

8.3 Deriving Video Semantics

8.3.1 Cut Detection

A very simple and very basic scheme for the investigation of video semantics is **cut detection**. A **cut** denotes the border between two **shots** in a movie. During a shot the camera films continuously.

One distinguishes **hard cuts** and **soft cuts** (fade out, fade in, dissolve,....).

The most important use of cut detection is to decompose a video into smaller units which can then be analyzed. A sample application is the use of shots as the “atomic” units for storage and retrieval in a video archive.

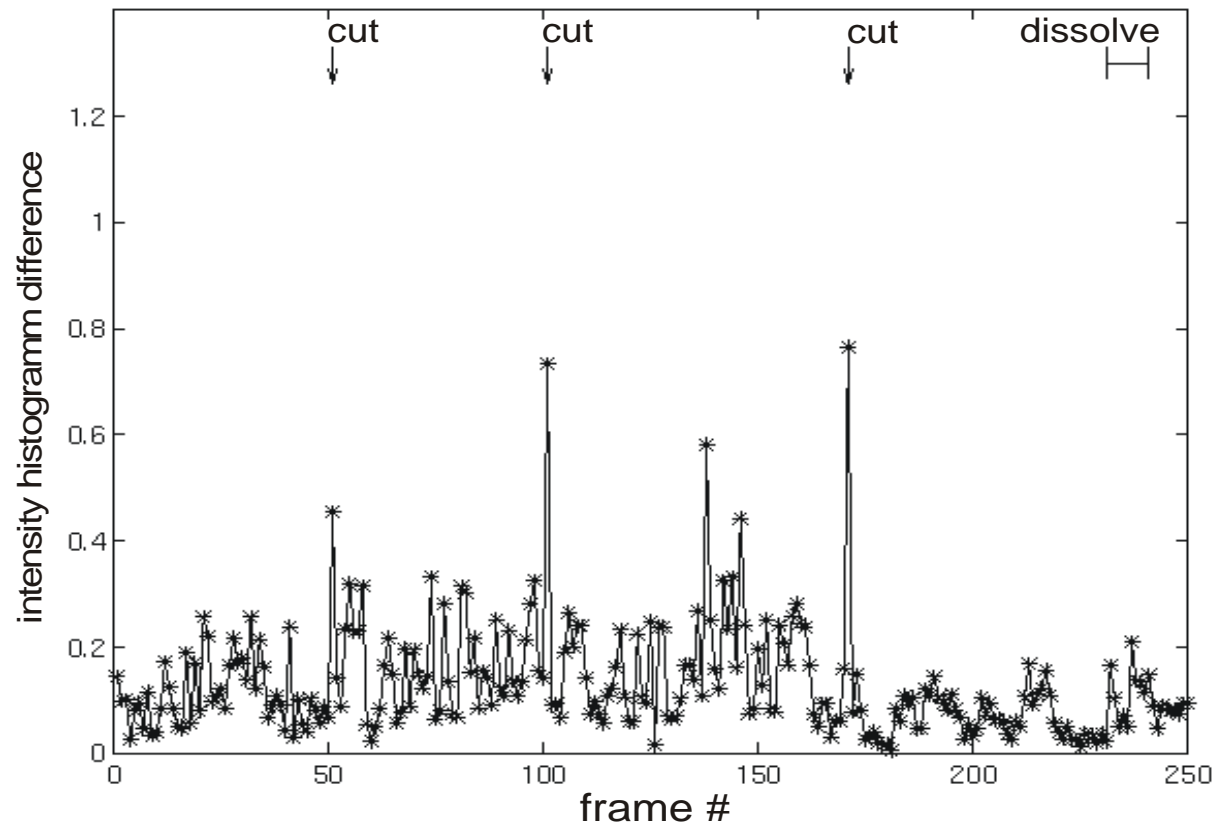
Cut Detection with Color Histograms

The simplest approach to cut detection is by means of color histograms: If the color histograms of two successive video frames i and $i+1$ differ at least by a threshold value T , a hard cut is detected.

Let $H(r,g,b,i)$ be the value of the histogram for a color triple (r,g,b) in a frame i , i.e., the number of pixels in frame i that have color (r,g,b) . A cut is detected if and only if:

$$\sum_{r,g,b} (H(r,g,b,i) - H(r,g,b,i+1))^2 \geq T$$

Example: Cut Detection with Gray-scale Histogram Differences



Typical Detection Errors

If we use color histogram differences, the rate of successful detection of hard cuts is 90% to 98% for a typical video.

