International Journal of Electrical and Electronics Engineering (IJEEE)
Vol.1, Issue 1 Aug 2012 46-57
© IASET

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

A REVIEW ON DIFFERENT METHODS OF VIDEO SHOT BOUNDARY DETECTION

[1]RAVI MISHRA & [2]SANJAY KUMAR SINGHAI

[1]EEE department, SSCET,Gunwani, Bhilai Nagar (C.G), India

[2]Electrical Engg. department, Govt. Engg. College, Bilaspur (C.G), India

ABSTRACT

Many algorithms have been proposed for detecting video shot boundaries and classifying shot and shot transition types. Here we are using four different methods for comparison, using GIST, Segmentation, Color Histogram and Motion Activity Discriptor.

KEYWORDS: Video Shot, Detection Techniques, Visual Communication.

INTRODUCTION

Shot-boundary detection is the first step towards scene extraction in videos, which is useful for video content analysis and indexing. Shots are basic units of a video. There are many types of transitions between shots. Shot boundaries can be classified into two main categories: cut and gradual. A cut is an abrupt shot change that occurs over a single frame while a gradual is a slow change that occurs in a number of consecutive frames. With the gradual type, fades and dissolves are common. A fade is usually a change in brightness with one or several solid black frames in between, while a dissolve occurs when the images in the current shot get dimmer and the images of the next shot get brighter[1]. They are required for further video analysis such as:

- Person tracking, identification.
- high-level feature detection.

They provide cue about high-level semantics:

- In video production each transition type is choose carefully to support the content and context.
- For example dissolves occur much more often in feature films and documentaries, while wipe usually occur in news, sports and shows.

Problem: Given a video V consisting of n shots, find the beginning and the end of each shot. It is fundamental to any kind of video analysis and video application since it enables segmentation of a video into its basic components.

Many approaches have been proposed for shot boundary detection. The simplest approach is to compute the differences between the color distributions of consecutive frames and use a threshold to classify whether a hard cut occurs. In order to detect gradual transitions, edge change ratios or motion vectors can be used[1], [2]. Since these approaches use threshold based models for detection, their advantage is they are fast. Nevertheless, they are sensitive to changes in illumination and motion. Furthermore, they are difficult to generalize for new datasets. Recent works[3], [4] use machine learning methods for making decisions and have received impressive results.

The basis of detecting shot boundaries in video sequences is the fact that frames surrounding a boundary generally display a significant change in their visual contents. The detection process is then the recognition of considerable discontinuities in the visual-content flow of a video sequence. In the first step of this process, feature extraction is performed, where the features depict various aspects of the visual content of a video. Then, a metric is used to quantify the feature variation from k frame to frame k+l. The discontinuity value z(k, k+l) is the magnitude of this variation and serves as an input into the detector. There, it is compared against a threshold T. If the threshold is exceeded, a shot boundary between frames k and k+l is detected. To be able to draw reliable conclusions about the presence or absence of a shot boundary between frames k and k+l , we need to use the features and metrics for computing the discontinuity values z(k, k+l) that are as discriminating as possible. This means that a clear separation should exist between discontinuity-value ranges for measurements performed within shots and at shot boundaries.

Further in this paper section 2, shows the different features that can be used to measure visual discontinuity, section 3, shows the method use to implement video shot boundary detection algorithm, section 4, gives a comparison between between difference methods.

