FRAUNHOFER HHI AT TRECVID 2004: SHOT BOUNDARY DETECTION SYSTEM

Christian Petersohn

Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institut,
Einsteinufer 37, 10587 Berlin, Germany
(petersohn@hhi.fhg.de)

ABSTRACT

This paper describes the shot boundary detection and determination system developed at the Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institut, used for the evaluation at TRECVID 2004. The system detects and determines the position of hard cuts, dissolves, fades, and wipes. It is very fast and has proved to have a very good detection performance.

As input for our system, we use luminance pixel values of sub-sampled video data. The hard cut detector uses pixel and edge differences with an adaptive thresholding scheme. Flash detection and slow motion detection lower the false positive rate. Dissolve and fade detection is done with edge energy statistics, pixel and histogram differences, and a linearity measure. Wipe detection works with an evenness factor and double Hough transform. The difference between the submitted runs is basically only different threshold settings in the detectors, resulting in different recall and precision values.

1. INTRODUCTION

Huge amounts of video data are produced around the world each day. An automatic video shot detection and determination is important for tasks involving management, analysis, and search and retrieval of video data.

This paper gives an overview of the shot detection system developed at the Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institut. The different detection steps are described. Results on the TRECVID 2004 test set conclude the paper.

2. SYSTEM OVERVIEW

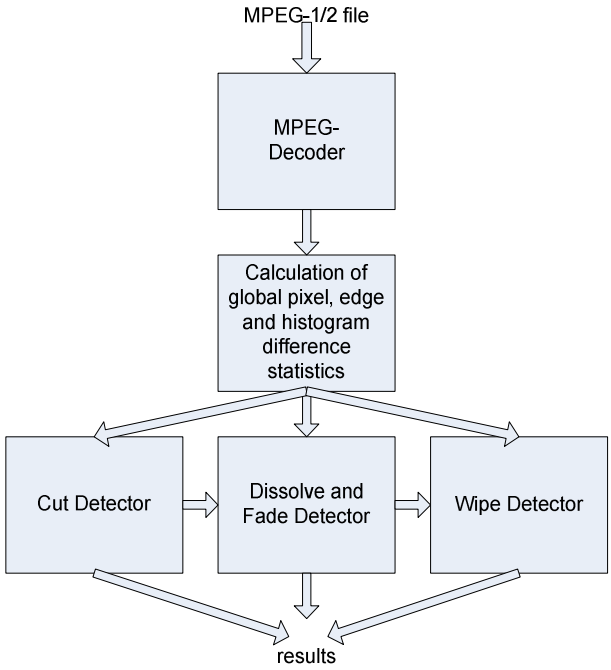

Fig. 1: System overview

An overview of the shot detection system is shown in Fig. 1. The first step is decoding the MPEG-file. Mpeg2dec (http://libmpeg2.sourceforge.net/) is used in this step. It is the fastest free MPEG-decoder we could find and it is available under the GNU General Public License (GPL). We made some modifications to the decoder during our research in order to enable the extraction of additional information such as DC-coefficients and motion vectors. The system used for TRECVID first performs a full decoding and works afterwards on video frames that have been sub-sampled by a factor of eight in x and y directions containing luminance

information only. This small input data rate used in the calculation of statistics and by the detectors enables very fast processing.

After decoding, pixel, edge, and histogram difference statistics are calculated for those features that are needed by one or more detectors for all frames. Features that are only needed in special situations (e.g. flash detection) are calculated when they are actually used. The various detectors are comprised of different stages. The first stage always marks shot boundary candidates. Each subsequent stage only examines the candidates from the preceding stage and tests whether additional criteria are met. To reduce the number of candidates to test, more complex operations are always done in the latest stage possible.

