Performance evaluation of Fade and Dissolve Detection algorithm for different video sequences

S. A. Chavan, Miss. Rajeshree B. Telang, S. G. Akojwar.

Author Details
Miss. Rajeshree B. Telang, Department of Electronics & Telecommunication, Dr. Bhauasaheb Nandurkar College of Engg. & Tech., Yavatmal (M.S.), India, Email: rajeshree.telang@rediffmail.com
Rajeshree B. Telang is currently pursuing master’s degree program in Digital Electronics engineering in Amravati University, India.

S. A. Chavan is currently pursuing Ph.D. degree program in electronics engineering in Sant Gadge Baba Amravati University, India, Email: salimsahil97@rediffmail.com

S. G. Akojwar, Professor & Head of Department EXTC Engg., Rajiv Gandhi College of Engineering & Research, Chandrapur, India, Email: sudhirakojwar@rediffmail.com

Keywords
Shot Boundary Detection, Scene changes, Gradual Transition, Hard cuts, Fades, Dissolve, Adaptive threshold.

ABSTRACT
Automatic shot boundary detection has been an active research area for nearly a decade and has led to high performance detection algorithms for hard cuts, fades, dissolve, wipes. To extract valid information from videos, without any loss of information, much attention is being paid to video processing technology. Digital video is widely used in multimedia databases and requires effective retrieval techniques. Shot boundary detection is a common first step in analysing video content. Segmenting digital video into its constituent basic semantic entities, or shots, is an important step for effective management and retrieval of video data. To detect transitions between shots, many algorithms are developed which are highly effective on abrupt transition. The effective detection of gradual transitions is an especially a difficult task. In this paper, we present a gradual transition detection approach based on average frame similarity and adaptive threshold. We report good results on different video sequences, particularly for fades and dissolve.
1. INTRODUCTION

With the widespread usage of video in computer systems and networks, automatic video content summarization techniques are becoming a necessity. A concise and informative video summary will enable the user to quickly figure out general contents of video and help him/her to decide if it is worthwhile to watch through the whole sequence. For video content retrieval, a video summary will dramatically save the user’s time and efforts to spot the desired videos from a large volume of video collection. For this content to be usable, it must be easily accessible. An important first step is to identify and annotate section of interest.

Historically, identification and annotation have been performed by human annotators [13, 14]. Manual annotation provides the best quality but it is a long and tedious process. Automatic indexing method has the potential to solve this problem.

The first step that has to be achieved by automated indexing system is segmenting the video sequences into a collection of video shots [4]. A video shot is defined as an uninterrupted sequence of frames, corresponding to the same scene or event. The transition between adjacent shots can be abrupt-a cut- or gradual. A shot change where two consecutive frames belongs to different shots. The later involves a changeover between two shots using video editing techniques, such as dissolve, fades, and wipes [5].

Fades, dissolve, spatial edits, and other gradual transitions are more complex but less frequent, these are much common in entertainment footage such as movies and television serials. Accurate detection of cuts fades and dissolve is crucial to video segmentation; indeed Lienhart [7] reports that these transitions account more than 99% of all transitions across all types of video.

In this paper, we present novel approach to gradual transition detection in video based on Moving Query Window method that applies to the problem of shot boundary detection.

2. BACKGROUND

Recent automated techniques for detecting transitions between shots are highly effective on abrupt transitions. Automatic cut detection method uses the video indexing system [1] that explores fully automatic content analysis method for cut detection with multi-model feature extraction. Image comparison with motion compensation technique [26] also detects cuts transitions with a special module for detecting photographic flashes and filtering them as erroneous ‘cuts’ via motion-peak detector. Indeed the results are comparable to results obtained by human observers [2]. However, gradual transitions are more difficult to detect using automated system [8, 10].

Shot boundary detection techniques can be categorized as using compressed or uncompressed video. The former consider features of encoded footage such as DCT coefficients, macro blocks, or motion vectors [15, 16, 17]. These techniques are efficient because the video does not need to be fully decoded. Koprińska [6] provide an overview of existing approaches. But using encoded features directly, it results in lower precision. The exact transition boundaries may not be identifiable, or gradual transitions may not be distinguishable from object movements.