DIFFERENT FEATURES THAT CAN BE USED TO MEASURE VISUAL DISCONTINUITY

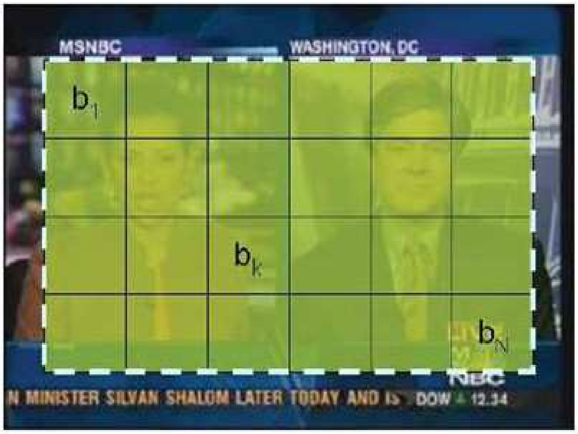
Pixel Differences

Two common approaches:

- Calculate pixel-to-pixel difference & Compare the sum with a threshold
- Count the number of pixels that change in value more than some threshold & Compare the total number against a second threshold.

Statistical Differences

- Divide image into regions
- Compute statistical measures from these regions (e.g., mean, standard deviation)
- Compare the obtained statistical measures



Compression Differences

- Use differences in the discrete cosine transform (DCT) coefficients of JPEG compressed frames as the measure of frame similarity.
- Avoid the need to decompress the frames

Edges Differences

- The edges of the objects in the last frame before the hard cut usually cannot be found in the first frame after the hard cut,
- The edges of the objects in the first frame after the hard cut in turn cannot be found in the last frame before the hard cut.
- Use Edge Change Ratio (ECR) to detect hard cuts Edges/contours: Based on edge change ratio (ECR). Let σn be the number of edge pixels in frame n, and Xnin and Xn-1out the number of entering and exiting edge pixels in frames in frames n and n-1, respectively. The edge change ratio ECRn between frames n-1 and n is defined as:

$$ECR_n = \max(X_n^{in} / \delta_n, X_{n-1}^{out} / \delta_{n-1})$$

Motion Vector

Use motion vector to determine discontinuity.

Image *I*

Image *I -Rotated*

Illustration of optical flow

Histograms Comparison

The most common method used to detect shot boundaries.

- Provides good trade-off between accuracy and speed
- The simplest histogram method computes gray level or color histograms of the two images. If the bin-wise difference between the two histograms is above a threshold, a shot boundary is assumed.
- Several extensions available: Using regions, region weighting, different distance metrics.

METHOD USE TO IMPLEMENT VIDEO SHOT BOUNDARY DETECTION ALGORITHM

Many algorithms have been proposed for detecting video shot boundaries and classifying shot and shot transition types. Few published studies compare available algorithms, and those that do have looked at limited range of test material. A comparison of several shot boundary detection and classification techniques can be done using their variations including histograms, discrete cosine transform, motion vector, and block matching methods. The performance and ease of selecting good thresholds for these algorithms are evaluated based on a wide variety of video sequences with a good mix of transition types. Threshold selection requires a trade-off between recall and precision that must be guided by the target application.

Video Shot Boundary Detection Using Color Histogram

The approach here takes each (352 * 288) frame of the video and computes a color histogram using a total of 192 different colour bins where each bin contains the percentage of pixels from the whole frame. When this is done, the color histogram for each frame is checked against the histogram for the one following and a similarity metric is used to compute the likeness between the two adjacent frames. When the sequence of similarity values is analysed, a shot will show a big change in the sequence where the shot boundary occurs, so by looking for these peaks in the differences, a shot boundary can be detected. The peak in the values is required to be above some minimum value before it will be detected as a shot cut Video programmes will have many different types of shot boundaries, and those where the shot boundary occurs over a number of frames may not be detected by the method above. For example, a fade or dissolve occurring over a 3 second period in, say, a cookery or gardening program will span a total of 75 frames and the incremental difference between adjacent frames in this sequence will be quite small. Furthermore, it is possible that the two separate shots in such a transition may have similar colouring and hence colour histograms, anyway. Such omissions are difficult to avoid using colour histogram based segmentation, but nevertheless this method is very popular and if used correctly it is very reliable and accurate. On the downside it is slow to compute because each frame of the digital video has to be decoded and calculations run on it to extract color values.