This method, however, fails when the colors change significantly between two adjacent frames even if no cut is present. **False hits are much more common than misses.**

Examples

- the light is switched on in a room
- an explosion occurs
- the camera pans very quickly.

Cut Detection with the Edge Change Ratio

Generally, the edges in the first frame after a cut differ largely from the edges in the last frame before the cut. Thus the ECR can be used to detect hard cuts.

Let ECR_i be the edge change ration between frame i and frame $i+1$. A cut is detected if and only if

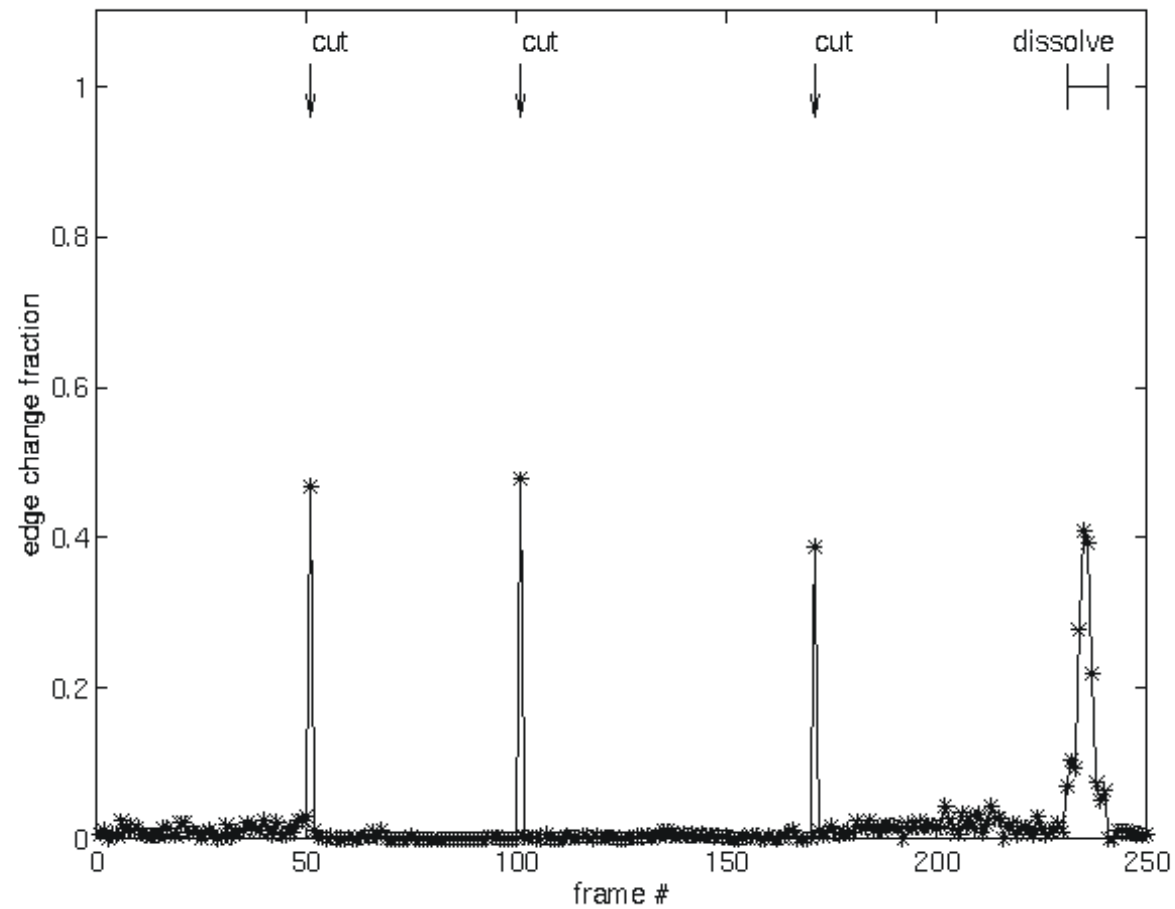
$$ECR_i \geq T$$

where T is a threshold.

