Visual Table of Contents
Automatic Structure Detection for Videos

Books have detailed tables of content that enable you to go directly to the chapter or subchapter you’re looking for without leafing through the whole volume. Yet thus far there are no comparable indexes or directories for videos in the Internet or TV recordings: searching for particular sequences or scenes is usually a time-consuming, laborious process. Within the framework of the THESEUS research program – New Technologies for the Internet of Services – the Fraunhofer Heinrich Hertz Institute has developed a patented software module for recognition of video structure that segments videos into single scenes, shots and sub-shots. Such segmentation enables rapid and accurate detection of video sequences.

Benefits
- Visual table of contents
- Exact frame-target access to disparate semantic video segments
- Browsing & editing functionalities
- Streamlined video searching through temporally defined video segments
- Foundation for further video annotation and archiving

Challenges
At present, searching for a particular video sequence in video material is extremely complicated and time-consuming. The user must know the exact location of the sequence; otherwise there is no alternative to a laborious manual search through the whole video. The challenge is how to find a certain scene, particular shots or smaller visual unit rapidly and without any complicated search techniques.

Supported by
Federal Ministry of Economics and Technology
THESEUS

Visual Table of Contents
The software module automatically generates metadata which describe the boundaries between scenes, shots and sub-shots. This means that the module is capable of identifying the structure of the video. Shot detection is not just limited to identification of hard cuts but can also register gradual transitions like fades, dissolves and wipes. The software is significantly faster than real-time, and offers both C and C++ interfaces for easy integration in existing software products.

Hierarchical video structure

**Technical Background**

The software module specially developed for video structure detection is a dynamic link library which can recognize the temporal units of a video like scenes, shots and sub-shots. Scenes are the largest units; they could be interviews or a scene on the beach. Shots are the next smaller unit, consisting of video frames continuously recorded with a single camera operation. If these shots include numerous movements of the camera or objects, the visual content of the shot is modified. Sub-shots are the smallest visually consistent unit within a single shot.