Video Shot Boundary Detection Using Edge Detection

This approach looks not at the colour differences, but at the differences between the edges detected in adjacent frames. Each frame in the digital video is turned into a greyscale image and Sobel filtering applied to detect edges. The method looks for similar edges in adjacent frames to detect a shot boundary. The principle behind the edge detection approach is that it can counter problems caused by fades and dissolves and other transitions which are invariant to gradual colour changes. With edge detection, even when there are gradual transitions between shots, there should always be a pair of adjacent frames where an edge is detected in one, and not in the other and identifying an occurrence of this on a large scale locates a shot transition. Like the colour-based method above, this will require a minimum difference between adjacent frames to detect a shot cut but it has the advantage of not being fooled by a large colour change for example when we get a camera flash. In such a case the colour-based methods will register a large difference in colour and as a result it will detect a shot boundary but the edge detection method will be looking for pixel edge differences. The downside of this technique is that it needs to decode each frame and as a result may compute slowly.

Video Shot Boundary Detection Using Using Macroblocks

The third of our methods for shot boundary detection, unlike the others, is based on processing the encoded version of the video. One of the features of MPEG-1 encoding is that each frame is broken into a fixed number of segments called macroblocks and there are three types of these macroblocks, I-, P- and B-. The classification of the different macroblock types is done at the encoder, based on the motion estimation and efficiency of the encoding.

I-macroblocks are encoded independently of other macroblocks and are used when the segment being encoded is very different in appearance from other segments around it in the same frame, or the corresponding segment in previous or following frames. P-macroblocks do not encode the region in the frame but instead have a *motion vector* which indicates the difference between the current block being encoded and the corresponding block in the previous frame. This is refered to as motion compensation and is used in video coding where objects remain stationary or move only a short distance between adjacent frames. B-macroblocks are bi-directionally predicted blocks and use both forward and backward prediction.

Different types of video sequence lead to different uses of I-, B- and P-macroblocks and different types of shot transitions in particular tend to have particular characteristic uses of macroblocks. Our use of macroblocks in this work is to see if we can identify and use these characteristics to determine shot boundaries directly from the compressed form of the video. For example, if a frame type that would be expected to contain backward predicted blocks does not actually have any, then it could be suggested that the future picture changes dramatically and this could indicate a shot boundary. Because it operates on the encoded form directly, it is likely to be more efficient than the other methods we use.

Video Shot Boundary Detection Using Segmentation Method

Although many methods have been proposed for this task, finding a general and robust shot boundary method that is able to handle the various transition types caused by photo flashes, rapid camera movement and object movement is still challenging. We present a novel approach for detecting video shot boundaries in which we cast the problem of shot boundary detection into the problem of text segmentation in natural language processing. This is possible by assuming that each frame is a word and then the shot boundaries are treated as text segment boundaries (e.g. topics).

The text segmentation based approaches in natural language processing can be used. The shot boundary detection process for a given video is carried out through two main stages. In the first stage, frames are extracted and labelled with pre-defined labels. In the second stage, the shot boundaries are identified by grouping the labelled frames into segments. We use the following six labels to label frames in a video: NORM FRM (frame of a normal shot), PRE CUT (pre-frame of a CUT transition), POST CUT (post-frame of a CUT transition), PRE GRAD (pre-frame of a GRADUAL transition), IN GRAD (frame inside a GRADUAL transition), and POST GRAD (post-frame of a GRADUAL transition).

Given a sequence of labelled frames, the shot boundaries and transition types are identified by looking up and processing the frames marked with a non NORM FRM label.

Video Shot Boundary Detection Using Motion Activity Descriptor

The motion activity is one of the motion features included in the visual part of the MPEG-7 standard. It also used to describe the level or intensity of activity, action, or motion in that video sequence. The main idea underlying the methods of segmentation schemes is that images in the vicinity of a transition are highly dissimilar. It then seeks to identify discontinuities in the video stream. The general principle is to extract a comment on each image, and then define a distance[5] (or similarity measure) between observations. The application of the distance between two successive images, the entire video stream, reduces a one-dimensional signal, in which we seek then the peaks (resp. hollow if similarity measure), which correspond to moments of high dissimilarity.