3. HARD CUT DETECTION

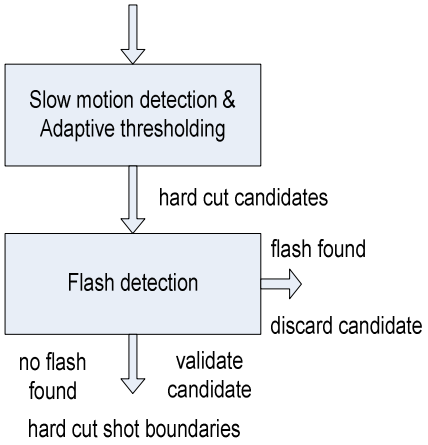

Fig. 2: Hard cut detection

For hard cut detection (Fig. 2), we use the edge and pixel differences between consecutive frames that are calculated after simple motion compensation. To be able to adapt to the varying statistics in a video file, adaptive thresholding is used to mark hard cut candidates. Next, candidates are tested with a flash detector that examines the pixel and edge differences for pairs of frames with different temporal distances around the hard cut candidate. If any of the differences is too small, the candidate is rejected. As an additional step, we test for slow motion passages before marking candidates, as we found that our adaptive thresholding scheme produced too many false positives for those sections.

4. DISSOLVE AND FADE DETECTION

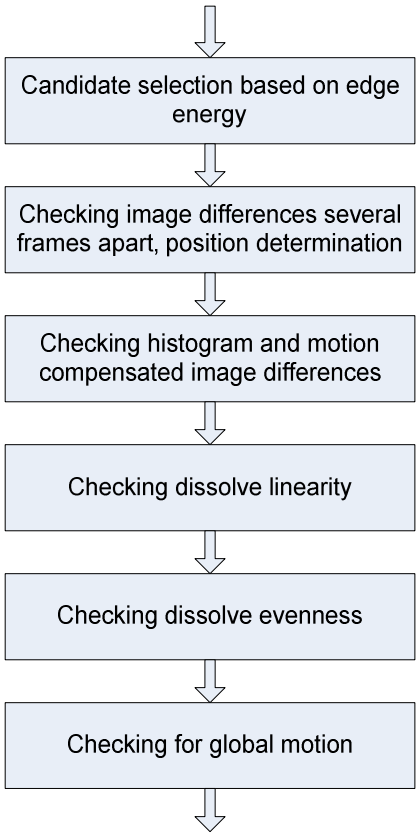

Fig. 3: Dissolve detection

The dissolve detector (Fig. 3) consists of six stages [1]:

1. Selection of dissolve candidates by searching for U-shapes in the edge energy diagram (Fig. 4)

2. Checking image differences several frames apart and determination of position and duration of the dissolve candidate (Fig. 5)
3. Checking histogram differences and motion compensated image differences
4. Checking dissolve linearity
5. Checking dissolve evenness
6. Checking global motion with cross correlation

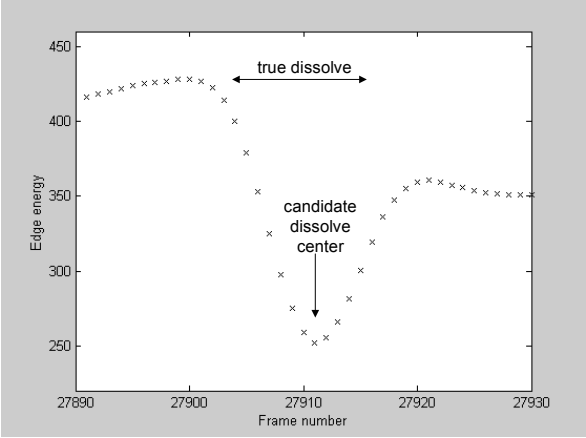

Fig. 4: Edge energy diagram (stage 1)

