



Video and XML

ã 2002. Visual Century Research, S.L. All rights reserved.

The information contained in this document may not be distributed, modified, reproduced in whole or in part without the prior written permission of Visual Century Research, S.L.

Lineal and hierarchic video concepts

When a series of sequential pictures, very similar between them, is shown to the human eye at sufficient speed, they are not perceived as separate images but as a smoothly moving animation. This is the basis for film and video. The cinema and European televisions use 25 frames per second to create this illusion, whereas in the United States the frame rate used is 30 per second.

If we observe the structure of a video we will see that it is usually formed by several fragments that correspond to different camera shots. Shots are the result of the continuous operation of a camera. The images in a shot are usually very similar, so it is possible to obtain a simple representation of the whole shot using a single image, or key frame.



Fig 1: Key frame extraction from the shots in a video

If we define a scene as a set of consecutive shots with some kind of logic relationship between, we can say that a video is made of scenes, that is, parts or fragments with semantic relevance. Some examples of scene are a piece of news, an advertisement or an interview. Therefore, it is possible to hierarchically represent a video with scenes, scenes with shots, and shots with key frames.



FIG 2: Sequential key frames

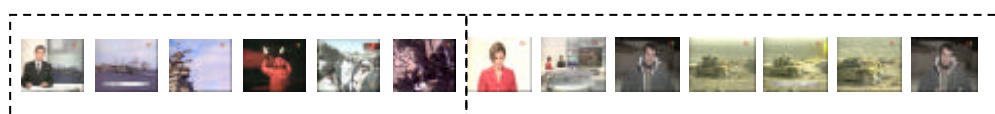


Fig 3: Grouping of key frames into scenes

Although video images are purely sequential in time, they can be structured hierarchically in the shape of a tree. This structure considers the shots to be the smaller units and builds up scenes from them. This way, scenes are generated as groups of shots with a logic relationship between them. And in the same way as scenes are built from shots, we can say that a video document is formed by several scenes.

If we take the example of a news program, we can represent the whole video with a representative image or key frame, such as the clock that appears before the news program begins. Each piece of news forms a scene, that is, a content unit with its own meaning, and can again be represented with an image. Finally, within each piece of news, we have a number of key frames, and we can see what it is about just by looking at them. The resulting hierarchical structure will look similar to this one:

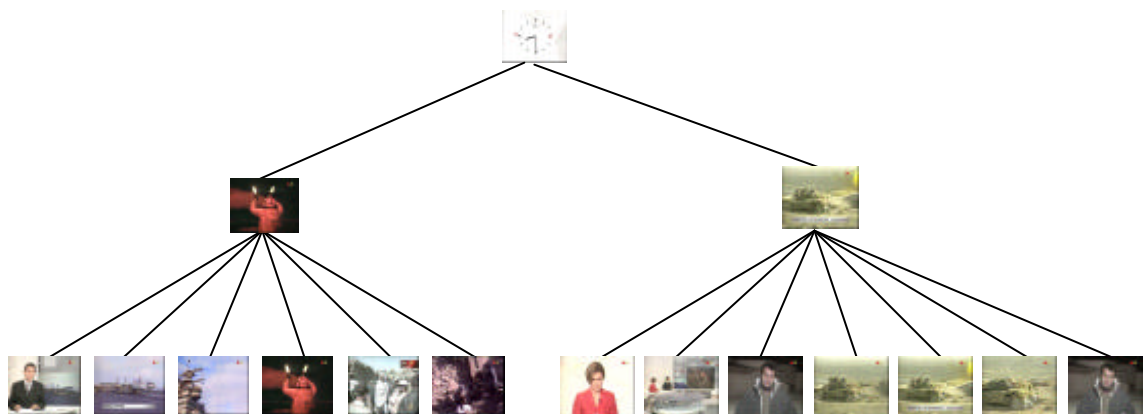


Fig 4: Example of hierarchical video structure

Video and XML

The most natural way to store the hierarchical video structure seen above is by using XML, or eXtensive Markup Language. XML gives total flexibility to build hierarchical structures and annotate all the information concerning the video. During the annotation process the key frames are used to structure the video in a tree-like shape. Video content is then annotated, and the generated metadata stored in XML format in with the corresponding metadata. The digitized video is also stored in a separate video repository.

Having all metadata in XML format facilitates the search of information in the database, making it quicker and more efficient. XML format also offers the possibility of customizing and personalizing search results as well as the ways to present the information, adapting them to the needs and desires of the client.

The storage of information in XML format allows the distribution of video content via a number of different channels. Once the material has been digitized, analyzed and annotated, thanks to the flexibility of XML format it can be reused and published in many different ways. This applies not only to the channels in existence today, but also to the ones that may appear in the future.