In this work, we used the extraction of key frames method based on detecting a significant change in the activity of motion. To jump 2 images which do not distort the calculations but we can minimize the execution time. First we extract the motion vectors between image i and image i+2 then calculates the intensity of motion, we repeat this process until reaching the last frame of the video and comparing the difference between the intensities of successive motion to a specified threshold. The idea can be visualised in figure below.

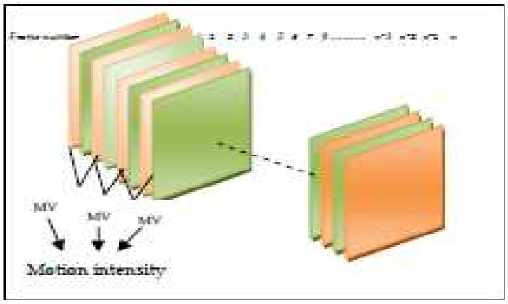


Fig:- The Idea of Video Segmentation Using Motion Intensity.

Video Shot Boundary Detection Using Independent Component Analysis

A new method for shot boundary detection using independent component analysis (ICA) is presented. By projecting video frames from illumination-invariant raw feature space into low dimensional ICA subspace, each video frame is represented by a two-dimensional compact feature vector. An iterative clustering algorithm based on adaptive thresholding is developed to detect cuts and gradual transitions simultaneously in ICA subspace. This method has the following major steps: (i) Raw feature generation from illumination-invariant chromaticity histograms; (ii) ICA feature extraction; (iii) Dynamic clustering for shot detection.

Video Shot Boundary Detection Using QR-Decomposition & Gaussian Transition Detection

The algorithm utilizes the properties of QR-decomposition and extracts a block-wise probability function that illustrates the probability of video frames to be in shot transitions. The probability function has abrupt changes in hard cut transitions, and semi-Gaussian behaviour in gradual transitions. The algorithm detects these transitions by analyzing the probability function.

Video Shot Boundary Detection Using Similarity Analysis

In this method we detect scene boundaries by considering the self-similarity of the video across time. For each instant in the video, the self similarity for past and future regions is computed, as well as the cross-similarity between the past and future. A significantly novel point in the video, i. e. a scene boundary, will have high self-similarity in the past and future and low cross-similarity between them.

Video Shot Boundary Detection Using Gist

Gist[6],[7] has been shown to characterize the structure of images well while being resistant to luminance change and also small translation. We use gist representation to model the global appearance of the scene. Gist treats the scene as one object, which can be characterized by consistent global and local structure. This method differs from prior work that characterizes scene by the identity of objects present in the image. Gist has been shown to perform well on scene category recognition[6]. Gist also provides good contextual prior for facilitating object recognition task[7]. Within one shot, due to the motion, appearance or disappearance of objects, the color histogram may not be consistent. Gist captures the overall texture of the background while ignoring these small change due to foreground objects.

In order to compute gist-features we resize the image into 128x128 pixels. The filters used to compute gist are divided into 3 scales and 8 orientation. After convolving with each of the 24 filters, images are equally divided into 16 blocks, and the average is taken for each block. This results in a 384-dimensional (16 x 24) vector. The dimensionality of the features is reduced by using PCA. We retain the top 40 components as our gist representation. To represent the color-content of the image, we compute a global histogram with 10 uniformly placed bins in RGB as well as HSV color-space. This results in 60 additional dimensions encoding the color. We observe that abrupt shot boundaries are marked by sharp-changes in the gist and color features. However, the absolute change in features are not consistent across videos or even across different shot boundaries within the same videos. Thus, a simple pair-wise frame difference will not work in order to detect shot- boundaries.