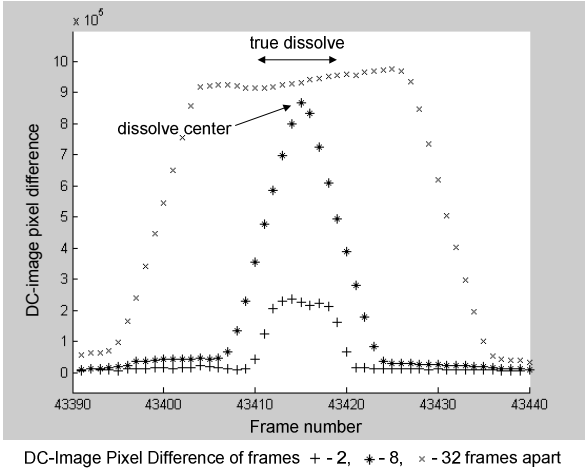

Fig. 5: Image differences curves (stage 2)

The linearity check in stage 4 basically works with three images S, C, and E for the dissolve start, center, and end and compares pairwise differences using triangle inequality. During a dissolve, the difference between S and E should be as large as the sum of the differences between S and C and C and E. Whereas during motion etc., the difference between S and E tends to be smaller.

Stage 5 works with an evenness measure we defined that is related to the variance of difference images. Changes in between frames during a dissolve tend to be more evenly distributed than during object motion or object fade out/in (see Fig. 6)

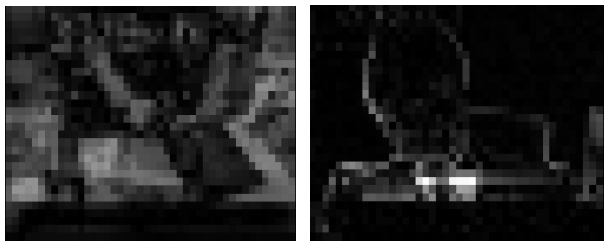

Fig. 6: Difference image during a dissolve (left) and a false dissolve candidate with object motion and headline fade-out (right)

During dissolve detection, fades get marked as shot boundary candidates. If edge energy is below a threshold close to zero, we mark it as a fade and correct the shot boundaries accordingly.

5. WIPE DETECTION

For TRECVID, we used an initial version of our wipe detector (Fig. 7). To mark wipe candidates, it uses an evenness factor that exploits the observation that, during a wipe, spatial zones of change (Fig. 9) move through the image [2]. This corresponds to low minima in the evenness factor curve (Fig. 8).

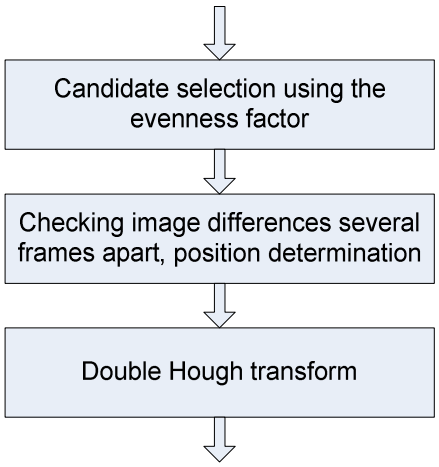

Fig. 7: Wipe detection

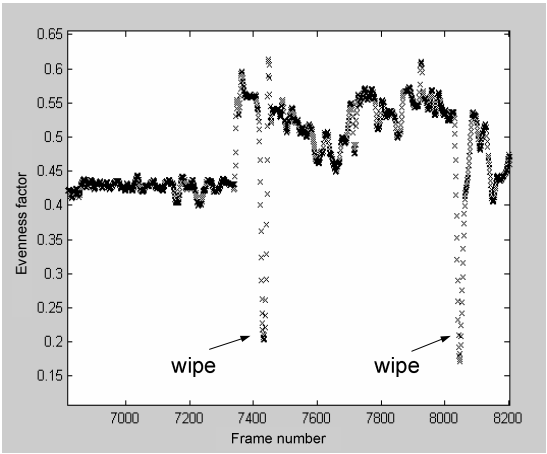

Fig. 8: Evenness factor curve with two wipes

In the second stage, differences between images several frames apart are checked. In the third stage, we use a double Hough transform on the wipe candidate's set of difference images. The first Hough transform detects linear segments in the difference images (i.e. boundaries in the images during the wipe). The accumulators in the transformed Hough spaces (Fig. 10) are set to one if a line is found. All Hough spaces for the difference images are added (Fig. 11), and the result is, itself, Hough-transformed. This is done to find patterns in the movement of the boundaries during the wipe.

