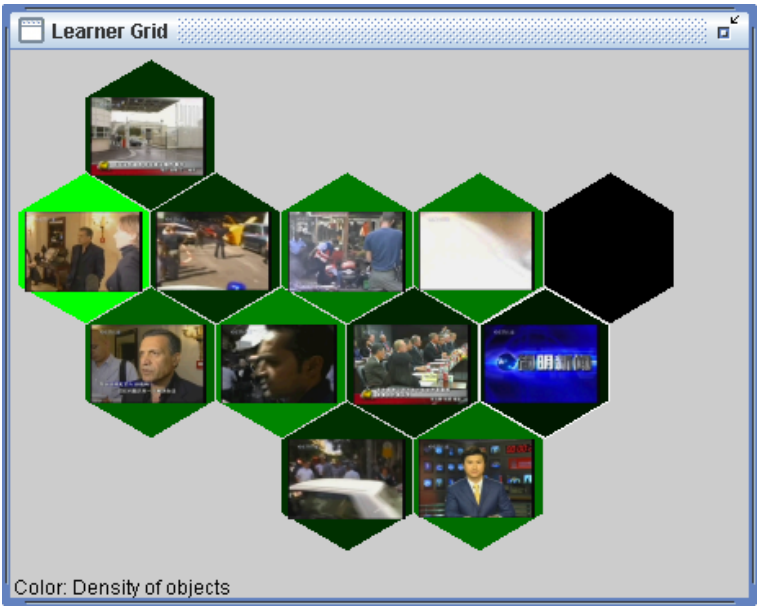

Fig. 1. Growing self-organizing map after learning

2.1 Opening and Pre-processing a Video

First, the data is loaded. This can be either the loading of a new raw video in a supported format, in which case the shot boundary detection and feature extraction steps have to be performed. Alternatively, a video including these information which was pre-processed in a former session can be opened. The user can change important variables of these steps, like the threshold for a shot boundary. Then, the the learning phase of the growing self-organizing map starts. The map has a hexagonal topology and can contain empty clusters. We usually start with a small grid size (2*2 or 3*3) and visualize the evolution of the learning process. The result of this step is illustrated in Fig. 2. The different shades of green indicate the density of shots in a certain cell, i.e. the number of shots assigned to it. The map can be retrained if the obtained clustering seems unsatisfactory.

2.2 Navigating Through a Video

A click on one cell (Fig. 1-1) opens the the corresponding shot list window in the interface. Simultaneously, the temporal position of the shots who are assigned to the chosen cell are projected in the form of black extensions on the time bar. After selecting one keyframe from the shot list (Fig. 1-2), the color of the cells in the map changes from green to shades of red. Now, the colour indicates the distance of the cells from the currently selected shot. Cells being very similar to the selected shot are coloured in dark red while cells being less similar are coloured with a brighter red. A main advantage of self-organizing