COMPARISON OF VARIOUS METHODS OF VIDEO SHOT BOUNDARY DETECTION

Here we are using four different methods involved in video shot boundary detection for comparison, GIST, Color Histogram, Segmentation and Motion Activity Descriptor.

Color Histogram Method

This technique is able to differentiate abrupt shot boundaries by the analysis of color histogram differences and smooth shot boundaries but temporal color variation. This provides a simple and fast algorithm able to work in real-time with reasonable high performances in a video indexing tool.

The algorithm firstly extracts abrupt shot discontinuities by the analysis of the color histogram difference, and then it detects possible gradual discontinuities for the following discrimination from motion activity as they behave a smooth discontinuity pattern. The simplicity of the method relies on the low complexity of the computation of the color histogram difference and that the most of the subsequent analysis is performed over the resulting 1-D signal. Unlike techniques based on pixel difference between adjacent frames, color histogram techniques reach a higher degree of independence to object motion within a frame sequence. A shot is detected by comparing the present value with a threshold value. The threshold value for cut detection used in the first derivative signal is obtained from the percentage of the maximum possible value of the color histogram difference. A cut is only detected when an abrupt variation overcome a certain percentage of color change in a video sequence.

The advantage of this method is that it takes into account the global variation of the image, being therefore less sensitive to camera or object movement.

Segmentation Method

The simplest approach as in color histogram is to compute the differences between the color distributions of consecutive frames and use a threshold to classify whether a hard cut occurs. In order to detect gradual transitions, edge change ratios or motion vectors can be used [1], [2]. Since these approaches use threshold based models for detection, their advantage is they are fast. Nevertheless, they are sensitive to changes in illumination and motion. Furthermore, they are difficult to generalize for new datasets. Recent works [3], [4] use machine learning methods for making decisions and have received impressive results on the test videos.

Given a sequence of labeled frames, the shot boundaries and transition types are identified by looking up and processing the frames marked. To label a frame in a video, we must firstly extract the features for that frame and then use a classifier, which has been trained in advance by the annotated frames, to classify it into one of the six categories NORM_FRM(frame of a normal shot), PRE CUT (pre-frame of a CUT transition), POST CUT (post-frame of a CUT transition), PRE GRAD (pre-frame of a GRADUAL transition), IN GRAD (frame inside a GRADUAL transition), and POST GRAD (post-frame of a GRADUAL transition).

SVM is used in this method for shot boundary detection. Support Vector Machines (SVM) is a statistical learning method based on the structure risk minimization principle[8]. It has been very efficiently proved to be useful in many pattern recognition applications[9]. In the case of binary classification, the objective of the SVM is to find the best separating hyperplane with a maximum margin. To train this classifier, we manually annotated frames in the training data. Using the trained classifier, we can label a sequence of frames.

Since the classifier occasionally produces false predictions due to the variations caused by photo flashes, rapid camera movement, and object movement, only using the perfect match between the predefined patterns and sub-sequences usually skips many of the true shot boundaries. Instead, we use a more flexible matching algorithm in which a match is declared if a Since the classifier occasionally produces false predictions due to the variations caused by photo flashes, rapid camera movement, and object movement, only using the perfect match between the predefined patterns and sub-sequences usually skips many of the true shot boundaries. Instead, we use a more flexible matching algorithm in which a match is declared if a portion of the predefined pattern is found in the input subsequence.

Motion Activity Discriptor

Video segmentation based on motion [10] is a new research area. Motion is a salient feature in video, in addition to other typical image features such as color, shape and texture. The motion activity is used in different applications such as video surveillance, fast browsing, dynamic video summarization, content-based querying etc. This information can for example be used for shot boundary detection, for shot classification or scene segmentation [11].

The magnitude of the motion vectors represents a measure of intensity of motion activity that includes several additional attributes that contribute towards the efficient use of these motion descriptors in a number of applications. Intensity of activity [12] is expressed by an integer in the range (1-5) and higher the value of intensity, higher the motion activity.