A drawback of this approach is that it requires the computation of motion compensation for a video: fast camera panning or zooming or fast object movements lead to a high ECR for successive frames even if there is no cut.

Object motion or camera operations can be distinguished from a hard cut since they always last for several frames.

Example: Cut Detection with ECR

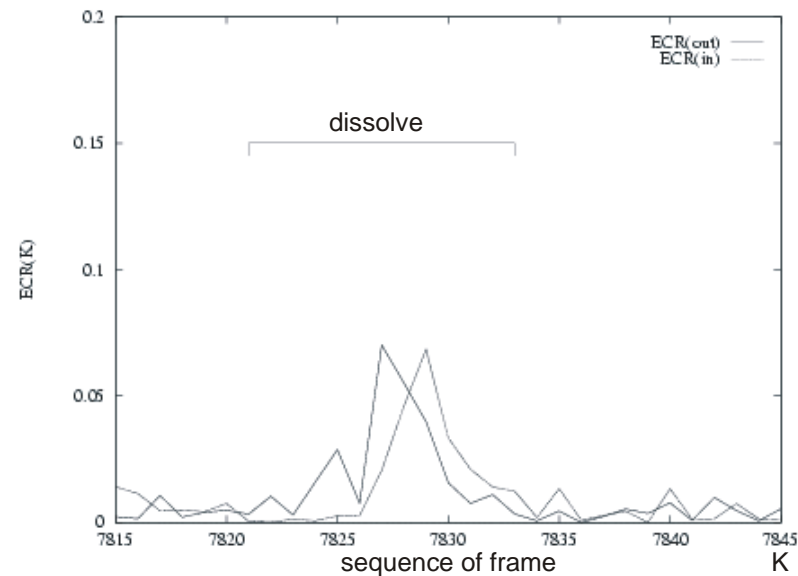


Detecting Soft Cuts

A fade between successive scenes is much harder to detect than a hard cut. One possibility is to try to detect a characteristic behaviour of the **edge change ratio** (ECR) in the area of fade outs, fade ins and dissolves.

Example:

During a **dissolve**, the edges of the old shot disappear with a constant ratio. At the same time, the edges of the new shot gradually appear. We observe a characteristic pattern of the ECR:

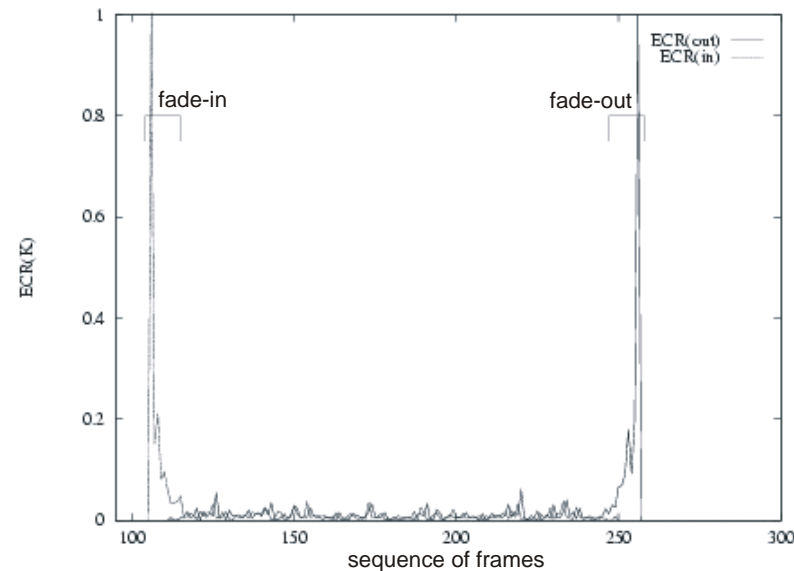


Fade-ins and Fade-outs (1)

In a similar fashion, fade-ins and fade-outs in a video can be determined: When fading out, the number of pixels located on an edge will be zero in the last frame of the fade. When fading in, the number of pixels located on an edge is zero in the first frame.

Example:

ECR during a fade-in and a fade-out.



Could We Detect Soft Cuts With Color Histograms?

It is obvious that soft transitions between **color histograms** are frequent in a video, and can have many reasons. Thus, it is practically impossible to detect soft cuts with color histogram differences.