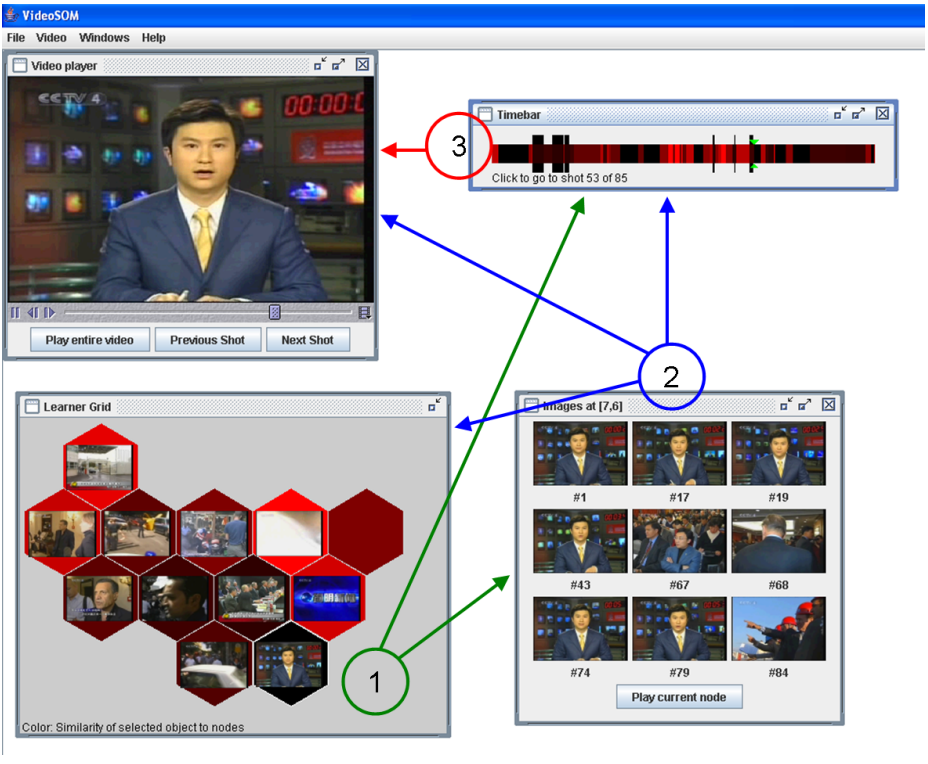

Fig. 2. User Interface with typical interactions. (1) Selection of a cell from the SOM. (2) Selection of a specific shot. (3) Selection of a temporal position.

maps is the fact that objects that are assigned to cells close to each other, in the low-dimensional space, are also close to each other in the high-dimensional space. But this does not mean that objects with a small distance in the high-dimensional space are necessarily assigned to cells separated by a small distance on the map. However, we overcome this problem with the visualisation schema presented above, starting with a specific shot, we will easily identify similar shots in dark red nodes. This improves significantly the navigation possibilities compared to the navigation support provided by other clustering schemas.

From user interaction perspective the map is limited to the following actions (Fig. 1-1): select nodes and communicate cluster assignment and colour information to the time bar. Nevertheless it is a very powerful tool which is especially useful for presenting a structured summarization of the video to the user.

The time bar changes its color synchronously with the map and visualizes the same colors for each shot. Thus, it provides a temporal view of similar keyframes. Furthermore, we added black bars at the positions where the shots of the currently selected cluster are located. The current shot is also played in the video player window. Simultaneously, more shot lists can be obtained by clicking on another cell in the map. The play current node operation merges all shots

from the current node into one single video sequence and plays it. Clicking once on the time bar plays the shot at the given position (Fig. 1-3). A double click forces the system to change the currently selected shot resulting in renewing the distances displayed in the self-organizing map and the time bar. In fact, this corresponds to selecting a shot from the shot list but we do not necessarily have to know in which cell the shot is located. Furthermore, the black bars indicating the shots assigned to the currently selected cell are adjusted.

3 System Requirements

VideoSOM is implemented in Java and was tested under Microsoft Windows XP as well as Mandrake Linux operating system. Apart from the Java Virtual Machine, it requires the Java Media Framework (JMF including the mp3plugin) and Java Advanced Imaging (JAI) libraries installed. Although the application itself is platform independent, we recommend to run it under MS Windows using the appropriate Windows Performance Pack versions of these libraries, since the Linux and cross-platform versions do not implement all features, especially the variety of implemented video codecs is reduced significantly. Consequently, all video codecs supported by the libraries can be loaded into VideoSOM. There are no specific hardware requirements, i.e. a standard personal computer is sufficient.

References

1. Bärecke, T., Kijak, E., Nürnberger, A., Detyniecki, M.: Video navigation based on self-organizing maps. In: to appear in Proc. of Int. Conference on Image and Video Retrieval (CIVR 2006), Springer (2006)
2. Kohonen, T.: Self-Organizing Maps. Springer-Verlag, Berlin Heidelberg (1995)
3. Nürnberger, A., Detyniecki, M.: Visualizing changes in data collections using growing self-organizing maps. In: Proc. of Int. Joint Conference on Neural Networks (IJCANN 2002), IEEE (2002) 1912–1917