Most of the approaches work on uncompressed video with frame difference as a measure for shot boundary detection. It rely on the property that adjacent frames within one shot are usually similar. By evaluating inter frame differences, and searching for significant dissimilarities, transition between shots can be detected.

There are several methods to measure the difference between frames. The popular approach is ‘pixel-by-pixel’ comparison method. A simple way to detect a qualitative change between a pair of images is to compare the corresponding pixels in two frames to determine how many pixels have changed. While this method shows good results [3], it is computationally expensive and sensitive to camera motion, camera zoom, intensity variation, and noise.

Approaches that use clustering algorithm monitor frame similarity, and identify frames that belong to a scene change. Gunsel [18] proposed scene change detection method using unsupervised clustering to eliminate the data dependency of threshold selection. Colour image segmentation is a fundamental task in many computer vision problems. A common is to use fuzzy iterative clustering algorithms that provide a partition of pixels into a given number of clusters. C. Lo [19] proposed a fuzzy c-means algorithm which relies on a new efficient cluster centre initialization and colour quantization allowing faster and more accurate convergence such that it is suitable to segment very large colour images. But they are time consuming and sensitive to initialization and noise.

Most popular techniques on uncompressed video summarize frame content using histograms. Such approaches represent a frame, or parts of a frame, by the frequency distribution of features such as colour or texture. The simplest histogram method computes grey level or colour histogram of the two images. If the bin-wise difference between two histogram is above a threshold, a shot boundary is assumed. Approaches using global histogram [9, 12, 14] represent each frame as a single vector. Truong [12] proposed conventional cut detection method using colour histogram differences by utilizing an adaptive threshold computed from a local window on the luminance histogram differences curve. Approaches using localized histogram [27] generate separate histograms for subsections of each frame. The twin-comparison algorithm first proposed by Zhang et al. [14] is the basis of several proposed approaches for detection of gradual transitions [13, 24, 25]. Here, a low threshold is applied to detect groups of frames that belong to a possible gradual
transition. The accumulative inter-frame distance is calculated for these frames. A gradual transition is reported if the accumulated inter-frame distance exceeds a second, higher threshold.

Quenot et al. [20, 21] use direct image comparison for cut detection. To reduce false positives, motion compensation is applied before image comparison. A separate flash detection module is used to further reduce false positives. Gradual transitions are detected by checking whether the pixel intensity in adjacent frames approximately follows a linear, non-constant function. Recent work in TRECVID indicates that histograms seem to be the favoured way to represent feature data. Adams et al. [1] propose a video retrieval system which employs a combination of three-dimensional RGB colour histograms and localized edge gradient histograms for shot boundary detection. Recent frames are held in memory to compute adaptive thresholds. The system proposed by Hua et al. [28] uses global histograms in the RGB colour space. Pickering et al. [29] use a detection algorithm which employs localized RGB colour histograms. Each frame is divided into nine blocks and the median between the nine block distances is computed. A transition is detected when the median distance exceeds a fixed threshold. Wu et al. [30] propose a shot boundary detection algorithm which calculates frame-to-frame difference based on luminance information and histogram similarity in the RGB colour space. Flash and motion detectors are used to reduce false positives.

Nagasaki and Tanaka [9] compared several simple statistics based on grey level and colour histograms. They found the best results by breaking the images into 16 regions, using a χ2 test on colour histograms of those regions, and discarding the eight largest differences to reduce the effects of object motion and noise.

Another approach involves applying transforms to the frame data. Cooper, Foote, Adcock & Casi [22] represent frames by their low-order DCT coefficients, and calculate the similarity of each frame to the frames surrounding it. The frames before and after a cut would have high similarity to past and future frames respectively, but low similarity across the boundary. Miene, Hermes, Ioannidis & Herzog [23] use FFT coefficients calculated from a grey scale version of the frame for their comparisons. A detailed overview of existing techniques is provided by Koprinska & Carrato [6].