The motion features are extracted using the motion vectors. Block motion techniques are employed to extract the motion vectors. To extract the motion intensity, the motion of the moving objects has to be estimated first. This is called motion estimation. The commonly used motion estimation technique in all the standard video codecs is the block matching algorithm (BMA). Block matching is used to retrieve an initial estimate of the image displacement. To obtain a dense displacement field, matching with adaptive block sizes was implemented. In this typical algorithm, a frame is divided into blocks of M × N pixels or, more usually, square blocks of N2 pixels [13]. Then, we assume that each block undergoes translation only with no scaling or rotation. The blocks in the first frame are compared to the blocks in the second frame. Motion Vectors can then be calculated for each block to see where each block from the first frame ends up in the second frame. For every video sequence we determine the number of shots, the number of shots correctly reported, the number of false detections and the number of non reported shots.

For each sequence we also draw the curve of the distances between the successive frames. These curves are mainly used to determine the threshold values, but they also give an idea about the capacity of the used technique in detecting transitions.

GIST Method

In this method we apply gist and color features to represent the individual frames. Gist [6],[7] has been shown to characterize the structure of images well while being resistant to luminance change and also small translation. Thus, it is ideally suited for our application of shot boundary detection. The most intuitive way to detect shot change is to compute the pixel difference of two consecutive frames, if the overall difference is larger than some threshold, it is detected as a shot boundary [14],[15]. This method results in a large number of vectors and is susceptible to camera motion and object motion. The sensitivity to motion is reduced by extracting features from the whole frame, among which histogram is the mostly used [16],[17],[18]. The disadvantage of global features is it tends to have low performance at detecting the boundary of two similar shots. To balance the tradeoff of resistance to motion and discriminating similar shots, a region based feature is proposed. Region-based method divides each frame into equal-sized blocks, and extracts a set of features per block [19],[20]. Based on the assumption that color content doesn’t change rapidly within but across shots, color is the mostly used features [14], others are edges and textures [19].

Assume that features computed in the first stage are similar within one shot but vary across shots, one can compute the distance of feature vectors of adjacent frames, and compare it against a threshold. A distance higher than the threshold usually corresponds to a hard cut. This fixed threshold is hard to detect gradual transitions because of its slow-changing nature [19]. In addition, most videos have variations in each frame and also sudden change within one shot, such as appearance of a new object. Computing the distance of multiple frames will cancel these variations [19]. Another method is to vary the threshold depending on the average distance within the current shot or the statistics of the whole video [15]. The problem was also formulated as a binary classification task in which two classes are ”transition” and ”no transition”. Within one shot, due to the motion, appearance or disappearance of objects, the color histogram may not be consistent. Gist captures the overall texture of the background while ignoring these small change due to foreground objects. A classifier trained to detect such rare occurrence would be biased to treat all frames as non-transition. But the classifier approach is impractical. Instead, we treat the task of detecting shot boundaries as an outlier detection problem. We use a simple z-score metric to determine outliers. The first strategy measures the z-score across all dimensions (assuming they are independent). The second strategy counts the number of dimensions in which the given data-point is an outlier. In practice, we observe that both strategies perform well. We define detected interval as the interval of the group of consecutive detected frames, and its width the number of frames it contains. For each detected interval, we mark it as cut if its width is higher than some threshold wt, as gradual transition otherwise.

This method proposed gist and color features for shot-boundary detection in video. This method posed the task as an outlier detection problem in the gradient domain and proposed a simple z-score based approach to solve the problem. We observe that gist is better than color features for detecting abrupt transitions. However, color performs better for gradual transitions.

CONCLUSIONS

In this paper we presented various different approaches used for video – shot boundary detection process and give a comparison between some of the most commonly employed methods. We can conclude from the above discussion that all these methods has some or the other problem and some of its advantages.

The method that seemed to produce good resuts were the GIST method. As it used two different approaches color and gist advantages at different stages.