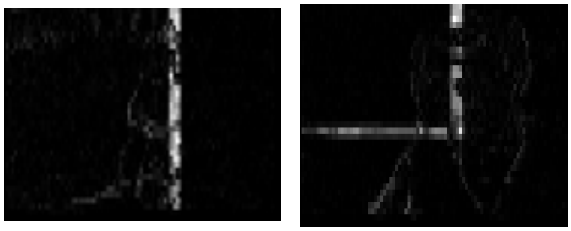

Fig. 9: Difference images during a wipe

Fig. 10: Hough transform of difference image

Fig. 11: Added binarized Hough spaces, input for second Hough transform

7. RESULTS AT TRECVID

The following charts show the shot boundary detection results for all transitions (Fig. 12), cuts (Fig. 13), and gradual transitions (Fig. 14 and Fig. 15). Our results are marked in blue. The results of the other teams are shown in yellow.

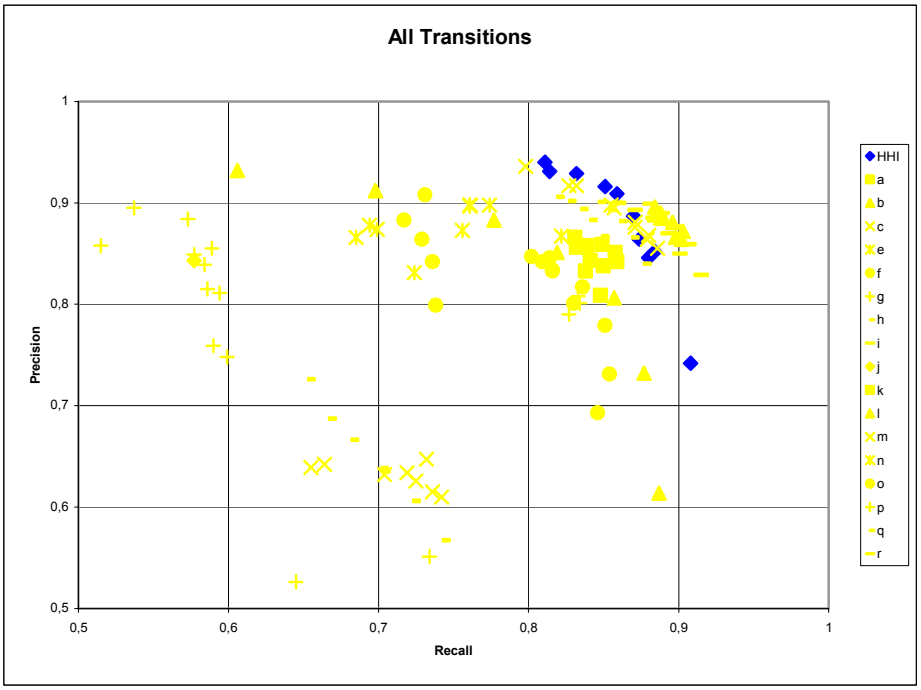

Fig. 12: Precision vs. recall for all transitions

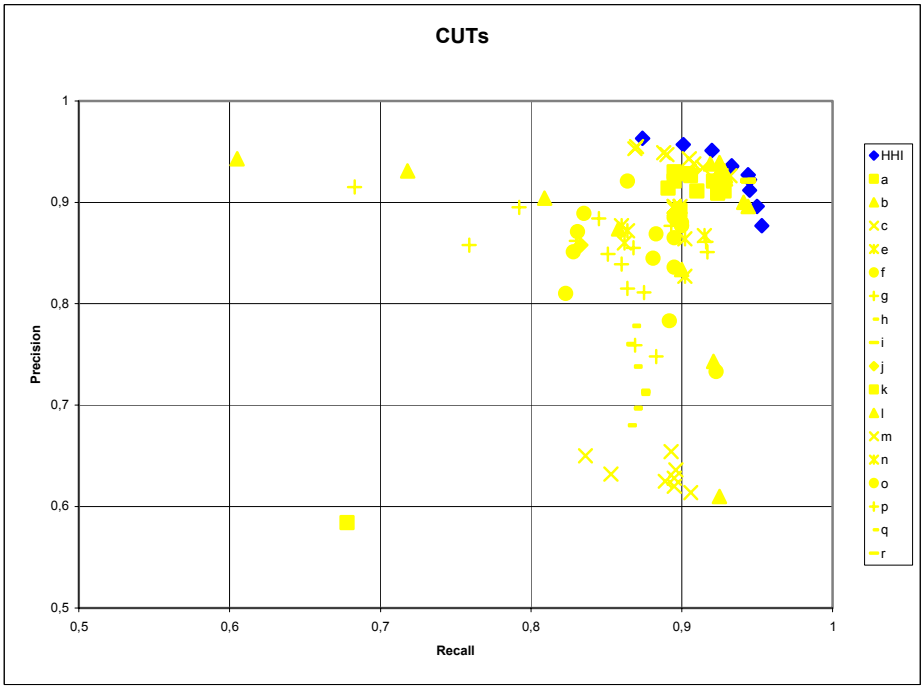

Fig. 13: Precision vs. recall for cuts

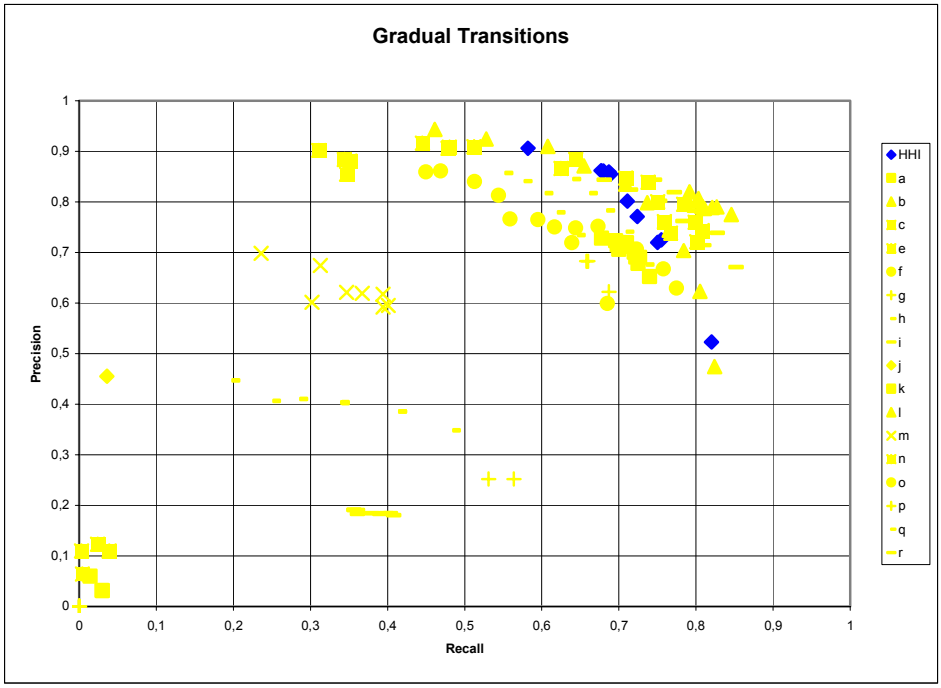

Fig. 14: Precision vs. recall for gradual transitions

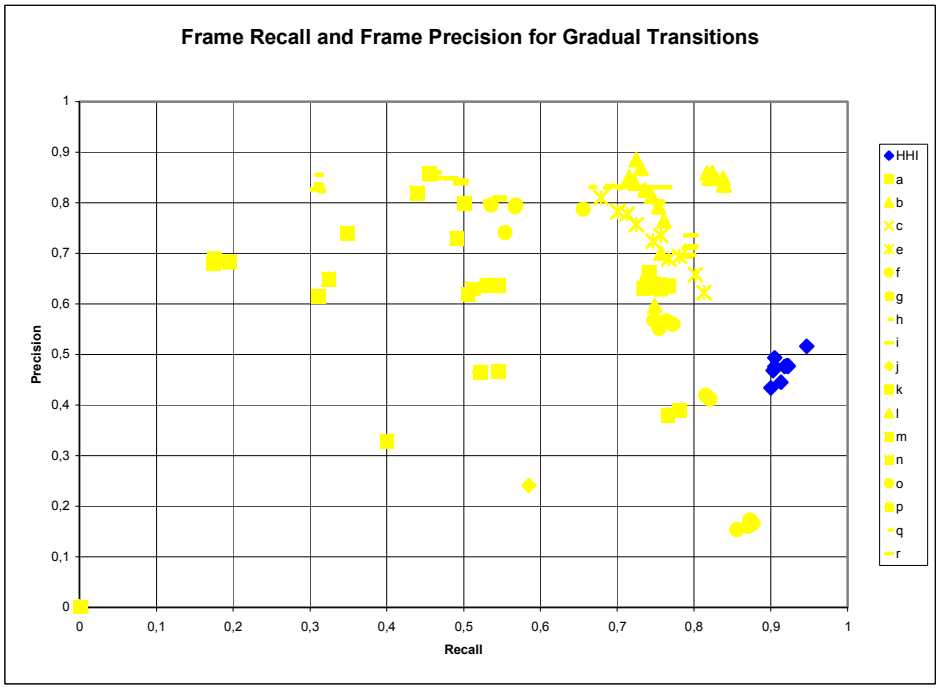

Fig. 15: Frame precision vs. frame recall for gradual transitions