8.3.2 Action Intensity

The intensity of action in a video shot is an interesting parameter. For example, it can be used to distinguish between different genres of TV broadcasts, such as newscasts vs. music clips.

Action intensity can easily be estimated by means of **motion vectors**: one computes the average absolute value of all vectors in each frame of the shot; this includes both the motion of objects within the scene and camera operations. If the video was compressed using motion vectors, this physical parameter can be computed with minimum cost.

The **edge change ratio** ECR can also be used as an indicator for action: long and relatively static shots have a low ECR whereas shots with a lot of motion have a high ECR.

8.3.3 Detection of Camera Operations

The notion of a camera operation encompasses panning, zooming, etc. It is characteristic for camera operations that they affect all pixels of a frame in a similar, deterministic manner.

Example 1

When the camera pans, all pixels move into the same direction, by the same amount, between two frames.

Example 2

When the camera zooms in, all pixels (except the one in the center of the frame) move from the center to the boundaries.

Algorithm for the Detection of Camera Operations

Algorithm Detect-Camera-Operations

- Use the motion vectors of the underlying compression algorithm (e.g., MPEG-2), or compute the optical flow for a sequence of frames.
- Determine if the direction and the length of the motion vectors fit to the scheme of a pre-defined camera operation.

The detection of camera operations works well when the shot under consideration contains no or little object motion. When camera operation and object motion appear simultaneously (which is quite common in practice), automatic detection becomes very difficult.

8.3.4 Text Detection

Goal

Extraction of text appearing in a video. Motivation: text is rich in semantics.

We distinguish between artificially generated text (such as a movie title) and text appearing in a scene.

Algorithm

- Detect regions of the frame containing text,
- cut them out,
- run an OCR algorithm (optical character recognition).

Text Detection(2)

Characteristics of generated text in a video

- monochrome
- rigid
- in the foreground
- the letters have a minimum and maximum size
- either stationary or moving linearly
- high contrast to background
- appears repeatedly in successive frames
- letters appear in groups.

Text Segmentation (1)

Text is monochrome



original frame of a video

segmentation into regions

Text Segmentation (2)

Apply thresholds for the size of the letters and the contrast.



previous result

after thresholding

Text Segmentation (3)

The text must be either stationary or moving linearly (here: horizontally).



previous result

after application of
the rule of horizontal
motion

Experimental Results

Detection of text regions in a video

test video	number of frames	number of letters	successful findings
title sequences	7372	6423	99%
advertisements	6858	1065	99%
newscast	18624	1411	97%

(from the dissertation of Rainer Lienhart, Praktische Informatik IV, U. Mannheim)