3. GRADUAL TRANSITION WITH MOVING QUERY WINDOW

In this section, we focused on the approach for the detection of gradual transitions based on Moving Query Window technique. The moving query window [11] caters for the fact that gradual transitions usually extend over several frames by evaluating the average inter-frame distance in a set of frames, rather than examining only individual frames. Many researches previously proposed moving query window technique, but it is effective only for cut detection. The moving query window technique performs comparison on a set of frame to detect abrupt transitions. This method proceed through a video, and take each frame in turn as a pivot, and consider a fixed-size window of frames encompassing current frame. This moving window consisting of two equal-sized half windows on either side of current frame as illustrated in figure 1. After that, by employing one dimensional global colour histogram to represent frame content, frame similarity is evaluated. Based on the histogram similarity, all frames in moving window are ranked to the current frame; the most similar frame is ranked highest. The number of frames from the preceding half window that are ranked in top half is monitored while advancing through the video. A cut is reported when this number exceeds an upper threshold and falls below a lower threshold within four consecutive frames. This approaches show effectiveness on cut detection, however without modification, this scheme is less effective on gradual transition.

![Image](image.png)

**Figure 1.** An equal number of frames on each side of the current frame — the pre-frames and the post-frames constitute the moving query window.

For the effective detection of gradual Transition, we have developed a novel extension of Moving query window approach. As we are knowing with the fact that, the cuts shows significant inter-frame distances within a few consecutive frames, the method of ranking frames in a query window works well for abrupt transitions. Observations have shown that this is not usually the case for gradual transitions, where inter-frame distances are typically smaller. This results in our approach being far less effective in detecting gradual transitions.

To overcome this problem, we define two sets of frames rather than comparing individual frames. The two sets of frames are on either side of current frame, refer as pre-frames and post-frames. Then the inter frame distance is determined between each of the two sets and the current frame. Also the averages inter frame distance between each of the two sets and current frame is calculated. This gives the final value that is the average distance between two sets and current frame. This computation results in two values, one each for the pre- and post-frame sets, and we use the ratio of these values — referred to as PrePostRatio — to detect gradual transitions.
The moving query window technique is examined with the following example: consider two shots, shot A and shot B. We have to detect the gradual transitions. We assume that dissolve starts from frame 12 and ends with frame 22. The moving window comprised of two equal sized with 10 sets of frames preceding and following the current frame. As we are assuming 10 frames preceding and following the current frame, means the 11th is the current frame, which belongs to shot A and the last frame before the transition starts. Frames 1-10 are pre-frames and similar to current frame as they belong to shot A. That means, there inter-frame distance to the current frame is relatively small. We assume it has the value 2. Frame 12-21 are the post-frames and mostly dissolve frames, means inter-frame distance to the current frame is comparatively high because they are relatively dissimilar to the current frame. Let us assume it has the value 10. Given the pre-frame average of 2 and post-frame average of 10, the pre-post ratio for the first row is 2/10=0.2.

As shown in figure 2, the current frame moves further into the dissolve in row two and three. For the second row, 15th is the current frame, and dissolve starts from frame 12 and ends with frame 22. Means, the PrePostRatio is slowly rising and steeply rising for third row. In fourth row, frame 22 is the current frame which is also the last frame of transition. This frame is very much similar to frames 23 to 32 that belong to shot B, producing low average inter-frame distance. Assume it has the value 2.

The pre frames that are formed by frames 12 to 21 are the dissolve frames, there average inter-frame distance is high, we assume a value of 10. Hence PrePostRatio for fourth row is 10/2=5. Once the window exist the transition completely, the ratio usually reverts to a relatively low value.

This behaviour is common for both dissolve and fades. By monitoring the PrePostRatio through a video clip, we can detect the minima and maxima that accompany the start and end of such transition.

Figure 2. An example of a dissolve. Before the transition, the PrePostRatio is minimal. It rises to a maximum as we proceed through the transition, before falling again afterwards.

Further improvement for gradual transition detection replaces all fixed threshold by adaptive methods to increase recall and make the system more applicable to different types of videos. Adaptive threshold is set by calculating average mean and standard deviation by maintaining the history of PrePostRatio values. It is sometimes necessary to adjust the level of threshold. For example poor quality and noisy videos produces smaller peaks in PrePostRatio curves, to reduce false detection caused by these peaks, we multiply the calculated threshold by factor called UTF-Upper Threshold Factor.