REFERENCES

1. J. Boreczky and L. A. Rowe, “Comparison of video shot boundary detection techniques,” in Proc. SPIE, Storage and Retrieval for Still Image and Video Databases IV, vol. 2664, 1996, pp. 170–179.

2. R. Lienhart, “Comparison of automatic shot boundary detection algorithms,” in Proc. SPIE, Storage and Retrieval for Still Image and Video Databases VII, vol. 3656, 1998, pp. 290–301.

3. J. Bescs, G. Cisneros, J. M. Martnez, J. M. Menndez, and J. Cabrera, “A unified model for techniques on video shot transition detection,” IEEE Transactions on Multimedia, vol. 7, no. 2, pp. 293–307, Apr 2005.

4. J. Yuan, H. Wang, L. Xiao, W. Zheng, J. Li, F. Lin, and B. Zhang, “A formal study of shot boundary detection,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 2, pp. 168–186, Feb 2007.

5. Ahmet Ekin and A.Murat Tekalp, “Automatic Soccer video analysis and summarization, ”In Proceedings of SPIE Conference on storage and Retrieval for Media Databases 2003, volume 5021, pages 339-350, Santa Clara, CA, 20-24 January 2003.

6. Oliva and A. Torralba. Modeling the shape of the scene: a holistic representation of the spatial envelope. Int. Journal of Computer Vision., 42, 2001.

7. Torralba, A.Modeling global scene factors in attention. J. Opt. Soc. Am. A Opt. Image Sci. Vis. 20, 2003.

8. V. Vapnik, Statistical Learning Theory. Wiley Interscience, 1998.

9. R. Duda, P. Hart, and D. Stork, Pattern Classification, 2nd ed. Wiley Interscience, 2000.

10. Amel Malek Abdelati, Jlassi bahaeddine, Abdessalem Ben Abdellali and Abdellatif Mtibaa, " Suivi d'objet en movement par la method Level Set variationnelle, " international workshop on system engineering design & application, October 24-26, 2008, Monastir, Tunisia.

11. Report ,"State of the Art of Content Analysis Tools for Video, Audio and Speech, " Werner Bailer (JRS), Franz Holler (JRS), Alberto Messina (RAI), Daniele Airola (RAI), Peter Schallauer (JRS), Michael Hausenblas (JRS), 10/3/2005.

12. S. Jeannin and A. Divakaran, "MPEG-7 visual motion descriptors, " IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, pp. 720-724, 2001.

13. ISHIGURO, T., and IINUMA, K. "Television bandwidth compression transmission by motion-compensated interframe coding," IEEE Commun. Mag., 1982, 10, pp.24–30

14. P. Campisi, A. Neri, and L. Sorgi. Automatic dissolve and fade detection for video sequences. In Proc. Int. Conf. on Digital Signal Processing,2002.

15. W.J. Heng and K.N. Ngan. An object-based shot boundary detection using edge tracing and tracking. Journal of Visual Communication and Image Representation, 12(3), 2001.

16. Jun Yu and M. D. Srinath. An efficient method for scene cut detection. Pattern Recognition Letters, 22(13), 2001.

17. Z. Cernekova, C. Kotropoulos, and I. Pitas. Video shot segmentation using singular value decomposition. In Proc. 2003 IEEE Int. Conf. on Multimedia and Expo, volume II, 2003.

18. Rainer Lienhart. Reliable dissolve detection. In Storage and Retrieval for Media Databases 2001, volume 4315 of Proc. of SPIE, 2001.

19. Sarah Porter, Majid Mirmehdi, and Barry Thomas. Detection and classification of shot transitions. In Proc. of the 12th British Machine Vision Conf., 2001.

20. J.M. Sanchez and X. Binefa. Shot segmentation using a coupled Markov chains representation of video contents. In Proc. Iberian Conf. on Pattern Recognition and Image Analysis, June 2003.