8.3.5 Face Detection

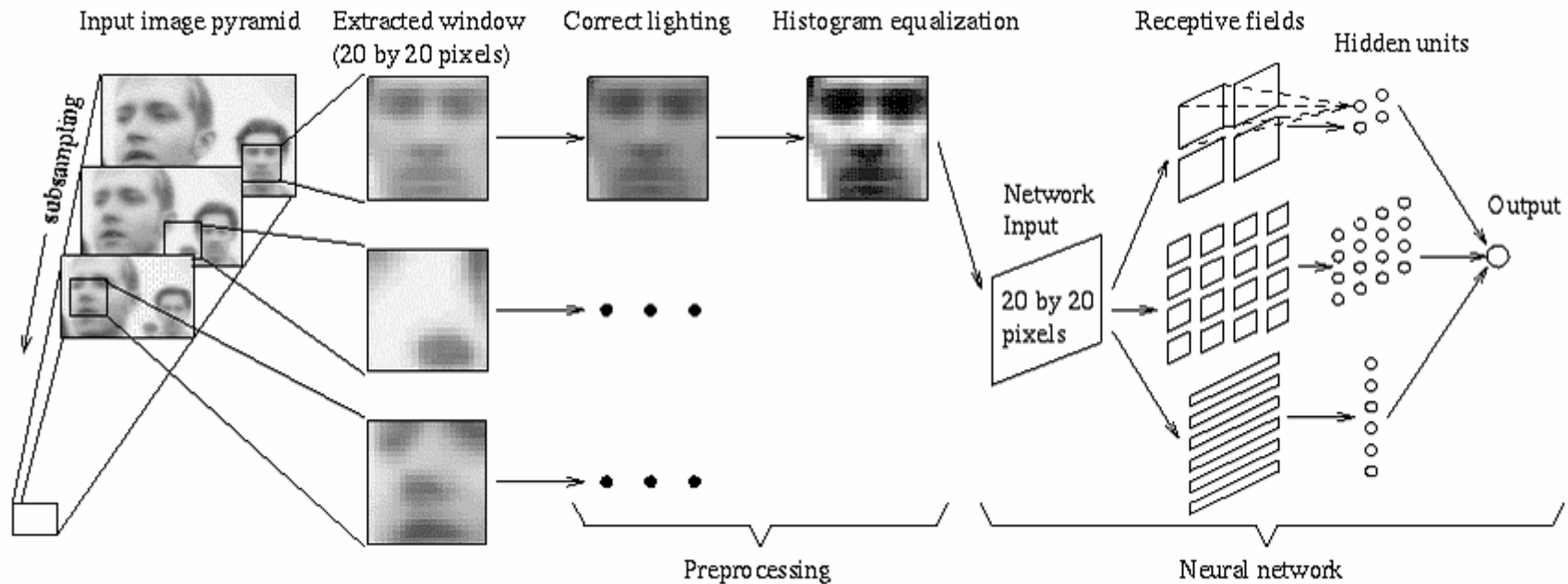
Goal

Detection of areas in a video frame that show the frontal view of a human face.

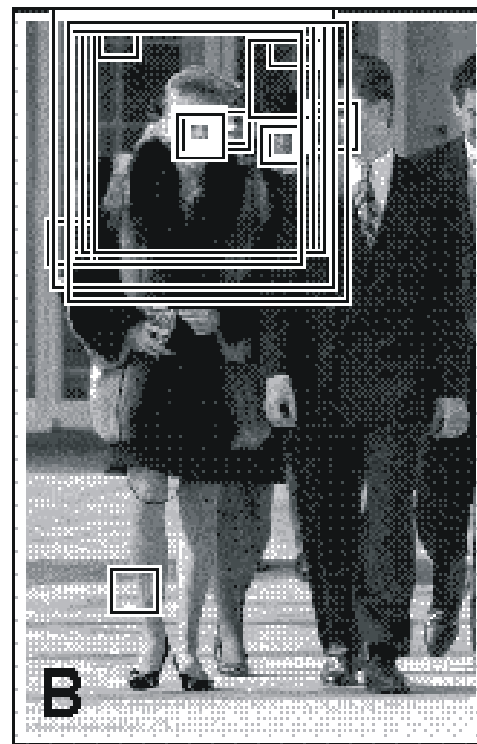
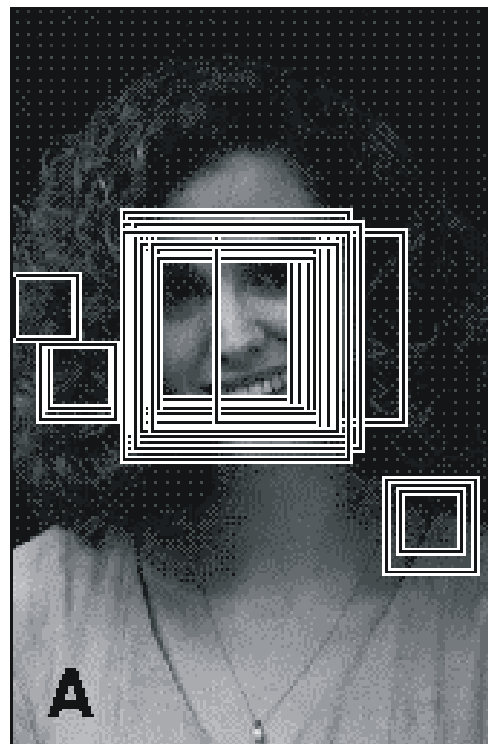
Approach

- Construct a neural network for the detection of texture
- Train this network with thousands of faces, where the line between the eyes as well as an orthogonal line towards the tip of the nose are marked
- Process an unknown image with search areas in varying sizes:
 - pre-processing / filtering / standardization of the illumination
 - test on the hypothesis “face” with the trained neural network.

Algorithm "Face Detection"



Multiple Detection



Visualization of Results