The most important algorithm parameter influencing the results is the number of pre- and post-frames on either side of the current frame, which we refer to as the Half-Window Size (HWS). The number of frames in the entire query window is then 2×hws, as shown in Figure 1. The current frame is not part of the query window.

4. EVALUATION PARAMETER

The performance measure of the algorithm is evaluated by calculating two parameters. Recall is the fraction of all known transitions that are correctly detected, while precision is the fraction of reported transitions that match the known transitions recorded in the reference data.

\[
\text{Recall} = \frac{N_{\text{Correct}}}{N_{\text{SCD}}} = \frac{N_{\text{Correct}}}{N_{\text{SCD}}} = \frac{N_{\text{Correct}}}{N_{\text{Correct}} + N_{\text{Missed}}} \times 100
\]

\[
\text{Precision} = \frac{N_{\text{Correct}}}{N_{\text{Correct}} + N_{\text{False}}} \times 100
\]

where, \(N_{\text{Correct}}\) is the number of correctly detected frames, \(N_{\text{False}}\) is the number of falsely detected frames, \(N_{\text{Missed}}\) is the number of missed frames, and \(N_{\text{SCD}}\) is total number of frames where transitions are occurred.
We developed our algorithm using the shot boundary detection task subset of varieties of different types of video. The detailed results for the video are shown in Table 1.

<table>
<thead>
<tr>
<th>Collection</th>
<th>Clip</th>
<th>Frames</th>
<th>Gradual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video 1</td>
<td>10</td>
<td>2800</td>
<td>501</td>
</tr>
<tr>
<td>Video 2</td>
<td>07</td>
<td>1500</td>
<td>277</td>
</tr>
<tr>
<td>Video 3</td>
<td>10</td>
<td>3100</td>
<td>165</td>
</tr>
<tr>
<td>Video 4</td>
<td>12</td>
<td>1452</td>
<td>276</td>
</tr>
</tbody>
</table>

Table 1. Details of our test collections.

5. RESULT

We have experimented with one-dimensional and three-dimensional histograms using the RGB and HSV colour spaces. We have found that gradual transitions are best detected with one-dimensional HSV colour histograms using 32 bins per colour component. All results reported in this paper are for this feature representation.

Table 2 shows results for detecting gradual transition for each of the four video collections using algorithm parameters that produce the best performance.

To reduce the effects of low video quality, camera motion, and compression artefact’s, a DMZ of one frame on either side of the current frame is applied. The best parameter values over all four video collections are HWS=14, DMZ=0, and UTF=1.

<table>
<thead>
<tr>
<th>Collection</th>
<th>Reference Transition</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fades</td>
<td>Dissolve</td>
<td>Fades</td>
</tr>
<tr>
<td>Video 1</td>
<td>266</td>
<td>235</td>
<td>81%</td>
</tr>
<tr>
<td>Video 2</td>
<td>136</td>
<td>141</td>
<td>83.91%</td>
</tr>
<tr>
<td>Video 3</td>
<td>146</td>
<td>19</td>
<td>72%</td>
</tr>
<tr>
<td>Video 4</td>
<td>181</td>
<td>95</td>
<td>67.92%</td>
</tr>
</tbody>
</table>

Table 2. Results grouped by transition type for the best run on each test collection. We observe much better performance for dissolves and fades than for other types of gradual transition.

5. CONCLUSION

Effective identification of gradual transition is important for video indexing and retrieval. In this paper, we have proposed an approach to gradual transition, based on moving query window technique. These methods evaluated inter frame distance between the two set of frames referred as pre and post frames and the current frame. This algorithm is very much effective for gradual transition with recall and precision. Also it is effectively detect start and end of gradual transition.

In future, by using localized histogram and edge tracking feature we count to overcome false detections.

References

National Conference On Research Trends In Electronics, Computer Science & Information Technology And Doctoral Research Meet, Feb 21st & 22nd