Our shot detection approach showed very good results even though we chose an approach with low computational complexity. Especially the hard cut detection performance is excellent.

8. COMPUTATIONAL COMPLEXITY

We measured execution times on a PC with a P4 Xeon 3.06GHz processor. Decoding times include MPEG-decoding only. Therefore, they are independent of parameters used during segmentation and the same for every run. Segmentation times include sub-sampling of the frames, extracting, and analysing feature statistics.

	Total time in sec	Decoding time in sec	Segmenta-tion time in sec
Run-1	996	600	396
Run-2	993	600	393
Run-3	968	600	368
Run-4	1001	600	401
Run-5	966	600	366
Run-6	984	600	384
Run-7	990	600	390
Run-8	991	600	391
Run-9	1003	600	403
Run-10	1011	600	411

Table 1: Time consumption

Total real playing time for the shot boundary detection test set is 20614 seconds (618409 frames). That means our system is about 20 times faster than real time on the test set on the computer we used.

9. CONCLUSION

The shot boundary detection system used for the evaluation at TRECVID consists of separate detectors for hard cuts, dissolves and fades, and wipes. It has a very low computational complexity only using sub-sampled luminance images as input. It showed to have a very good detection performance, especially for hard cuts.

REFERENCES

[1] C. Petersohn. "Dissolve Shot Boundary Determination", *Proc. IEE European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology*, pp. 87-94, London, UK, 2004

[2] C. Petersohn. "Wipe Shot Boundary Determination", *Proc. IS&T/SPIE Electronic Imaging 2005, Storage and Retrieval Methods and Applications for Multimedia*, pp. 337-346, San Jose, CA, 2005